

# mgcvUI : mgcv (廣義加法模型) 套件的互動式介面

繁體中文譯本，由人工智慧輔助翻譯並經編輯部審校；英文版為權威版本。

William Bert Craytor

第 1 卷，第 1 期 · 2026 年夏季

發表日期：2026 年 7 月 1 日

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

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

本開放取用文章依據 [創用 CC 姓名標示 4.0 國際授權條款 \(CC BY 4.0\)](#) 的條款發布。

**摘要。** mgcvUI 是 R 語言 mgcv 套件的圖形使用者介面 (graphical user interface)，該套件使用帶自動平滑度選擇 (smoothness selection) 的懲罰迴歸樣條 (penalized regression spline) 來擬合廣義加法模型 (generalized additive model, GAM)。它提供三種用途模式——通用預測建模、不動產估價與市場區域分析——並引導使用者完成資料匯入、平滑項設定、模型擬合、診斷圖與效應圖檢視以及報告下載。本文記錄了 mgcvUI 的資料格式要求、建模工作流程、輸出顯示介面以及完整的功能參考。

**關鍵詞：** 廣義加法模型，GAM，懲罰樣條，平滑化，mgcv，R，圖形使用者介面，不動產估價

出版者按 (2026 年 7 月)：付印前不久，mgcvUI 已併入 earthUI (0.11.0 及以後版本)，本文所述的 GAM 工作流程現與 MARS 和彈性網絡工作流程一同由 earthUI 提供。mgcvUI 仍以其最終發布形態 (0.3.0) 保留在 <https://github.com/wcraytor/mgcvUI>；後續開發將在 earthUI 中繼續。本文適用於已發布的 mgcvUI。

## 目錄

|          |                                   |           |
|----------|-----------------------------------|-----------|
| <b>1</b> | <b>引言</b>                         | <b>4</b>  |
| 1.1      | 歷史與背景                             | 4         |
| 1.1.1    | 廣義加法模型                            | 4         |
| 1.1.2    | mgcv 套件                           | 4         |
| 1.1.3    | 與 MARS (earth) 和彈性網絡 (glmnet) 的比較 | 4         |
| 1.1.4    | earthUI-mgcvUI 流程                 | 5         |
| 1.1.5    | 建議的工作流程                           | 5         |
| 1.2      | 什麼是 mgcvUI ?                      | 5         |
| 1.3      | 三種用途模式                            | 6         |
| 1.4      | 不動產專用功能                           | 6         |
| 1.5      | 入門                                | 6         |
| <b>2</b> | <b>MLS 輸入資料要求</b>                 | <b>8</b>  |
| 2.1      | 檔案格式與結構                           | 8         |
| 2.2      | 估價模式的必需欄位                         | 8         |
| 2.3      | 特殊欄位命名約定                          | 9         |
| 2.4      | 資料品質與完整性                          | 9         |
| 2.5      | 標的列的位置                            | 9         |
| <b>3</b> | <b>通用模式</b>                       | <b>10</b> |
| 3.1      | 概述                                | 10        |
| 3.2      | 側邊欄工作流程                           | 10        |
| 3.3      | 從 earthUI 匯入 (選用)                 | 11        |
| 3.4      | 主面板頁籤                             | 11        |

|          |                                            |           |
|----------|--------------------------------------------|-----------|
| 3.5      | 設定的持久化                                     | 12        |
| 3.6      | 深色模式                                       | 12        |
| 3.7      | 區域與地區設定                                    | 12        |
| <b>4</b> | <b>估價模式</b>                                | <b>14</b> |
| 4.1      | 標的列的處理                                     | 14        |
| 4.2      | 估價基準日與銷售時距                                 | 14        |
| 4.3      | 特殊欄位指定                                     | 14        |
| 4.4      | RCA 調整概覽                                   | 14        |
| <b>5</b> | <b>市場區域分析模式</b>                            | <b>15</b> |
| 5.1      | 與估價模式的差異                                   | 15        |
| 5.2      | 何時使用市場模式                                   | 15        |
| <b>6</b> | <b>變數選擇</b>                                | <b>16</b> |
| 6.1      | 目標變數                                       | 16        |
| 6.2      | 反應變數變換                                     | 16        |
| 6.3      | 預測變數表                                      | 16        |
| 6.4      | 資料型別偵測與覆寫                                  | 17        |
| 6.5      | 特殊欄位類型參考                                   | 17        |
| <b>7</b> | <b>參數選擇</b>                                | <b>18</b> |
| 7.1      | 參數預設組                                      | 18        |
| 7.2      | Family (分配族)                               | 18        |
| 7.3      | Method (方法)                                | 18        |
| 7.4      | Gamma (平滑度乘子)                              | 18        |
| 7.5      | 交叉驗證                                       | 19        |
| 7.6      | Select (收縮)                                | 19        |
| 7.7      | 預設基底                                       | 19        |
| 7.8      | 預設 k (基底維度)                                | 19        |
| 7.9      | 張量類型                                       | 19        |
| 7.10     | 允許的交互                                      | 20        |
| 7.11     | 因子-平滑交互                                    | 20        |
| 7.12     | 進階參數                                       | 20        |
| <b>8</b> | <b>擬合模型</b>                                | <b>21</b> |
| 8.1      | Fit 按鈕                                     | 21        |
| 8.2      | 擬合過程中發生了什麼                                 | 21        |
| 8.3      | 背景擬合與進度顯示                                  | 21        |
| <b>9</b> | <b>結果頁籤</b>                                | <b>22</b> |
| 9.1      | Data (資料)                                  | 22        |
| 9.2      | Equation (方程)                              | 22        |
| 9.3      | Summary (摘要)                               | 22        |
| 9.4      | Variable Importance (變數重要性)                | 22        |
| 9.5      | 貢獻圖                                        | 22        |
| 9.5.1    | 單變數平滑項 — $s(x)$                            | 23        |
| 9.5.2    | 因子-平滑項 — $s(x, \text{by} = \text{factor})$ | 23        |
| 9.5.3    | 交互項 — $ti(x_1, x_2)$ 與 $te(x_1, x_2)$      | 23        |
| 9.5.4    | 參數項 — 線性項與因子項                              | 23        |
| 9.6      | Correlation (相關性)                          | 23        |
| 9.7      | Diagnostics (診斷)                           | 23        |

|           |                                 |           |
|-----------|---------------------------------|-----------|
| 9.8       | RCA Adjustments (RCA 調整)        | 23        |
| 9.9       | Anova (變異數分析)                   | 24        |
| 9.10      | Mgcv Output (mgcv 輸出)           | 24        |
| 9.11      | Sign Check (符號檢查)               | 24        |
| 9.12      | Concurvity (共曲性)                | 24        |
| 9.13      | Basis Check (基底檢驗)              | 24        |
| <b>10</b> | <b>下載資料</b>                     | <b>25</b> |
| 10.1      | 輸出欄位                            | 25        |
| 10.2      | CQA 得分                          | 25        |
| <b>11</b> | <b>RCA 計算與下載</b>                | <b>26</b> |
| 11.1      | 開啟 RCA 對話方塊                     | 26        |
| 11.2      | CQA 得分內插                        | 26        |
| 11.3      | 輸出欄位                            | 26        |
| <b>12</b> | <b>銷售比較網格</b>                   | <b>27</b> |
| 12.1      | 可比案例選擇對話方塊                      | 27        |
| 12.2      | 活頁簿版式                           | 27        |
| 12.3      | 工作表保護與殘差儲存格                     | 27        |
| <b>13</b> | <b>下載報告</b>                     | <b>28</b> |
| 13.1      | 第 1 步：產生 Quarto 報告              | 28        |
| 13.2      | 第 2 步：轉換 Quarto 報告              | 28        |
| 13.3      | 報告內容                            | 28        |
| <b>14</b> | <b>匯出模型函數</b>                   | <b>29</b> |
| 14.1      | R 函數                            | 29        |
| 14.2      | Python 函數                       | 29        |
| 14.3      | C++ 標頭檔                         | 29        |
| 14.4      | SQLite 資料庫                      | 29        |
| <b>15</b> | <b>與 earthUI 和 glmnetUI 的比較</b> | <b>30</b> |
| 15.1      | 主要差異                            | 30        |
| 15.2      | 共有功能                            | 30        |
| 15.3      | 建議的工作流程                         | 30        |
| <b>16</b> | <b>示範資料集：Appraisal_1.csv</b>    | <b>31</b> |
| 16.1      | 說明                              | 31        |
| 16.2      | 各欄位                             | 31        |
| 16.3      | 建議的快速上手步驟                       | 32        |
| <b>17</b> | <b>系統需求與疑難排解</b>                | <b>33</b> |
| 17.1      | 受支援的平台                          | 33        |
| 17.2      | R 版本                            | 33        |
| 17.3      | 必需的 R 套件                        | 33        |
| 17.4      | 選用相依套件                          | 33        |
| 17.5      | 與平台相關的說明                        | 33        |
| 17.5.1    | macOS                           | 33        |
| 17.5.2    | Windows                         | 34        |
| 17.5.3    | Linux                           | 34        |
| 17.6      | 優雅降級                            | 34        |
| 17.7      | 疑難排解                            | 34        |

|                                                                            |    |
|----------------------------------------------------------------------------|----|
| 17.7.1 「PDF option is not available」(PDF 選項不可用)                            | 34 |
| 17.7.2 「Settings will not persist」(設定將不會儲存)                                | 35 |
| 17.7.3 「Diagnostics plot unavailable」(診斷圖不可用)                              | 35 |
| 17.7.4 「Report error: number of items to replace...」(報告錯誤：待替換的項數<br>.....) | 35 |
| 17.7.5 7880 連接埠已被佔用                                                        | 35 |
| 17.7.6 在 Linux 上安裝套件失敗                                                     | 35 |

## 1 引言

### 1.1 歷史與背景

#### 1.1.1 廣義加法模型

廣義加法模型 (generalized additive model, GAM) 由 Trevor Hastie 與 Robert Tibshirani 在其 1990 年的專著中正式提出。GAM 對線性模型作了推廣：把每一個線性項  $\beta_j x_j$  替換為一個平滑函數 (smooth function)  $f_j(x_j)$ ，於是模型成為：

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

每個  $f_j$  都用懲罰迴歸樣條 (penalized regression spline) 從資料中估計得到。這些平滑函數使模型無需使用者事先指定函數形式即可刻畫非線性關係。懲罰項控制每個平滑項的擺動程度 (wiggliness)，在保留足夠彈性以追蹤真實模式的同時防止過擬合。

#### 1.1.2 mgcv 套件

R 套件 `mgcv` (Mixed GAM Computation Vehicle) 由巴斯大學的 Simon Wood 開發 (Wood 2017)。它於 2001 年首次發布，如今已是 GAM 在 R 中的標準實作，並隨每一份 R 安裝一同發行。其底層方法見於一系列論文，涵蓋穩定的多平滑參數估計 (Wood 2004)、薄板迴歸樣條 (Wood 2003)、快速穩定的 REML 估計 (Wood 2011) 以及一般平滑模型的選擇 (Wood, Pya, and Säcken 2016)。主要特性包括：

