

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

© 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。 earthUI 是 R 語言 earth 套件的圖形使用者介面，該套件實作了多元自適應迴歸樣條 (MARS, multivariate adaptive regression splines)。它提供三種用途模式——通用預測建模、房地產估價和市場區域分析——並引導使用者完成資料匯入、模型配置、模型擬合，以及對 earth 診斷資訊和圖形輸出的解讀。本文記錄了 earthUI 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： MARS，多元自適應迴歸樣條，earth，R，圖形使用者介面，房地產估價

出版者按 (2026 年 7 月)：本文描述的是 earthUI 0.10.0。付印前不久，earthUI 0.11.0 吸收了配套應用 glmnetUI 和 mgcvUI (見隨後兩篇文章)，將 MARS、彈性網絡和 GAM 工作流程整合到單一應用程式中。本文所述的介面與工作流程不受此次整合的影響。

目錄

1	引言	3
1.1	什麼是 earthUI ?	3
1.2	一個套件，三個常式	4
1.3	三種用途模式	4
1.4	房地產專用功能	4
1.5	安裝 earthUI 與 vProlog	5
1.6	快速入門	5
2	MLS 輸入資料要求	7
2.1	檔案格式與結構	7
2.2	估價模式所需的欄	7
2.3	特殊欄命名約定	8
2.4	資料品質與完整性	8
2.5	估價對象列的位置	8
3	通用模式	9
3.1	概述	9
3.2	側邊欄 workflow	9
3.3	主面板頁籤	9
3.4	設定持久化	10
3.5	深色模式	10
3.6	語言環境與區域設定	10
3.6.1	支援的國家	11
3.6.2	紙張覆蓋設定	11
3.6.3	儲存預設設定	11

4	估價模式	12
4.1	估價對象列的處理	12
4.2	生效日期與銷售日齡	12
4.3	特殊欄指定	12
4.4	殘差貢獻（稀有特徵估計）	13
4.5	RCA 調整概述	13
5	市場區域分析模式	15
5.1	與估價模式的區別	15
5.2	何時使用市場模式	15
6	變數選擇	16
6.1	目標變數	16
6.2	預測變數表	16
6.3	資料類型檢測與覆蓋	16
6.4	Factor 與 Linear 標誌	16
6.5	特殊欄類型參考	17
6.6	多個目標變數	17
7	參數選擇	18
7.1	基本參數	18
7.2	前向過程	18
7.3	允許的交互	19
7.4	剪枝	19
7.5	交叉驗證	19
7.6	子集過濾	19
7.7	推薦參數值	20
7.7.1	前向過程參數	20
7.7.2	剪枝參數	21
7.7.3	交叉驗證參數	21
7.7.4	變異數模型	22
7.7.5	Earth 輸出檔案	22
7.8	設定預設值	22
8	擬合 Earth 模型	23
8.1	Fit 按鈕	23
8.2	擬合模態框與追蹤輸出	23
8.3	擬合日誌	23
9	下載資料	24
9.1	輸出欄	24
9.2	欄排序與 Excel 格式	24
9.3	CQA 評分	25
9.4	排序與估價對象列	25
9.5	mgcvUI 自動匯出	25
9.6	glmnetUI 匯入	25
9.6.1	主要優勢	25
9.6.2	重要提示：權重欄必須一致	26
9.6.3	工作流程	26