- 自動平滑度選擇 (smoothness selection)：每一項的平滑參數都作為模型擬合的一部分加以估計，可採用 REML (受限最大概似)、GCV (廣義交叉驗證) 或 ML (最大概似)。
- 多種基底類型：薄板迴歸樣條 (tp)、三次迴歸樣條 (cr)、P 樣條 (ps) 與 B 樣條 (bs)，各具不同性質。
- 張量積平滑：`te()` 與 `ti()` 可實現變數間的平滑交互，而不受維度詛咒之困。
- 變數選擇：`select = TRUE` 選項會附加一個額外懲罰，可將平滑項收縮至零，從而實際完成變數選擇。
- 共曲性診斷 (concurvity diagnostics)：度量平滑項中與多重共線性相對應的現象。

#### 1.1.3 與 MARS (earth) 和彈性網絡 (glmnet) 的比較

`earth` 應用程式及其兩個後處理模組 (`glmnet` 與 `mgcv`) 使用不同的建模引擎。下表概括了主要差異：

| 特性     | mgcv (GAM)                   | earth (MARS)              | glmnet (彈性網絡)              |
|--------|------------------------------|---------------------------|----------------------------|
| 模型類型   | 平滑非線性 (懲罰樣條)                 | 分段線性 (自適應樣條)              | 線性 (帶正則化)                  |
| 非線性    | 透過平滑函數自動實現 (薄板、三次等)          | 透過帶資料驅動節點的鉸鏈函數自動實現        | 僅能透過交互項或基底展開實現             |
| 變數選擇   | 選用, 透過雙重懲罰收縮實現 (select=TRUE) | 內建, 透過前向/後向逐步法配合 GCV 修剪實現 | 內建, 透過 $L_1$ 懲罰 (Lasso) 實現 |
| 交互項    | 張量積平滑 (te()、ti())            | 基於鉸鏈函數的乘積 (可解釋, 最高三階)     | 兩兩交叉乘積 (難以解釋)              |
| 可解釋性   | 平滑的偏效應曲線; 非常直觀               | g 函數按變數給出分段線性偏效應          | 係數即線性權重; 易於逐一說明            |
| 最適用於   | 平滑的非線性關係、彈性建模                | 自動偵測非線性與交互                | 高維資料、變數選擇、小樣本              |
| RCA 調整 | 透過偏效應給出平滑調整                  | 透過 g 函數給出分段線性調整           | 按變數給出線性調整                  |
| 過擬合控制  | REML/GCV 平滑參數估計              | 基於 GCV 的修剪                | 交叉驗證選取的 lambda             |

**MARS** (Multivariate Adaptive Regression Splines, 多元自適應迴歸樣條) 由 Jerome Friedman 於 1991 年提出。它透過從資料中自適應地選擇鉸鏈函數及其節點 (knot) 位置來建構分段線性模型。其 R 實作是 Stephen Milborrow 編寫的 **earth** 套件。

彈性網絡 (elastic net) 迴歸由 Friedman、Hastie 與 Tibshirani (2010) 實作於 **glmnet** 套件中, 它把 Lasso ( $L_1$ ) 懲罰與嶺迴歸 ( $L_2$ ) 懲罰結合起來, 同時完成變數選擇與係數收縮。

#### 1.1.4 earthUI-mgcvUI 流程

一套強有力的工作流程把 earthUI 的自動節點發現與 mgcvUI 的平滑估計結合起來。當 earthUI 匯出一個 .rds 結果檔案而 mgcvUI 將其匯入時:

1. **earth** 發現節點 —— MARS 為每個變數找出由資料驅動的鉸鏈位置 (變點)。
2. **mgcvUI** 以其為錨點 —— earth 的節點成為 GAM 中三次迴歸樣條 (cr 基底) 的基礎, 使平滑項一開始就擁有反映資料真實結構的節點位置。
3. **mgcv** 精化擬合 —— 懲罰樣條估計把 earth 的分段線性擬合加以平滑化, 在保留 earth 所發現的整體形狀的同時, 給出平滑的偏效應曲線。

這條流程把探索性建模 (earth) 與驗證性建模 (GAM) 銜接起來, 兼取兩種方法之長。

#### 1.1.5 建議的工作流程

對於不動產估價及其他要求模型可解釋、可辯護的類似應用:

1. 先用 **earthUI** 發現資料中的重要變數、非線性關係與交互作用。earth 的自動基底選擇與 g 函數能提供極好的初步洞見。
2. 再用 **mgcvUI** 精化, 如果你希望得到更平滑的偏效應。把 earthUI 的節點匯入 mgcvUI, 即可從分段線性模型順暢過渡到平滑模型。
3. 使用 **glmnetUI**, 當你需要激進的變數選擇 (Lasso)、預測變數數目接近或超過觀測數目, 或者你希望得到一個純線性且帶正則化、便於辯護的模型時。

#### 1.2 什麼是 mgcvUI?

mgcvUI 是 R 語言 mgcv 套件的圖形使用者介面 (graphical user interface)。它以本機 Shiny 應用程式的形式執行——無需登入、無需伺服器、也沒有帳號。你從 R 中啟動它, 匯入資料集 (CSV 或 Excel), 設定模型, 然後互動式地擬合 GAM。

自目前版本起，mgcvUI 不再是單獨安裝的套件：它隨 earthUI 套件一同發行，作為其中的 GAM 後處理模組（Post Processing Module），用 `earthUI::launch_mgcv()` 啟動。應用程式本身——以及本文所述的一切——並無改變；改變的只是安裝與啟動指令。三個例程（`earth`、`glmnet`、`mgcv`）共用一棵專案樹、一個設定資料庫以及同一份支撐基礎設施。

該應用程式提供完整的工作流程：資料匯入、含平滑項設定的變數設定、帶背景處理的模型擬合、診斷圖、平滑偏效應曲線，以及可下載的 Word、PDF 或 HTML 格式報告。

廣義加法模型把每個預測變數的線性關係  $\beta x$  替換為平滑函數  $f(x)$ 。這意味著：

- 無需指定函數形式——每種關係的形狀由資料本身決定。
- 平滑的偏效應——每個預測變數的貢獻是一條平滑曲線而非直線，便於看清某變數在不同取值處如何影響反應變數。
- 自動平滑度選擇——mgcv 透過 REML 或 GCV 確定每條曲線應有多擺動，在擬合優度與過擬合之間取得平衡。

### 1.3 三種用途模式

每個專案都有一個用途（purpose），在側邊欄第 1 節（**Project**，專案）中建立專案時選定，此後即被鎖定。目前作用中專案的用途決定哪些工具與介面元素出現；若要在別的模式下工作，請建立或開啟一個具有該用途的專案。三種模式為：

- **General**（通用）——適用於任何類型母體或資料集的 GAM 迴歸。這是預設模式，提供完整的 GAM 工作流程，不含任何領域專用的附加功能。
- **For Appraisal**（用於估價）——為不動產估價量身打造的 GAM 迴歸。增加了單一不動產估價所特有的功能，包括估價標的處理、特殊欄位指定以及殘差約束法（Residual Constraint Approach, RCA）。
- **Market Area Analysis**（市場區域分析）——為市場區域研究量身打造的 GAM 迴歸。增加了在既定市場中分析成組不動產的功能，並可選擇處理標的列。

在三種模式下，核心建模引擎完全相同——你始終是在透過 `mgcv::gam()` 擬合 GAM。用途設定只控制哪些附加工具與介面元素可用。

### 1.4 不動產專用功能

當選定 **For Appraisal** 或 **Market Area Analysis** 時，mgcvUI 會啟用若干為不動產分析而設計的功能：

- **特殊欄位指定**——每個預測變數都可以被標記上某種特殊角色，例如 `contract_date`、`dom`、`concessions`、`latitude`、`longitude`、`living_area`、`lot_size`、`actual_age`、`effective_age`、`area`、`site_dimensions` 或 `display_only`。這些指定決定該欄位在擬合與輸出過程中的處理方式。
- **Sale Age 欄位**（銷售時距）——當某欄位被指定為 `contract_date` 且提供了 `Effective Date`（估價基準日）時，mgcvUI 會計算一個 `sale_age` 欄位（成交日期與基準日之間的天數），並以它作為預測變數代入模型。
- **RCA 計算**（僅估價模式）——在估價模式下，模型擬合完成後 mgcvUI 可計算殘差約束法（RCA）輸出，給出逐個可比案例的調整額、淨調整與總調整彙總，以及估價標的的調整後銷售價格。

提示：*General* 用途模式適用於任何資料集——金融、科學、工程或不動產。估價模式與市場模式只是為不動產專業人士增添了一些便利功能。

### 1.5 入門

使用 mgcvUI 的步驟：

1. 在 R 中安裝 **earthUI**：`install.packages("earthUI")` ——**mgcv** 例程已包含在內。(開發版與選用元件請參見本期 **earthUI** 一文中的「Installing earthUI and vProlog」一節。)
2. 啟動應用程式：在 R 中執行 `earthUI::launch_mgcv()`。應用程式會在你的網頁瀏覽器中透過 7880 連接埠開啟。
3. 在側邊欄第 1 節建立或開啟專案。專案設定用途 (General、For Appraisal 或 Market Area Analysis) 與工作地點 (國家、州、縣、市) 及其輸入/輸出資料夾。專案存放在共用的 **regProj** 根目錄之下，並與同族應用 **earthUI** 和 **glmnetUI** 共用。開啟一個用途不同的專案，正是切換模式的方式。
4. 在第 2 節匯入資料：從目前作用中專案的輸入資料夾中選擇一個 CSV 或 Excel 檔案。把檔案放入該資料夾，然後點選 **Refresh** (重新整理) 以列出它們。
5. 從 **earthUI** 匯入 (選用) ——第 3 節允許你匯入一個 **earthUI** 結果 **.rds** 檔案，用 **earth** 發現的節點為 **GAM** 提供初值。若你沒有 **earthUI** 結果，跳過此步即可。
6. 設定變數 ——選擇目標變數與預測變數，設定資料型別，指派特殊欄位角色，並將變數標記為因子型或線性。
7. 設定 **mgcv** 參數 ——**family**、**method**、**gamma**、基底類型、**k**、張量類型、交互項以及其他選項。
8. 擬合模型 ——點選「Fit **Mgcv** **GAM** Model」(擬合 **mgcv** **GAM** 模型)，並在主面板中檢視結果。
9. 匯出 ——把預測結果下載為 Excel，產生並轉換 **Quarto** 報告，或 (在估價模式下) 計算 **RCA** 調整額與銷售比較網絡。

設定會自動儲存在一個 **SQLite** 資料庫中，並在你重新載入同一輸入檔案時復原。

## 2 MLS 輸入資料要求

對於不動產估價與市場分析工作流程而言，你的輸入資料通常來自多重上市服務（Multiple Listing Service，MLS）的匯出檔案。本章說明預期的檔案結構以及 mgcvUI 可以使用的各欄位。

### 2.1 檔案格式與結構

mgcvUI 接受 CSV 與 Excel (.xlsx、.xls) 檔案。匯入時，欄位名會自動轉換為 snake\_case——例如，「Living SqFt」變為 living\_sqft，「Contract Date」變為 contract\_date，「Sale Price」變為 sale\_price。這一規範化確保整個工作流程中欄位引用的一致性。匯入時使用的 CSV 分隔符與小數點符號由區域設定決定（參見第 3 章「Locale & Regional Settings」，區域與地區設定）。

你的資料檔案應是一張扁平表格，每個不動產佔一行，每個屬性佔一欄。檔案的第一列必須是欄位標題。

### 2.2 估價模式的必需欄位

雖然 mgcvUI 可以處理任意一組欄位，但完整的估價工作流程若具備以下各欄位將獲益更多：

| 欄位                     | 特殊類型            | 用途                              |
|------------------------|-----------------|---------------------------------|
| Sale Price（銷售價格）       | （目標變數）          | 模型的反應變數                         |
| Contract Date（成交日期）    | contract_date   | 用於計算 sale_age（距估價基準日的天數）        |
| Listing Date（掛牌日期）     | listing_date    | 在缺少 DOM 欄位時，與成交日期一起用於計算 DOM     |
| Days on Market（市場停留天數） | dom             | 市場停留天數；在匯出檔案中顯示                 |
| Concessions（成交讓價）      | concessions     | 銷售讓價；淨售價 = 銷售價格 - 讓價            |
| Living Area (SF)（居住面積） | living_area     | 啟用按平方英尺計的殘差（residual_sf、cqa_sf） |
| Lot Size（地塊面積）         | lot_size        | 宗地面積欄位                          |
| Site Dimensions（宗地尺寸）  | site_dimensions | 與地塊面積歸為一組                       |
| Latitude（緯度）           | latitude        | 四捨五入至 3 位小數                     |
| Longitude（經度）          | longitude       | 四捨五入至 3 位小數                     |
| Area ID（區域識別碼）         | area            | 市場區域 / 街區識別碼                    |
| Actual Age（實際年齡）       | actual_age      | 不動產年齡                           |
| Effective Age（有效年齡）    | effective_age   | 不動產的有效年齡                        |
| Address（地址）            | display_only    | 在匯出檔案中顯示；不進入模型                  |

試算表的欄位名可以使用外語——而「特殊」名稱使用英文，以便 R 程式對其作特殊處理。除此之外，你所給定的欄位名會出現在迴歸模型、圖形中，若是做估價，還會出現在輸出報告中。

並非所有欄位都是必需的。mgcvUI 會自行適應——若某欄位遺漏，相應功能便自動略去。不過，對於不動產定價模型，以下幾個欄位強烈建議：

1. **Sale Age**（銷售時距）——成交日期與估價基準日之間的天數。若使用跨越數年的銷售歷史，`sale_age` 往往起著核心作用。
2. **Living Area**（居住面積）——也常被稱作「Living Sqft」「GLA」（gross living area，總居住面積）。
3. **Total Bath Count**（衛浴總數）——全衛、四分之一衛、半衛與四分之三衛的總數。
4. **Garage Bays**（車庫車位）或 **Garage Area**（車庫面積）——車庫車位數或建築面積。
5. **Lot Size**（地塊面積）——不動產的土地面積。
6. **Longitude, Latitude**（經度、緯度），以及若有的話 **Area ID**（區域識別碼），用於刻畫地理上的價格變化。

## 2.3 特殊欄位命名約定

mgcvUI 依據欄位的特殊類型指定而非欄位名來識別各欄位。你可以在 MLS 匯出檔案中隨意命名欄位——關鍵在於你要在變數設定表（第 6 章）中指派正確的特殊類型。

## 2.4 資料品質與完整性

- **遺漏值 (NA)**：任一預測變數欄位或目標欄位中含 NA 的列會在擬合前自動刪除。mgcvUI 會把資料子集化為模型實際使用的那些欄位，因此未被使用的欄位中的 NA 不會導致丟列。
- **數值欄位**：銷售價格、居住面積、地塊面積等類似欄位必須為數值型。
- **因子欄位**：文字欄位、TRUE/FALSE 欄位以及 R 因子欄位會被自動視為分類變數——它們只能以因子形式進入模型。一個數值欄位（例如數值型的區域碼或風格碼）只有在你勾選其 **Factor**（因子）核取方塊時才會成為因子；mgcvUI 絕不會根據數值的取值範圍推斷其為分類變數，因此像 `bath_count` 這樣的離散數值預測變數仍保持連續（並被平滑化）。Special（特殊）角色與因子的判定彼此獨立。

提示：匯入後請檢視預測變數表中的 NA 欄位。遺漏值很多的欄位可能導致丟列。NA 佔比  $\geq 50\%$  的欄位會被標紅並自動排除。

## 2.5 標的列的位置

在 **Appraisal**（估價）模式下，第 1 列必須是估價標的。其餘各列均為可比銷售案例。標的列不參與模型擬合。擬合完成後，模型仍會為標的列產生預測值。

在 **Market Area Analysis**（市場區域分析）模式下，是否把估價標的置於第 1 列可自行決定。

在 **General**（通用）模式下沒有特殊的列處理——所有列一視同仁。

### 3 通用模式

#### 3.1 概述

啟動 mgcvUI 時，通用 (General Purpose) 模式是預設模式。它為任何資料集——而不僅僅是不動產資料——提供完整的 GAM 工作流程。你可以把 mgcvUI 用於科學資料、金融分析、工程研究，或任何預期存在平滑非線性關係的迴歸問題。

在 General 模式下，介面略去不動產專用功能 (特殊欄位、銷售時距、RCA)。側邊欄被精簡，專注於變數選擇、參數設定、模型擬合與匯出。

#### 3.2 側邊欄工作流程

側邊欄按編號組織為若干可折疊小節，引導你從專案選擇一路走到報告匯出。多數小節只有在開啟一個作用中專案後才出現，其中若干還與專案的用途相關。輸出資料夾由作用中專案推導得出——並沒有單獨的輸出資料夾欄位。

**1. Project (專案)** ——建立或開啟一個專案。專案設定用途與工作地點 (國家、州、縣、市)，以及它的輸入 (`<os>_in/`) 與輸出資料夾。專案存放在共用的 `regProj` 根目錄下 (預設為 `~/regProj`，可透過 `REGPROJ_ROOT` 環境變數或 Settings 下拉選單覆寫)，並與同族應用 `earthUI` 和 `glmnetUI` 共用，因此三者的模型可以並列存放。點選 **Close Project** (關閉專案) 即可切換。

**2. Import Data (匯入資料)** ——從作用中專案的輸入資料夾中選擇一個 CSV 或 Excel 檔案。先透過 Finder/檔案總管把檔案放入該資料夾，然後點選 **Refresh** (重新整理) 以列出它們。對於含多個工作表的 Excel 檔案，會出現一個工作表選擇器。欄位名會自動轉換為 `snake_case`。

**3. Import from earthUI (optional) (從 earthUI 匯入，選用)** ——匯入一個 `earthUI` 結果 `.rds` 檔案，用 `earth` 發現的節點位置為 GAM 提供初值 (同時也為變數選擇提供初值)。此步驟是選用的——若你沒有可匯入的 `earthUI` 結果，跳過即可。

**4. Variable Configuration (變數設定)** ——目標變數選擇器、反應變數變換 (無、log、log10)，以及帶 Include (納入)、Factor (因子) 和 Linear (線性) 核取方塊的預測變數表。Special (特殊) 欄位只在估價模式與市場模式下出現。詳見第 6 章。

**5. MgcV Call Parameters (mgcv 呼叫參數)** ——全部模型設定：參數預設組、family、method、gamma、交叉驗證、select、基底類型、k、張量類型、交互矩陣以及進階參數。完整的參數參考見第 7 章。

**6. Fit MgcV GAM Model (擬合 mgcv GAM 模型)** ——執行模型的按鈕。結果會顯示在右側的各個頁籤中。

**7. Download Output (下載輸出)** ——把預測值、殘差、CQA 得分以及各變數的貢獻匯出為 Excel 檔案，存入專案的輸出資料夾。模型擬合之後，在所有用途模式下均可使用。

**8. Calculate RCA Adjustments & Download (計算 RCA 調整額並下載)** ——為每個可比案例計算逐變數的 RCA 調整額，並透過 CQA 對估價標的殘差作內插。僅在估價模式與市場模式下、且已完成擬合後出現。

**9. Generate Sales Grid & Download (產生銷售網格並下載)** ——產生銷售比較網格 (Sales Comparison Grid)，按總調整百分比建議可比案例。僅在估價模式下出現。

**10. Generate Quarto Report (產生 Quarto 報告)** ——把一個自包含的 Quarto (`.qmd`) 報告包——含全部圖形與表格——寫入專案的輸出資料夾。模型擬合之後即可使用。

**11. Convert Quarto Report (轉換 Quarto 報告)** ——把產生的 `.qmd` 包渲染為 HTML、Word 或 PDF。PDF 需要安裝 LaTeX；若未偵測到 LaTeX，PDF 選項會被隱藏。參見第 13 章。

在其他模式下，與用途相關的步驟會被略去 (其餘步驟重新編號)：市場模式下沒有銷售網格步驟，通用模式下 RCA 與銷售網格兩個步驟都沒有。

### 3.3 從 earthUI 匯入 (選用)

側邊欄的第 3 節——**Import from earthUI**——允許你匯入一個 earthUI 結果 .rds 檔案。這就啟用了 earth-mgcv 流程：earth 由資料驅動得出的節點位置成為 GAM 平滑項的錨點。

匯入 earthUI 結果後，mgcvUI 會：

- 顯示一份摘要，列出目標變數、R 平方、項數以及每個變數的節點位置。
- 自動選中「Earth Pipeline」(earth 流程) 參數預設組 (三次迴歸樣條基底， $\gamma = 1.4$ )。
- 把目標變數設為與 earth 的目標變數一致。
- 為 earth 的預測變數預先勾選 Include。
- 為 earth 的分類變數預先勾選 Factor，為 earth 的線性預測變數預先勾選 Linear。
- 預先填入 earth 偵測到的交互項 (在交互矩陣中以黃色高亮顯示，並鎖定不允許手工更改)。
- 提供一個 **Export Knots CSV** (匯出節點 CSV) 按鈕，可把 earth 的節點下載為 CSV 檔案。

隨後，變數設定表會顯示一條說明，指出變數選擇 (Include、Linear、Factor) 與樣條節點是由 earth 模型提供初值的。與 glmnetUI 不同——在 glmnetUI 中 glmnet 的擬合被綁定到 earth 的基底上——mgcvUI 是依據你目前的選擇來擬合 GAM 的，earth 只是提供了一個起步優勢，而每一個控制項都仍可編輯。