10 RCA 計算與下載	27
10.1 開啟 RCA 對話方塊	27
10.2 CQA 評分插值	27
10.3 輸出欄	27
10.4 RCA Adjustments 頁籤	28
11 銷售比較網格	29
11.1 概述與用途	29
11.2 可比案例選擇模態框	29
11.3 網格佈局	29
11.4 售價、讓價優惠與 Net SP	30
11.5 分組列（位置、地塊、屋齡）	30
11.6 模型變數列	30
11.7 CQA Residual 與殘差特徵列	30
11.8 調整後售價公式	30
11.9 工作表保護與可編輯儲存格	31
11.10 在 Excel 中使用網格	31
12 下載報告	32
12.1 生成 Quarto 報告	32
12.2 轉換 Quarto 報告	32
12.3 報告內容	32
13 後處理模組：彈性網絡與 GAM	34
13.1 為什麼需要輔助擬合？	34
13.2 啟動與共享狀態	34
13.3 彈性網絡常式	34
13.4 GAM 常式	34
13.5 一個實用的工作流	34
14 示範資料集：Appraisal_1.csv	36
14.1 描述	36
14.2 資料欄	36
14.3 建議的快速上手步驟	37
15 實驗性功能：Prolog 文字提取與推導規則	38
15.1 它的作用	38
15.2 環境要求：vProlog (Trealla)，而非 SWI-Prolog	38
15.3 啟用與執行	38
15.4 推導規則	39
15.5 帶版本號的活動資料檔案	39
15.6 編程介面與隨附檔案	39
15.7 相關實驗：氣候區先驗	40
15.7.1 超越加利福尼亞：轄區包與市場區域模組	40

1 引言

1.1 什麼是 earthUI？

earthUI 是 R 語言 earth 套件的圖形使用者介面，該套件實作了多元自適應迴歸樣條 (MARS, multivariate adaptive regression splines) (Friedman 1991)。建模引擎是 Stephen Milborrow 的

earth 套件 (Milborrow 2024), Milborrow (n.d.[a]) 與 Milborrow (n.d.[b]) 對其有詳細的說明文件。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。您從 R 中啟動它，匯入資料集 (CSV 或 Excel)，配置模型，然後互動式地進行擬合。

該應用程式提供了完整的工作流程：資料匯入、變數配置、模型擬合、診斷圖、變數重要性、模型方程，以及可下載的 HTML、Word 或 PDF 格式報告。

1.2 一個套件，三個常式

從當前版本起，earthUI 套件還包含了它的兩個後處理模組，這兩個模組在早期版本中是獨立的套件 glmnetUI 和 mgcvUI。安裝 earthUI 即安裝了全部三個常式；每個常式都作為獨立的本地 Shiny 應用程式執行：

- `earthUI::launch()` ——本文所描述的 earth (MARS) 應用程式，執行在連接埠 7878 上。`earth()` 是核心估值引擎。
- `earthUI::launch_glmnet()` ——彈性網後處理模組 (連接埠 7879)，它匯入 earth 模型的基函數，並在 lasso / 彈性網正則化下重新估計它們 (參見下文「glmnetUI Import」)。
- `earthUI::launch_mgcv()` ——GAM 後處理模組 (連接埠 7880)，它以 earth 模型的節點位置作為 mgcv 平滑項的起點 (參見下文「mgcvUI Auto-Export」)。

這三個常式共享同一個專案樹、同一個設定資料庫和同一套支撐基礎設施，因此在一個應用程式中開啟的專案可以立即供其他應用程式使用。glmnet 和 mgcv 常式是對 earth 模型價值結論的佐證，而不是相互競爭的工作流程：預期的使用模式是先擬合並完善 earth 模型，然後在需要佐證時針對同一專案執行這些模組。

1.3 三種用途模式

earthUI 以專案的形式組織工作——專案是有名稱的實體，將您的輸入檔案、輸出和設定集中儲存在一個固定位置。每個專案都帶有一個 **Purpose** (用途)，在建立專案時於 New Project wizard (新增專案精靈) 中選擇。共有三種模式：

- **General** (通用) ——適用於任何類型的總體或資料集的 Earth 迴歸。這是預設模式。它提供完整的 MARS 建模工作流程，不含任何特定領域的附加功能。
- **For Appraisal** (估價用途) ——為房地產估價量身打造的 Earth 迴歸。增加了針對單一房產估值的功能，包括估價對象 (subject) 處理、特殊欄指定、殘差約束法 (RCA)，以及銷售比較網格 (Sales Comparison Grid) 的生成。
- **Market Area Analysis** (市場區域分析) ——為市場區域研究量身打造的 Earth 迴歸。增加了在既定市場中分析成組房產的功能，而不是對單一估價對象進行估值。