一個 **Clear** (清除) 按鈕可移除已匯入的 earth 資料，並重置為獨立執行模式。

### 3.4 主面板頁籤

資料匯入之後，主面板提供以下頁籤 (取決於模型的頁籤在擬合完成後才有內容)：

| 頁籤                          | 內容                                                        |
|-----------------------------|-----------------------------------------------------------|
| Data (資料)                   | 已匯入資料的互動式 DataTable。在估價模式與市場模式下，分為估價標的表與可比銷售表。            |
| Equation (方程)               | 用 MathJax 渲染的完整 GAM 公式，以及一張平滑函數定義表。                       |
| Summary (摘要)                | 關鍵指標 ( $R^2$ 、CV $R^2$ 、已解釋離差、AIC、n、平滑項個數) 以及平滑項 / 參數項表格。 |
| Variable Importance (變數重要性) | F 統計量 (平滑項) 與 $ t $ 值 (參數項) 的水平長條圖，以及排序表。                 |
| Contribution (貢獻)           | 互動式 plotly 平滑偏效應曲線，帶信賴帶與 earth 節點標記。                      |
| Correlation (相關性)           | 數值型預測變數相關係數的熱圖 (擬合前即可檢視)。                                 |
| Diagnostics (診斷)            | 殘差對擬合值、Q-Q 圖、直方圖、反應值對擬合值 (透過 gratia 產生)。實際值對預測值散點圖。       |
| RCA Adjustments (RCA 調整)    | RCA 分析直方圖 (估價 / 市場模式，計算完成後)。                              |
| Anova (變異數分析)               | 參數項與平滑項合併的變異數分析表。                                         |
| Mgcv Output (mgcv 輸出)       | 原始文字：計時、公式、模型摘要、family、共曲性。                               |
| Sign Check (符號檢查)           | earth 與 GAM 的方向一致性檢查 (當已匯入 earth 結果時)。                    |
| Concurvity (共曲性)            | 整體與兩兩共曲性矩陣，帶顏色編碼。                                         |
| Basis Check (基底檢驗)          | <code>gam.check()</code> 的基底維度充分性檢驗。                      |

### 3.5 設定的持久化

mgcvUI 會自動把你的設定儲存到一個 SQLite 資料庫中，並以輸入檔案名為鍵。當你重新載入同一檔案時，所有設定都會復原：目標變數選擇、預測變數核取方塊、資料型別、mgcv 參數以及交互矩陣。用途是作用中專案的屬性而非全域設定，最近使用的輸入檔案按專案分別記住，並在專案重新開啟時自動載入。

### 3.6 深色模式

點選右上角的月亮 / 太陽圖示，可在 Nord Light 與 Nord Dark 兩種主題之間切換。主題偏好儲存在 localStorage 中，跨會話保持。所有介面元素都會隨所選主題相應變化。

### 3.7 區域與地區設定

mgcvUI 透過一套基於國家的區域設定 (locale) 系統支援國際化的數字與 CSV 格式約定。標題欄中的 **Settings** (設定) 下拉選單為 30 多個受支援國家提供 **Country** (國家) 與 **Paper** (紙張) 選擇器。每種預設設定以下各項：

- **CSV 分隔符** —— 美國 / 英國 / 日本用逗號 (,)，歐洲大部分地區用分號 (;)。
- **小數點符號** —— 句點 (.) 或逗號 (,)。
- **千位分隔符** —— 逗號 (美國 / 英國 / 日本)、句點 (德國 / 義大利 / 西班牙)、空格 (芬蘭 / 法國 / 波蘭 / 波羅的海國家)，或撇號 (瑞士)。
- **紙張尺寸** —— Letter (美國 / 加拿大 / 墨西哥) 或 A4 (其他所有地區)。

點選 **Save**（儲存）即可把你的區域偏好全域存儲。Import Data 小節還提供一個按檔案生效的 **Locale**（區域）選擇器，供採用不同約定的一次性匯入使用。同一個 Settings 下拉選單中還有 **regProj root folder**（regProj 根資料夾）欄位，用於重新指定 mgcvUI、earthUI 與 glmnetUI 共用的專案樹位置。

提示：圖形座標軸與斜率標籤上的數字格式會自動適應你的區域設定。德語區域用句點作千位分隔（200.000），芬蘭用空格（200 000），瑞士用撇號（200'000）。介面不顯示任何貨幣符號——mgcvUI 與貨幣無關。

## 4 估價模式

當作用中專案的用途為 **For Appraisal**（用於估價）時，mgcvUI 會把自身設定為面向單一不動產估價。第 3 章所述的全部功能依然可用；本章只講估價模式特有的附加內容。

### 4.1 標的列的處理

在估價模式下，資料集的第 1 列即估價標的，其餘各列為可比銷售案例。你的輸入檔案必須據此組織（參見第 2 章）。估價標的的銷售價格可以留空，也可以填任意值——mgcvUI 在擬合時會自動把它當作 NA 處理。

擬合完成後，模型會為標的列產生預測值，這是 RCA 調整工作流程的基礎。

### 4.2 估價基準日與銷售時距

在估價模式與市場模式下，變數設定小節中會出現一個 **Effective Date**（估價基準日）欄位（預設為當天日期）。若你在 Special 下拉選單中把某欄位指定為 `contract_date`，mgcvUI 就會計算一個 `sale_age` 欄位——即每筆交易的成交日期與基準日之間的整天數。該欄位會作為預測變數加入模型。

當估價基準日改變時，`sale_age` 會自動重新計算。

### 4.3 特殊欄位指定

在估價模式與市場模式下，變數設定表中的每個預測變數都會出現一個 **Special**（特殊）下拉選單。特殊類型及其作用的完整參考見第 6 章。

### 4.4 RCA 調整概覽

**Calculate RCA Adjustments & Download**（計算 RCA 調整額並下載）按鈕（在估價模式與市場模式下、擬合完成後可見）會為每個可比案例相對於估價標的計算由市場推導出的調整額。完整的 RCA 工作流程見第 11 章。

提示：你為估價標的指定的 *CQA* 得分決定了殘差分配中有多大比例被歸屬於估價標的。得分 5.00 會把估價標的置於各可比案例的中位數位置。

## 5 市場區域分析模式

當作用中專案的用途為 **Market Area Analysis**（市場區域分析）時，mgcvUI 提供與估價模式相同的不動產專用功能（特殊欄位、銷售時距），但其取向是分析一組不動產，而非評估單一估價標的的價值。

### 5.1 與估價模式的差異

- 標的列的處理更為彈性 —— 在市場模式下，第 1 列不會被自動當作估價標的。
- 沒有銷售網格 —— 「Generate Sales Grid & Download」步驟不可用。

### 5.2 何時使用市場模式

在以下情形下適合使用市場區域分析模式：

- 為某個街區或市場區域建立 GAM 模型，以理解非線性的價值驅動因素
- 分析建築面積、年齡、地塊面積與區位等變數如何影響一組不動產的銷售價格
- 為市場狀況分析或街區界定準備支撐材料
- 處理這樣一份資料集：你想要特殊欄位功能（銷售時距），但不需要銷售網格工作流程

提示：市場模式對一般性的不動產迴歸也很有用——當你想要特殊欄位功能，但不需要單一不動產估價工作流程時。

## 6 變數選擇

側邊欄的第 4 節——**Variable Configuration**（變數設定）——是你選定哪些欄位參與模型、以及它們如何被處理的地方。

### 6.1 目標變數

第 4 節頂部的 **Target (response) variable**（目標 / 反應變數）下拉選單列出資料集中的每一個數值欄位。選擇其中一欄作為反應變數（例如銷售價格）。目標欄位會自動從預測變數清單中排除。

### 6.2 反應變數變換

在目標變數選擇器下方，**Response Transform**（反應變數變換）下拉選單提供三個選項：

- **None (raw)**（無，原值）——直接使用反應變數。當預測變數與反應變數在原始尺度上近似可加時，此選項合適。
- **Log (natural)**（自然對數）——在擬合前施加自然對數變換。反變換之後，這會使調整額成為比例式（%）而非絕對額（\$）。這對銷售價格尤其有用，因為一筆 \$10,000 的調整在不同價格水準上含義大不相同。
- **Log10**（以 10 為底的對數）——施加以 10 為底的對數變換。

提示：對數變換會把時間調整變為比例式（%）而非絕對額（\$）。對於價格跨度很大的不動產模型，對數變換往往能產生性態更好的殘差。

選擇對數變換後，反應變數中  $\leq 0$  的取值會被自動濾除。所有輸出（預測值、貢獻、殘差）都會自動反變換回原始尺度。

### 6.3 預測變數表

在目標變數選擇器下方，一張可捲動的表格列出其餘所有欄位。核取方塊欄位使用旋轉的直排表頭以求緊湊。欄位的順序為：

| 欄位       | 說明                                                                       |
|----------|--------------------------------------------------------------------------|
| Variable | 欄位名（等寬字型，過長時以省略號截斷）。帶邊框儲存格。                                              |
| Type     | 資料型別下拉選單：numeric、integer、character、factor、logical、Date、POSIXct。帶邊框下拉選單。  |
| Include  | 核取方塊——把該欄位作為預測變數納入模型。直排表頭。                                               |
| Factor   | 核取方塊——按分類因子處理（為每個水準產生一個單獨的係數）。直排表頭。                                      |
| Linear   | 核取方塊——強制使用線性關係（不用平滑項，只用 $\beta x$ ）。直排表頭。                                |
| Special  | 下拉選單（僅估價 / 市場模式）——參見下文的特殊欄位類型參考。帶邊框下拉選單。                                 |
| NAs      | 遺漏值的個數與百分比，帶顏色編碼：紅色（ $\geq 50\%$ ）、橙色（ $\geq 20\%$ ）、灰色（ $> 0$ ）。帶邊框儲存格。 |

因子、平滑與線性之別：預設情況下，被納入的數值變數會得到一個平滑項  $f(x)$ 。勾選 **Factor** 會產生一個分類項（每個水準一個係數）。勾選 **Linear** 會產生一個簡單線性項  $\beta x$  以取代平滑項。同時被標記為 Factor 與 Linear 的變數按 Factor 處理。

## 6.4 資料型別偵測與覆寫

mgcvUI 在匯入時會自動偵測資料型別，可識別數值、整數、邏輯、因子與日期欄位。你可以透過更改 **Type** 下拉選單來覆寫任何一項偵測結果。更改類型會影響該欄位在模型中的處理方式。

## 6.5 特殊欄位類型參考

在估價模式與市場模式下，**Special** 下拉選單提供以下選項：

加權：

- `weight` —— 觀測權重欄位（只允許一欄；權重為 0 的列不參與擬合）

日期與時間類型：

- `contract_date` —— 觸發根據估價基準日自動計算 `sale_age`
- `sale_age` —— 標示一個已預先計算好的銷售時距欄位（距基準日的天數），按原樣使用
- `listing_date` —— 在計算市場停留天數時作為備選來源
- `dom` —— 標示市場停留天數欄位

記錄類型：

- `sale_type` —— 標示用於區分估價標的與可比銷售案例的記錄類型欄位

貨幣類型：

- `concessions` —— 標示銷售讓價（賣方補貼、買方優惠等）

面積與區位類型：

- `latitude` —— 取值自動四捨五入至 3 位小數
- `longitude` —— 與緯度同樣的捨入處理
- `area` —— 市場區域或街區識別碼
- `living_area` —— 啟用按平方英尺計的殘差計算（`residual_sf` 與 `cqa_sf`）
- `lot_size` —— 宗地面積欄位
- `site_dimensions` —— 與地塊面積歸為一組

年齡類型：

- `actual_age` —— 不動產年齡欄位
- `effective_age` —— 不動產的有效年齡

顯示類型：

- `display_only` —— 該欄位會包含在 Excel 匯出中，但完全不參與模型擬合。可用於地址欄位、MLS 編號、宗地編號或其他參考性資料。

提示：被指定為 `display_only` 的欄位仍保留在資料集中並出現在 Excel 匯出裡，但會完全從預測變數表中排除。可用於 ID 欄位、地址或其他參考欄位。

## 7 參數選擇

側邊欄的第 5 節——**Mgcv Call Parameters** (mgcv 呼叫參數)——提供 GAM 的全部設定選項。每個參數都帶有一個藍色說明圖示 (?)，其提示框給出說明。

### 7.1 參數預設組

頂部的一個下拉選單提供兩種預設組，為常見工作流程設定合理的預設值：

| 預設                                           | 基底        | 設定                                    |
|----------------------------------------------|-----------|---------------------------------------|
| Standalone<br>(discovery) (獨立執行，探索)          | tp (薄板)   | k = 自動 (10)，gamma = 1.2，select = TRUE |
| Earth Pipeline<br>(refinement) (earth 流程，精化) | cr (三次迴歸) | k = 自動 (10)，gamma = 1.4，select = TRUE |

匯入 earthUI 節點時會自動選中 **Earth Pipeline** 預設組。earth 節點的整合必須使用三次迴歸樣條 (cr) 基底，因為只有它允許指定確切的節點位置。

### 7.2 Family (分配族)

為你的反應變數選擇分配族：

| Family                  | 適用情形                   |
|-------------------------|------------------------|
| <b>gaussian</b>         | 連續型反應變數 (例如銷售價格)。最為常用。 |
| <b>Gamma</b>            | 變異數與平均數成比例的正值連續型反應變數。  |
| <b>poisson</b>          | 計數資料 (例如成交筆數)。         |
| <b>binomial</b>         | 二元結果。                  |
| <b>inverse.gaussian</b> | 嚴重右偏的正值資料。             |

### 7.3 Method (方法)

平滑參數的估計方法：

| Method         | 說明                       |
|----------------|--------------------------|
| <b>REML</b>    | 受限最大概似 (預設，建議)。對過擬合最為穩健。 |
| <b>GCV.Cp</b>  | 廣義交叉驗證。傾向於選出略為複雜的模型。     |
| <b>ML</b>      | 最大概似。與 REML 相似，但可能欠平滑。   |
| <b>P-REML</b>  | Pearson REML 變體。         |
| <b>P-ML</b>    | Pearson ML 變體。           |
| <b>GACV.Cp</b> | 廣義近似交叉驗證。                |

### 7.4 Gamma (平滑度乘子)

gamma 參數會對平滑準則中的有效自由度作乘法放大，從而促成更平滑 (擺動更少) 的擬合。預設值為 1.2。

| 取值      | 效果                   |
|---------|----------------------|
| 1.0     | 標準平滑度                |
| 1.2–1.4 | 略比預設更平滑（建議）          |
| 2.0+    | 平滑得多——適用於雜訊較大的資料或小樣本 |

較大的  $\gamma$  值透過更重地懲罰複雜度來防範過擬合。Earth Pipeline 預設採用 1.4，以便在精化 earth 的節點時施加額外的平滑。

## 7.5 交叉驗證

勾選時（預設勾選），mgcvUI 會在擬合後執行 10 折交叉驗證（cross-validation）以計算 CV R 平方。這給出了樣本外預測能力的一個誠實估計。

## 7.6 Select（收縮）

勾選時（預設勾選），會為每個平滑項附加一個額外懲罰，可將其收縮至零。這就實現了自動變數選擇——不重要的平滑項實際上被移出模型。其實作方式是在 `gam()` 中設定 `select = TRUE`。

## 7.7 預設基底

平滑項所用的樣條基底函數（basis function）類型：

| 基底 | 全稱     | 說明                         |
|----|--------|----------------------------|
| tp | 薄板迴歸樣條 | 預設。各向同性（無需安放節點）。是很好的通用選擇。  |
| cr | 三次迴歸樣條 | 允許顯式安放節點。整合 earth 節點時必須使用。 |
| ps | P 樣條   | 節點等距的 B 樣條基底，配以差分懲罰。計算效率高。 |
| bs | B 樣條   | 彈性的 B 樣條基底。                |

## 7.8 預設 k（基底維度）

基底維度  $k$  控制每個平滑項擺動程度的上限。取值為 0（預設）表示「自動」——mgcvUI 會視情形取  $k = 10$  或 earth 節點的個數。

- 較小的  $k$ （3–5）：曲線非常平滑，接近直線。
- 中等的  $k$ （8–15）：彈性適中（通常已足夠）。
- 較大的  $k$ （20 以上）：彈性極高——若無強正則化則有過擬合之虞。

提示：Summary 頁籤中的有效自由度（EDF）告訴你每個平滑項實際用掉了多少自由度。若 EDF 接近  $k - 1$ ，說明基底可能太小——應加大  $k$ 。若 EDF 遠小於  $k$ ，說明該平滑項受到了良好的懲罰。

## 7.9 張量類型

對於連續變數之間的交互，有兩種張量積類型可用：

| 類型 | 函數         | 說明                                              |
|----|------------|-------------------------------------------------|
| ti | ti(x1, x2) | 張量交互——只對交互效應建模，主效應由單變數 $s()$ 項分別處理。就可解釋性而言更受建議。 |
| te | te(x1, x2) | 張量積——對包含主效應在內的整個聯合效應建模。在 RCA 調整中較難分解。           |

### 7.10 允許的交互

一個可折疊的 **Allowed Interactions**（允許的交互）小節會為所有被納入的平滑（非線性、非因子）預測變數顯示一個上三角核取方塊矩陣。每個核取方塊啟用相應變數對之間的張量積平滑。

- **Allow All**（全部允許） / **Clear All**（全部清除）——批次切換按鈕。
- 點選變數名——切換該變數的全部交互。
- 右鍵點選變數名——「Block from main effect」（從主效應中遮蔽）——把該變數標記為僅用於交互（不產生單變數平滑項）。

匯入 earthUI 節點時，earth 偵測到的交互會被預先勾選並鎖定（以黃色背景高亮）。若 earth 模型使用了  $\text{degree} = 1$ （無交互），則會顯示一條提示資訊。

#### 警告

啟用張量交互會大幅提高模型複雜度。請在你擁有領域知識、或有 earth 偵測到的證據支援時才使用交互項。務必核實所得調整額是合理且可以解釋的。

### 7.11 因子-平滑交互

當因子變數與平滑變數同時被納入時，會出現一個單獨的 **Factor-by-Smooth Interactions**（因子-平滑交互）矩陣。每個核取方塊會產生一個  $s(x, \text{by} = \text{factor})$  項——即為每個因子水準產生一條單獨的平滑曲線。

### 7.12 進階參數

一個可折疊的「Advanced」（進階）小節展示若干附加設定：

- **Optimizer**（最佳化器）——平滑參數估計所用的演算法：`outer/newton`（預設）、`outer/bfgs` 或 `efs`（擴展 Fellner-Schall 法）。
- **Scale**（尺度參數）——尺度參數。0（預設）表示由資料估計。對於尺度已知的模型可設為一個確定值。
- **Discrete covariate method**（離散化協變數方法）——針對大型資料集的快速離散化擬合（`discrete = TRUE`）。
- **Threads**（執行緒數）——並行執行緒的數目（1-32）。僅在 `discrete = TRUE` 時有效。

## 8 擬合模型

### 8.1 Fit 按鈕

側邊欄第 6 節中有 **Fit MgcV GAM Model** (擬合 mgcv GAM 模型) 按鈕。點選它即以你目前的設定執行模型。

### 8.2 擬合過程中發生了什麼

當你點選 Fit 時，mgcvUI 會：

1. 校驗設定 —— 檢查是否至少有一個預測變數、以及至少 10 個完整列。若完整列不足 50%，會給出警告（並列出造成大部分 NA 的那些欄位）。
2. 準備資料 —— 只保留與模型相關的欄位，刪除含 NA 的列，應用權重（若已指定），排除標的列（在估價模式下），並施加反應變數變換。
3. 建構 GAM 公式 —— 為數值預測變數構造平滑項  $s(x, bs="cr", k=10)$ ，為分類變數構造因子項，為被標記為 Linear 的變數構造線性項，並為所選交互構造張量積。earth 的節點會作為顯式的節點位置嵌入到 cr 基底的項中。
4. 處理邊界情形 —— 剔除常數變數，把二值預測變數轉換為因子，把  $k$  限制在不同取值個數減 1 以內，並在需要更多節點時用資料分位數補充 earth 的節點。
5. 擬合 GAM —— 以構造好的公式、family、method、gamma、select 及其他參數呼叫 `mgcv::gam()`。
6. 執行交叉驗證（若已啟用） —— 10 折交叉驗證，計算樣本外 R 平方。
7. 更新所有結果頁籤 —— Summary、Equation、Contribution 圖、Diagnostics 以及其他所有頁籤都會被填充。

### 8.3 背景擬合與進度顯示

當 callr 套件可用時，模型擬合會在一個背景行程中執行，以保持應用程式的回應性。一個暗色主題的終端式進度浮層會顯示：

- 計時器，顯示已耗費的秒數 / 分鐘數。
- 按顏色編碼的追蹤日誌：藍色表示狀態，黃色表示交叉驗證折的進度，紅色表示錯誤，綠色表示結果。
- 一個 **Abort**（中止）按鈕，可在需要時終止擬合行程。
- 一個 **Close**（關閉）按鈕，在擬合完成後出現。

若 callr 不可用，擬合將同步執行（應用程式會凍結，直到擬合完成）。

擬合成功後，Fit 按鈕上會出現一個白色勾號，同時狀態列顯示：「R-sq = X, CV R-sq = X, Dev = X%, AIC = X, n = X.」

## 9 結果頁籤

### 9.1 Data (資料)

以互動式 DataTable 顯示已匯入的資料，支援橫向捲動，每頁 15 列。在估價模式與市場模式下，該頁籤分為一張 **Subject Property**（估價標的，第 1 列）表和一張 **Comparable Sales**（可比銷售案例，其餘各列）表；在 General 模式下只顯示一張表。

每個儲存格都只顯示一行文字，並按欄寬截斷，這樣很寬的自由文字欄位（例如不動產備註）就不會把整列拉得很長。點選可選取其中一列；雙擊任一儲存格可彈出視窗，顯示其完整、未經截斷的內容。