在這三種模式下，核心建模引擎完全相同——您始終是在擬合一個 Earth (MARS) 模型。用途設定控制的是哪些附加工具和介面元素可用。

每個專案的用途是固定的。專案的用途——連同其國家、州、縣和城市——在專案建立時即被鎖定。切換到另一個專案會將 earthUI 重置為乾淨的預設狀態 (已匯入的資料、模型結果、頁籤、變數配置和 earth 參數全部清空)，然後恢復該專案針對其用途儲存的配置。

1.4 房地產專用功能

當選擇 **For Appraisal** 或 **Market Area Analysis** 時，earthUI 會啟用若干為房地產分析設計的功能：

- **特殊欄指定**——每個預測變數都可以被標記為某種特殊角色，例如 `contract_date`、`dom`、`concessions`、`latitude`、`longitude`、`living_area`、`lot_size`、`actual_age`、`effective_age`、`area`、`site_dimensions` 或 `display_only`。這些指定控制該欄在擬合、輸出和銷售網格生成過程中的處理方式。完整參考請參見第 6 章。
- **經緯度取整**——被指定為 `latitude` 或 `longitude` 的欄會自動取整到小數點後 3 位，以防止過擬合。

- **Sale Age 欄**——當某欄被指定為 `contract_date` 且提供了 Effective Date (生效日期) 時，earthUI 會計算一個 `sale_age` 欄 (銷售日期與生效日期之間的天數)，並將其作為預測變數代入。
- **RCA 計算 (僅限估價模式)**——在估價模式下，模型擬合完成後，earthUI 可以計算殘差約束法 (RCA) 輸出，生成逐個可比銷售案例的調整、淨 / 總調整匯總，以及估價對象的調整後銷售價格。
- **銷售比較網格 (僅限估價模式)**——生成一個多工作表的 Excel 活頁簿，其中包含可比案例調整、殘差特徵分解和調整後銷售價格的公式。參見第 11 章。

提示：*General* 用途模式適用於任何資料集——金融、科學、工程或房地產。估價模式和市場模式只是為房地產專業人士增加了便利功能。

1.5 安裝 earthUI 與 vProlog

earthUI 已發布在 CRAN 上，因此標準安裝只需在 R 中執行一行命令：

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

本文描述的是當前版本。如果您閱讀本文時 CRAN 上仍是較舊的版本，可以從 GitHub 安裝開發版 (注意 `subdir` 參數——該套件位於儲存庫的 `pkg/` 資料夾中)：

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

vProlog (可選)。實驗性的備註處理功能 (參見本文末尾附近的「Prolog Text Extraction」) 使用 vProlog 套件——一個自包含的嵌入式 Prolog 引擎。它被有意設計為非強制依賴：earthUI 在沒有它的情況下也能安裝並執行，Prolog 相關功能只是在其安裝之前保持停用狀態。vProlog 通過作者的 *r-universe* 儲存庫分發，該儲存庫為 Windows 和 macOS 提供預編譯的二進位檔案 (無需編譯器工具鏈)：

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

不需要系統級的 Prolog 安裝 (無論是 SWI-Prolog 還是其他)——Trealla 引擎已編譯進套件本身。本文中的其他所有內容在沒有 vProlog 的情況下均可正常使用。