### 9.2 Equation (方程)

分兩部分顯示已擬合的 GAM 模型：

模型方程（由 MathJax 渲染）：

- 左端顯示反應變數及其相應的變換（log、log10）與連結函數。
- 右端顯示截距值、以  $f_i$ (變數) 形式給出的平滑項、按變數分組並標明水準數的因子項，以及帶係數值的線性項。
- family 與 method 顯示在方程上方。

**Smooth Function Definitions**（平滑函數定義）表：

- 項的公式字串（例如 `s(living_sqft, bs="cr", k=8)`）
- 變數名
- 類型（s、te、ti 或 linear）
- 基底函數類型
- $k$  值
- earth 節點位置（若已匯入）

### 9.3 Summary (摘要)

頂部橫向排列六張指標卡片：

- **R-squared** —— 已解釋離差的比例
- **CV R-squared** —— 交叉驗證 R 平方（若已啟用交叉驗證）
- **Dev. Explained %** —— 已解釋離差的百分比
- **AIC** —— 赤池資訊準則
- **n** —— 所用觀測的個數
- **Smooths** —— 平滑項的個數

卡片下方：

- **Smooth Terms** 表（平滑項）—— 項、EDF（有效自由度）、Ref.df、F 統計量、p 值
- **Parametric Terms** 表（參數項）—— 項、估計值、標準誤、t 值、p 值

### 9.4 Variable Importance (變數重要性)

一張水平長條圖，顯示各模型項的相對重要性：

- 平滑項：按 F 統計量排序（藍色條）
- 參數項：按  $|t|$  值排序（綠色條）

圖下方的 DataTable 列出項、類型、EDF、統計量與 p 值，按統計量降序排列。

### 9.5 貢獻圖

每一個對預測值有貢獻的模型項都有一張互動式 plotly 貢獻圖，顯示在一張帶邊框的卡片中。該頁籤會為所有類型的項給出圖形：

### 9.5.1 單變數平滑項 — $s(x)$

- 藍色曲線，帶 95% 信賴帶
- 灰色散點，顯示偏殘差（用於未作對數變換的模型）
- 紅色垂直虛線，位於 earth 節點處（若已匯入 earth 結果）
- 懸停提示：變數取值、貢獻（\$）、95% 信賴區間、變化率（每單位的斜率）

對於經對數變換的模型，縱軸會用公式  $\bar{y} \times (e^{f(x)} - 1)$  反變換為以美元計的貢獻，使曲線可以直接按美元數額來解讀。

對於緯度／經度變數，斜率按每 0.001 度而非每 1 度給出，因為前者是更有意義的地理增量。

### 9.5.2 因子—平滑項 — $s(x, \text{by} = \text{factor})$

每個因子水準一條彩色曲線，疊繪在同一張圖上。這有助於看清某個預測變數的效應如何隨組別（例如不同街區）而變化。

### 9.5.3 交互項 — $ti(x_1, x_2)$ 與 $te(x_1, x_2)$

用熱圖顯示二維貢獻曲面。色標為：紅色（負）→ 白色（零）→ 藍色（正）。懸停時同時顯示兩個變數的取值與貢獻數額。

### 9.5.4 參數項 — 線性項與因子項

- 數值型線性項：資料點的散點圖，疊加擬合直線
- 因子項：長條圖，顯示每個因子水準的平均貢獻

## 9.6 Correlation（相關性）

一張熱圖矩陣，顯示所有數值型預測變數之間的 Pearson 相關係數。資料匯入後即可檢視——無需擬合模型。

- 顏色漸變：藍色（-1）→ 白色（0）→ 紅色（+1）
- 每個儲存格中標出相關係數值
- 有助於在擬合之前識別共曲性（concurvity，多重共線性在平滑項中的對應現象）

## 9.7 Diagnostics（診斷）

兩組診斷視覺化：

診斷面板（透過 `gratia::appraise` 產生）：四張圖——殘差對擬合值、Q-Q 圖、殘差直方圖，以及反應值對擬合值。

實際值對預測值散點圖：觀測值對預測值，附一條 45 度虛線參考線。對於經對數變換的模型，取值會被反變換回原始尺度。

## 9.8 RCA Adjustments（RCA 調整）

RCA 計算完成後顯示三張直方圖（僅限估價 / 市場模式）：

- **Residual Adjustment %**（殘差調整百分比）——殘差調整額佔銷售價格百分比的分配
- **Net Adjustment %**（淨調整百分比）——淨調整額的分配
- **Gross Adjustment %**（總調整百分比）——總調整額的分配

每張直方圖的副標題中給出平均數、中位數與標準差，並以虛線標出平均數與中位數的參考位置。

### 9.9 Anova (變異數分析)

一張合併的變異數分析表，把 GAM 摘要中的參數項與平滑項匯合在一起。顯示項、類型（參數項或平滑項）以及所有相關統計量。

### 9.10 Mgcvc Output (mgcv 輸出)

原始文字輸出，內容包括：

- 擬合計時
- GAM 公式
- 模型的 print 輸出
- 完整的 `summary(model)` 輸出
- family 與連結函數
- 整體共曲性度量

### 9.11 Sign Check (符號檢查)

當已匯入 earthUI 節點時，該頁籤會把每個變數在 earth 模型中的方向與其在 GAM 中的方向作比較：

- **earth\_direction**：遞增、遞減或混合（依據鉸鏈函數的符號判定）
- **gam\_direction**：遞增、遞減或混合（由平滑項在 100 點網格上的導數估計得出）
- **consistent**：TRUE/FALSE
- **warning**：對任何不一致之處的說明

狀態列會顯示「All directions consistent」（所有方向一致）或存在不一致的變數個數。

### 9.12 Concurvity (共曲性)

共曲性是多重共線性在平滑項中的對應現象。該頁籤顯示：

- 整體共曲性表：各列分別對應 worst / observed / estimate 三種度量，各欄對應每個平滑項。帶顏色編碼：白色 ( $< 0.5$ )、黃色 ( $0.5-0.8$ )、紅色 ( $> 0.8$ )。
- 兩兩共曲性表：最壞情形的兩兩矩陣，採用相同的顏色編碼。

「worst」一列中高於約 0.8 的取值表明，該平滑項可以被其他平滑項很好地逼近，提示存在冗餘。

### 9.13 Basis Check (基底檢驗)

`mgcv::gam.check()` 的輸出：針對每個平滑項的基底維度充分性檢驗。若某個平滑項的  $k$  過小，該檢驗會以一個較小的  $p$  值將其標示出來，提示你應當加大  $k$ 。

## 10 下載資料

擬合完成後，**Download Output (Excel)**（下載輸出，Excel）按鈕（側邊欄第 7 節）會把含預測值與診斷資訊的 Excel 檔案匯出到作用中專案的 mgcv 輸出資料夾。

### 10.1 輸出欄位

| 欄位                 | 說明                                                 |
|--------------------|----------------------------------------------------|
| basis              | 反變換後的截距值（所有列相同）                                    |
| <var>_contribution | 各平滑項對預測值的貢獻。對於對數模型，會按比例分配為美元數額。                    |
| est_<target>       | 模型預測值（例如 est_sale_price），若適用則已從對數尺度反變換             |
| residual           | 實際值 – 預測值                                          |
| calc_residual      | 校驗用：實際值 – 各項貢獻之和（反變換後）                             |
| cqa                | 狀況–品質–吸引力（Condition-Quality-Appeal，CQA）得分（0–10 分制） |
| residual_sf        | 殘差 / 居住面積（若已指定 living_area）                        |
| cqa_sf             | 由 residual_sf 排序計算得到的 CQA                          |

在估價模式下，第 1 列（估價標的）的 residual/cqa/cqa\_sf 均置為 NA。當指定了 living\_area 欄位時，各列按 residual\_sf 降序排列。

成功完成後，下載按鈕上會出現一個白色勾號。

### 10.2 CQA 得分

CQA 把每一列的殘差與其餘所有列相比較，按 0–10 分制排序評分：

- 高 CQA（約 9–10）：成交價遠高於預測值——很可能狀況 / 品質 / 吸引力較優
- 低 CQA（約 0–1）：成交價遠低於預測值——很可能狀況較差或屬於困境出售
- CQA 約為 5：接近中位數

## 11 RCA 計算與下載

RCA (Residual Constraint Approach, 殘差約束法) 工作流程在 **Appraisal** (估價) 與 **Market Area Analysis** (市場區域分析) 模式下、模型擬合之後可用。估價標的必須位於第 1 列。

### 11.1 開啟 RCA 對話方塊

點選側邊欄第 8 節中的 **Calculate RCA Adjustments & Download** 按鈕，會彈出一個模式對話方塊，其中包含：

- **Score type** (得分類型)：CQA 或 CQA per SF (若已指定 living\_area)。按平方英尺計的選項會用居住面積對估價標的的殘差作縮放。
- **Subject CQA Score** (估價標的 CQA 得分)：接受 0.00 至 10.00 之間的取值。該欄位初始為空；若已匯入過 earthUI 模型，你在 earthUI 中使用的得分類型與取值會從共用的專案設定中沿用過來。

點選 **Generate** (產生) 即可計算並下載。Excel 檔案會寫入作用中專案的 mgcv 輸出資料夾 (<os>\_out\_mgcv)。

### 11.2 CQA 得分內插

1. 先對各可比案例的 CQA 得分與殘差排序
2. 用線性內插 (approx()) 把你給定的 CQA 值映射為一個殘差
3. 估價標的價值 = 模型預測值 + 內插所得殘差
4. 權重為 0 的列按與估價標的具有相同殘差來處理

### 11.3 輸出欄位

| 欄位                  | 說明                     |
|---------------------|------------------------|
| subject_value       | 模型預測值 + 內插所得殘差         |
| subject_cqa         | 使用者輸入的 CQA 得分          |
| <var>_adjustment    | 估價標的貢獻 - 可比案例貢獻 (以美元計) |
| residual_adjustment | 估價標的殘差 - 可比案例殘差        |
| net_adjustments     | 全部調整額之和                |
| gross_adjustments   | 全部調整額絕對值之和             |
| residual_adj_pct    | 殘差調整額佔銷售價格的百分比         |
| net_adj_pct         | 淨調整額佔銷售價格的百分比          |
| gross_adj_pct       | 總調整額佔銷售價格的百分比          |
| adjusted_sale_price | 實際銷售價格 + 淨調整額          |

計算成功完成後，按鈕上會出現一個白色勾號。

## 12 銷售比較網格

銷售比較網格 (Sales Comparison Grid) 僅在 **Appraisal** (估價) 模式下、且已計算 RCA 調整額之後可用。它會產生一份格式化的 Excel 活頁簿，適合納入估價報告之中。

### 12.1 可比案例選擇對話方塊

點選 **Generate Sales Grid & Download** 按鈕會開啟一個模態對話方塊，列出所有可比銷售案例。可比案例分為兩組：

- **Recommended comps** (建議的可比案例，已預先勾選)：總調整額 < 銷售價格的 25%，按銷售時距排序 (最近的在前)。最多 30 個。
- **Other comps** (其他可比案例，預設不勾選)：其餘的可比案例，按總調整百分比排序。

選定你的可比案例後，點選 **Generate Sales Grid** 即可建立活頁簿。

必需的特殊類型：必須已指定 `contract_date` 與 `living_area`。建議的特殊類型：`latitude`、`longitude` 與 `lot_size`——沒有它們網格也能產生，但某些欄位會留空。

### 12.2 活頁簿版式

產生的活頁簿含多個工作表，每張表並排容納 3 個可比案例：

- 表頭列：估價標的與可比案例的地址、銷售價格、成交日期以及市場停留天數
- 分組列：區位 (緯度、經度、區域)、宗地面積 (地塊面積、宗地尺寸) 與年齡 (實際年齡、有效年齡)，帶組標題與子條目
- 模型變數列：每個平滑項一列，顯示估價標的貢獻、各可比案例貢獻以及調整額
- 殘差特徵列：空白且未鎖定的儲存格，供估價師填寫模型未能刻畫的特徵 (例如狀況、品質、景觀)
- 彙總列：淨調整百分比、總調整百分比與調整後銷售價格，均帶有可即時計算的 Excel 公式

### 12.3 工作表保護與殘差儲存格

每張工作表都受到保護以防誤改，但殘差特徵的取值儲存格被顯式解鎖。這既讓估價師能夠為模型之外的特徵輸入取值，又保留了由公式驅動的調整額計算。

## 13 下載報告

報告產生採用兩步式 Quarto 工作流程，取代了早期版本中的一步式報告。

### 13.1 第 1 步：產生 Quarto 報告

**Generate Quarto Report**（產生 Quarto 報告）按鈕（模型擬合後可用）會把一個自包含的 .qmd 包寫入作用中專案的 mgcv 輸出資料夾。該包含有 Quarto 原始檔、預先渲染好的圖形檔案（PNG 與 PDF）、一個裝載了全部表格與模型元資料的 report\_data.rds，以及用於 Word 輸出的 reference.docx 樣式模板。由於該包是自包含的，你可以在渲染之前編輯 .qmd 原始檔，或把它與其他 Quarto 文件組合起來。

圖形在產生該包時即已預先渲染；若某張圖渲染失敗，程式會寫入一張佔位圖，包中其餘部分照常產生。

### 13.2 第 2 步：轉換 Quarto 報告

**Convert Quarto Report**（轉換 Quarto 報告）小節透過 quarto::quarto\_render() 把一個 .qmd 檔案渲染為一種或多種輸出格式。它預設指向最近產生的那個包，但 **Browse**（瀏覽）按鈕允許你選擇任意 .qmd 檔案。可用格式有三種：

- **HTML** —— 自包含的 HTML 文件。在所有平台上均可使用，無需額外軟件。建議作為預設選項。
- **Word (.docx)** —— 使用隨包提供的 reference.docx 樣式模板渲染。適合編輯與分發。
- **PDF** —— 需要安裝 LaTeX（參見下文的系統需求）。若未偵測到 LaTeX，PDF 選項會自動從格式核取方塊中隱去。

早期的一步式渲染器 render\_gam\_report() 與 export\_gam\_docx() 仍保留，可供程式化呼叫。

### 13.3 報告內容

產生的報告包括：

- 資料集與模型：資料檔案、樣本數、family、method 以及擬合資訊
- 結果摘要： $R^2$ 、 $CV R^2$ 、已解釋離差、AIC 及相關的擬合指標
- 模型方程：模型方程，其中平滑項用  $f_i$  記號表示、線性項給出顯式係數，並附平滑函數定義。較長的方程在 PDF 中會跨多行折排。
- 摘要表：平滑項（EDF、F、p 值）與參數項（估計值、標準誤、t、p 值）
- 已啟用的交互：模型中所含張量交互項與因子—平滑交互項的矩陣（僅在存在交互項時顯示）
- 變數重要性：按 F 統計量（平滑項）或  $|t|$ （參數項）排序的長條圖
- 平滑效應（所有類型的項）：
  - 單變數平滑圖（互動式圖形的 ggplot2 靜態版本）
  - 因子—平滑的多線條圖
  - 交互熱圖（ $t_i$  與  $t_e$  項）
  - 參數項圖（數值型用散點圖，因子型用長條圖）
- 預測變數相關性熱圖
- 診斷：殘差診斷以及實際值對預測值的散點圖
- mgcv 摘要：已擬合模型的完整文字輸出
- 記號附錄：一個不編號的結尾小節，說明報告中所用的價值貢獻函數記號

成功產生後，Generate Quarto Report 按鈕上會出現一個白色勾號。

## 14 匯出模型函數

mgcvUI 可以把已擬合 GAM 的平滑函數匯出為多種語言的獨立程式碼。每個平滑項都在一個 200 點網格上求值，並以帶線性內插的查表形式匯出，從而使該模型可在 R 之外使用，且不依賴 mgcv。

在目前版本中，函數匯出是透過程式化方式進行的。匯出用的輔助函數接受一個已擬合的結果物件（由 `fit_gam()` 產生），並寫出下述檔案：

- `generate_r_code()`、`generate_python_code()` 與 `generate_cpp_code()` ——各語言的原始檔
- `export_functions_sqlite()` ——SQLite 查表資料庫
- `export_functions_zip()` ——把這四種格式的任意組合打包成一個 zip 壓縮包

### 14.1 R 函數

檔案：`gam_functions.R`

- 頭部註解，含反應變數、 $R^2$ 、CV  $R^2$  與模型公式。
- 一個 `INTERCEPT` 常數。
- 每個平滑變數對應一個 `approxfun()`，帶預先算好的 `x/y` 向量。
- 線性項係數以具名常數形式給出。
- 用法：把截距、各個 `g_var(value)` 以及各個 `coef_var * value` 相加。

### 14.2 Python 函數

檔案：`gam_functions.py`

- `numpy` 匯入語句。
- `INTERCEPT` 常數。
- 每個變數對應：`_x_var` 與 `_fx_var` 陣列，以及使用 `np.interp()` 的 `g_var(x)` 函數。
- 線性項係數以 `COEF_*` 常數形式給出。

### 14.3 C++ 標頭檔

檔案：`gam_functions.hpp`

- 僅含標頭檔，使用 `#pragma once`。
- `gam_functions` 命名空間。
- 使用 `std::lower_bound` 的 `interp()` 輔助函數。
- 每個變數對應：一個帶 `static const` 陣列的內聯函數。
- 線性項係數以 `constexpr double` 形式給出。

### 14.4 SQLite 資料庫

檔案：`gam_functions.sqlite`

- 表 `smooth_grids`：欄位為 (`variable`, `x`, `fx`) ——即 200 點查表。
- 表 `model_info`：`response`、`r_squared`、`cv_rsqr`、`dev_explained`、`aic`、`n_obs`、`family`、`method`、`formula`。
- 表 `linear_terms`：`variable`、`coefficient`。
- 表 `intercept`：截距值。

## 15 與 earthUI 和 glmnetUI 的比較

earth 應用程式及其後處理模組構成一套用於迴歸建模的工具集。它們共用相同的資料格式、特殊欄位類型、RCA 工作流程與示範資料集，但使用不同的建模引擎。

### 15.1 主要差異

| 特性       | mgcvUI              | earthUI             | glmnetUI      |
|----------|---------------------|---------------------|---------------|
| 方法       | GAM (mgcv)          | MARS/Earth (earth)  | 彈性網絡 (glmnet) |
| 關係形態     | 平滑非線性               | 分段線性                | 線性            |
| 變數選擇     | select=TRUE 收縮      | 逐步修剪                | Lasso 懲罰      |
| 交互項      | 張量積 (te/ti)         | 鉸鏈函數乘積 (2-3 階)      | 兩兩乘積          |
| 偏效應      | 帶信賴區間的平滑曲線          | g 函數 (分段)           | 線性斜率          |
| earth 流程 | 從 earth 匯入節點        | 節點的來源               | 匯入鉸鏈函數基底      |
| 程式碼匯出    | R、Python、C++、SQLite | R、Python、C++、SQLite | —             |
| 報告格式     | Word、PDF、HTML       | Word、PDF、HTML       | Word、PDF、HTML |

有一項 earthUI 獨有的差異值得一提：earthUI 含有一個實驗性的 Prolog/DCG 功能，可從自由文字的 MLS 備註中提取結構化的不動產屬性。

### 15.2 共有功能

三個工具都提供：

- 共用的 regProj 專案樹（預設為 ~/regProj）——專案、輸入檔案以及按方法分列的輸出在同族應用之間共用
- CSV 與 Excel 匯入，並作 snake\_case 欄位名轉換
- 支援 30 多個國家的區域設定，涵蓋 CSV 分隔符、小數點符號與紙張尺寸
- 特殊欄位指定 (contract\_date、living\_area、area、latitude、longitude 等)
- 依據估價基準日自動計算 sale\_age
- 深色 / 淺色模式切換 (Nord 主題)
- 設定的持久化 (SQLite 與 / 或 localStorage)
- 相關性熱圖、變數重要性、診斷
- CQA 評分與 RCA 調整工作流程
- 帶可比案例選擇與格式化 Excel 輸出的銷售比較網格
- 相容同一批示範資料集 (Appraisal\_1.csv)

### 15.3 建議的工作流程

- 先用 earthUI 發現重要變數、非線性關係與交互作用。earth 的自動建模提供了最佳的起點。
- 再用 mgcvUI 精化，如果你希望得到更平滑的偏效應。匯入 earthUI 的節點即可無縫過渡。
- 使用 glmnetUI，當你需要一個帶激進變數選擇或係數符號約束的純線性模型時。
- 在同一份資料上比較三者，以評估非線性模型所增添的複雜度是否值得。

## 16 示範資料集：Appraisal\_1.csv

### 16.1 說明

該示範資料集與 earthUI 和 glmnetUI 共用。它位於 mgcvUI 原始碼的 demo\_mls/ 資料夾中；若已安裝 earthUI，則可用：

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

該檔案包含約 1,500 條住宅銷售記錄，取自一份模擬的 MLS 匯出，標的不動產位於第 1 列，可比銷售案例位於其餘各列。資料代表一個多區域市場中的獨棟住宅銷售，不動產面積、年齡與區位各不相同。

這並非真實資料，但以北加州一個實際存在的街區為藍本。所有識別資訊均已改寫或刪除。

### 16.2 各欄位

| 欄位                  | 類型        | 特殊類型          | 說明                                   |
|---------------------|-----------|---------------|--------------------------------------|
| sale_type           | character | sale_type     | 用於區分估價標的與可比銷售案例的記錄類型碼                |
| weight              | numeric   | weight        | 觀測權重 (0 = 不參與擬合)                     |
| id                  | numeric   | display_only  | 內部記錄 ID                              |
| property_id         | numeric   | display_only  | MLS 不動產識別碼                           |
| listing_id          | character | display_only  | MLS 掛牌編號                             |
| parcel_number       | character | display_only  | 縣估價官宗地編號 (APN)                       |
| street_address      | character | display_only  | 不動產地址                                |
| city_name           | character | display_only  | 市                                    |
| postal_code         | character | display_only  | 郵遞區號                                 |
| county_name         | character | display_only  | 縣                                    |
| contract_date       | Date      | contract_date | 銷售合約日期 (用於計算 sale_age)               |
| sale_age            | numeric   | —             | 從合約日期至估價基準日的天數 (已預先計算)               |
| coe_date            | Date      | display_only  | 交割日期                                 |
| listing_status      | character | display_only  | 掛牌狀態 (例如「Sold」, 已售)                  |
| sale_price          | numeric   | (目標變數)        | 銷售價格——反應變數                           |
| rent                | numeric   | —             | 月租金 (用於多目標模型)                        |
| list_price          | numeric   | display_only  | 掛牌價格                                 |
| original_list_price | numeric   | display_only  | 初始掛牌價格                               |
| living_sqft         | numeric   | living_area   | 以平方英尺計的總居住面積                         |
| ppsf                | numeric   | display_only  | 每平方英尺單價                              |
| beds_total          | integer   | —             | 臥室數                                  |
| baths_total         | numeric   | —             | 衛浴總數 (例如 2.5 = 2 個全衛 + 1 個半衛)        |
| lot_size            | numeric   | lot_size      | 以平方英尺計的地塊面積                          |
| area_id             | integer   | area          | MLS 區域識別碼                            |
| area_text           | character | display_only  | 區域名稱                                 |
| age                 | numeric   | actual_age    | 以年計的不動產年齡                            |
| year_built          | integer   | display_only  | 建成年份                                 |
| latitude            | numeric   | latitude      | 緯度 (入模前四捨五入至 3 位小數)                  |
| longitude           | numeric   | longitude     | 經度 (入模前四捨五入至 3 位小數)                  |
| latitude6           | numeric   | display_only  | 全精度緯度                                |
| longitude6          | numeric   | display_only  | 全精度經度                                |
| garage_spaces       | integer   | —             | 車庫車位數                                |
| fp_count            | integer   | —             | 壁爐數                                  |
| no_of_stories       | numeric   | —             | 樓層數                                  |
| style               | character | —             | 建築風格                                 |
| view                | character | —             | 景觀類型 (例如「Neighborhood」街區景、「Hills」山景) |
| days_on_market      | integer   | dom           | 市場停留天數                               |
| listing_date        | Date      | listing_date  | 掛牌日期                                 |

| 欄位               | 類型      | 特殊類型        | 說明   |
|------------------|---------|-------------|------|
| sale_concessions | numeric | concessions | 賣方讓價 |

### 16.3 建議的快速上手步驟

1. 啟動 mgcv 例程：`earthUI::launch_mgcv()`
2. 在第 1 節中新建一個用途為 **For Appraisal** 的專案
3. 把 `Appraisal_1.csv` 放入該專案的輸入資料夾，並在第 2 節中選中它（必要時點選 **Refresh**）
4. 選擇 `sale_price` 作為目標變數
5. 如需比例式調整，可將 Response Transform 設為 **Log (natural)**
6. 按上表所示指派特殊類型
7. 納入以下預測變數：`sale_age`、`living_sqft`、`baths_total`、`lot_size`、`area_id`（作為因子）、`age`、`latitude`、`longitude`、`garage_spaces`
8. 保留預設設定：`REML`、`gamma = 1.2`、`tp` 基底、`k = 自動`、`select = TRUE`
9. 點選 **Fit Mgcvm GAM Model**
10. 檢視 Summary、Contribution（平滑圖）與 Diagnostics 頁籤
11. 下載輸出（第 7 步），檢視 CQA 排名
12. 以約 ~5.00 的 CQA 得分計算 RCA 調整額（第 8 步）
13. 用建議的可比案例產生銷售比較網格（第 9 步）

## 17 系統需求與疑難排解

### 17.1 受支援的平台

mgcvUI 可在 **macOS**、**Windows** 與 **Linux** 上執行。本應用程式在 macOS 配合 RStudio 的環境中開發，也主要在該環境中測試。與平台相關的說明見下文。

### 17.2 R 版本

要求  $R \geq 4.1.0$ 。強烈建議使用 RStudio Desktop (2023.06 或更高版本) ——它自帶 **Quarto** 與 **pandoc** (渲染報告所需)，並為啟動本應用程式提供了便利的環境。

### 17.3 必需的 R 套件

以下套件會作為相依套件自動安裝：

| 套件                | 用途                          |
|-------------------|-----------------------------|
| shiny             | Web 應用程式框架                  |
| mgcv              | GAM 模型擬合                    |
| valengrCore       | 共用的估價核心 (特殊角色、earthUI 設定沿用) |
| ggplot2, plotly   | 靜態與互動式貢獻圖                   |
| gratia            | 平滑項提取與診斷                    |
| DT                | 互動式資料表                      |
| quarto, rmarkdown | Quarto 報告的產生與渲染             |
| readxl            | Excel 檔案匯入                  |
| shinyFiles        | 檔案與資料夾瀏覽器                   |
| RSQLite, DBI      | 設定與專案資料庫 (SQLite)           |
| jsonlite          | 變數狀態的序列化                    |

### 17.4 選用相依套件

| 元件                                     | 何時需要                                                                                          |
|----------------------------------------|-----------------------------------------------------------------------------------------------|
| <b>LaTeX</b> (TinyTeX、MiKTeX 或 MacTeX) | 僅用於產生 PDF 報告。若未偵測到 LaTeX 安裝，PDF 選項會自動從格式下拉選單中隱去。安裝指令： <code>tinytex::install_tinytex()</code> |
| <b>sysfonts + showtext</b>             | 供 ggplot2 圖形使用的 Roboto Condensed 字型。若不可用或處於離線狀態，應用程式會自動退回到系統無襯線字型。                            |
| <b>earth</b> ( $\geq 5.3.0$ )          | 僅在使用 earthUI 匯入管道時需要。獨立使用時不需要。                                                                |
| <b>callr</b>                           | 背景模型擬合 (保持應用程式的回應性)。沒有它則擬合約步執行。                                                               |
| <b>writexl, openxlsx</b>               | Excel 下載 (模型輸出、RCA、銷售網格)。遺漏時應用程式會提示。                                                          |
| <b>bslib</b>                           | Bootstrap 5 主題 (Nord Light/Dark)。                                                             |
| <b>officer</b>                         | 早期的程式化 Word 匯出 ( <code>export_gam_docx()</code> )。                                            |

### 17.5 與平台相關的說明

#### 17.5.1 macOS

問題最少。系統庫建議使用 Homebrew 安裝。若需要 PDF 報告：

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

### 17.5.2 Windows

配合 RStudio Desktop 執行良好。要點：

- 企業／受限機器：若暫存目錄或使用者資料目錄受到限制，設定的持久化與報告產生可能失敗。應用程式會在 R 控制台中顯示警告，但仍繼續執行，只是不儲存設定。
- **LaTeX**：從 R 中安裝 TinyTeX（指令與上文 macOS 相同），或從 [miktex.org](http://miktex.org) 安裝 MiKTeX。
- 過長的檔案路徑：Windows 有 260 個字元的路徑長度上限。請保持工作目錄路徑簡短。

### 17.5.3 Linux

各發行版之間差異最大。Ubuntu/Debian 使用者在編譯 R 套件之前可能需要先裝好系統庫：

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

若需 PDF 報告：使用 `tinytex::install_tinytex()`，或從系統套件管理器安裝 `texlive-full`。

對於無顯示設備的無頭伺服器，`plotly` 互動式圖形在報告中會以靜態圖形作為退回方案渲染。`Shiny` 應用程式本身仍需要一個網頁瀏覽器連接。

## 17.6 優雅降級

`mgcv`UI 的設計使其即便在缺少選用元件時也能工作：

| 遺漏的元件                            | 行為                                                        |
|----------------------------------|-----------------------------------------------------------|
| 無 LaTeX                          | PDF 選項從報告格式下拉選單中隱去。HTML 與 Word 仍然可用。                      |
| 無網絡 / 字型載入失敗                     | Roboto Condensed 字型由系統無襯線字型替代。控制台記錄一條提示資訊。                |
| 無 RSQLite / 唯讀檔案系統               | 設定不會跨會話儲存。控制台顯示警告。除此之外應用程式正常執行。                           |
| <code>gratia::appraise</code> 失敗 | 診斷頁籤與報告中顯示佔位資訊。可在控制台中改用 <code>mgcv::gam.check()</code> 。  |
| 暫存目錄不可寫                          | 報告產生失敗，並給出明確的錯誤訊息。可透過把 <code>TMPDIR</code> 環境變數設為可寫位置來解決。 |

## 17.7 疑難排解

### 17.7.1 「PDF option is not available」(PDF 選項不可用)

未偵測到 LaTeX 安裝。請在 R 中執行：

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

然後重啟應用程式，PDF 選項便會出現。

### 17.7.2 「Settings will not persist」(設定將不會儲存)

SQLite 資料庫目錄無法建立或無法寫入。請檢查以下位置的權限：

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

### 17.7.3 「Diagnostics plot unavailable」(診斷圖不可用)

`gratia::appraise()` 函數在複雜模型(含許多張量交互項、非標準分配族)上偶爾會失敗。這是一個已知的上游問題。變通辦法：在 R 控制台中執行 `mgcv::gam.check(model)` 以直接檢視診斷圖。

### 17.7.4 「Report error: number of items to replace...」(報告錯誤：待替換的項數……)

這是早期版本中一個與 `gratia` 套件相關的已知問題。請確保你安裝的是最新版本的 `mgcvUI`。目前版本會捕獲該錯誤，並在產生報告時為診斷部分放置一個佔位內容。

### 17.7.5 7880 連接埠已被佔用

若應用程式因 7880 連接埠被佔用(來自上一次會話)而無法啟動，請執行：

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

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

### 17.7.6 在 Linux 上安裝套件失敗

缺少系統庫。請安裝上文 Linux 一節中列出的開發標頭檔，然後重新執行 `install.packages()`。

## 參考文獻

- Wood, S. N. (2003). "Thin-plate regression splines". In: *Journal of the Royal Statistical Society (B)* 65.1, pp. 95–114. DOI: [10.1111/1467-9868.00374](https://doi.org/10.1111/1467-9868.00374).
- (2004). "Stable and efficient multiple smoothing parameter estimation for generalized additive models". In: *Journal of the American Statistical Association* 99.467, pp. 673–686. DOI: [10.1198/016214504000000980](https://doi.org/10.1198/016214504000000980).
- (2011). "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models". In: *Journal of the Royal Statistical Society (B)* 73.1, pp. 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*. 2nd ed. Chapman and Hall/CRC.
- Wood, S. N., N. Pya, and B. Säfken (2016). "Smoothing parameter and model selection for general smooth models (with discussion)". In: *Journal of the American Statistical Association* 111, pp. 1548–1575. DOI: [10.1080/01621459.2016.1180986](https://doi.org/10.1080/01621459.2016.1180986).