1.6 快速入門

使用 earthUI 的步驟：

1. 安裝套件——按上文所述從 CRAN 安裝。
2. 啟動應用程式：在 R 中執行 `earthUI::launch()`，或從命令行執行 `Rscript -e 'earthUI::launch()'`。應用程式將在您的網頁瀏覽器中開啟，連接埠為 7878。您也可以直接在瀏覽器中造訪 `http://localhost:7878` 來開啟應用程式。(後處理模組以同樣的方式啟動：`launch_glmnet()` 使用連接埠 7879，`launch_mgcv()` 使用連接埠 7880。)
3. 建立或開啟專案——在側邊欄的第 1 節中進行。New Project wizard 會記錄專案名稱、Purpose (General、For Appraisal 或 Market Area Analysis)、位置 (國家、州、縣、城市)，以及可選的初始資料檔案，該檔案會被複製到專案的 `in/` 資料夾中。
4. 匯入資料——在第 2 節中進行，該節列出活動專案 `in/` 資料夾中的 CSV 和 Excel 檔案。您通過 Finder 或檔案總管手動新增的檔案，在點選 Refresh (重新整理) 後會顯示出來。
5. 配置變數——選擇目標變數和預測變數，設定資料類型，並指定任何特殊欄角色。
6. 設定 **Earth** 參數——`degree`、`nprune`、子集過濾器以及其他 `earth()` 參數。
7. 擬合模型——點選「Fit Earth Model」(擬合 Earth 模型)，並在主面板中查看結果。

8. 匯出——將預測結果下載為 Excel，生成 Quarto 報告包，或（在估價模式下）計算 RCA 調整和銷售比較網格。

配置通過第 3 節和第 4 節中的「Save current as default」（將當前設定儲存為預設）按鈕儲存在伺服器端的專案資料庫中，以專案和用途為鍵（參見第 3 章「Settings Persistence」）。

2 MLS 輸入資料要求

對於房地產估價和市場分析工作流程，您的輸入資料通常來自多重掛牌服務（MLS，Multiple Listing Service）的匯出檔案。本章描述預期的檔案結構以及 earthUI 可以使用的欄。

2.1 檔案格式與結構

earthUI 接受 CSV 和 Excel (.xlsx、.xls) 檔案。匯入時，欄名會被自動轉換為 snake_case——例如，「Living SqFt」變為 living_sqft，「Contract Date」變為 contract_date，「Sale Price」變為 sale_price。這種規範化確保了整個工作流程中欄引用的一致性。匯入時使用的 CSV 分隔符號和小數點符號由區域設定決定（參見第 3 章「Locale & Regional Settings」）。

您的資料檔案應為平面表格，每個房產佔一行，每個屬性佔一欄。檔案的第一列必須包含欄標題。

2.2 估價模式所需的欄

雖然 earthUI 可以處理任何欄集合，但完整的估價工作流程（RCA 調整 + 銷售比較網格）在您的 MLS 匯出檔案中包含以下欄時效果最佳：

欄	特殊類型	用途
Sale Price	(目標變數)	模型的反應變數
Contract Date	contract_date	用於計算 sale_age（距生效日期的天數）
Listing Date	listing_date	在沒有 DOM 欄時，與合約日期一起用於計算 DOM
Days on Market	dom	在市天數；顯示在銷售網格中
Concessions	concessions	銷售讓步；銷售網格中 Net SP = Sale Price – Concessions
Living Area (SF)	living_area	啟用每平方英尺殘差 (residual_sf、cqa_sf)
Lot Size	lot_size	在銷售網格中歸入「Site Size Dimensions」列
Site Dimensions	site_dimensions	在銷售網格中與地塊面積歸為一組
Latitude	latitude	取整到小數點後 3 位；用於鄰近度計算和 Location 組
Longitude	longitude	取整到小數點後 3 位；用於鄰近度計算和 Location 組
Area ID	area	在銷售網格中歸入「Loc: Long Lat Area」列
Actual Age	actual_age	在銷售網格中歸入「Actual Age Effective Age」列
Effective Age	effective_age	在銷售網格中與實際屋齡歸為一組
Parcel Number	(無)	顯示在銷售網格的 APN 列中
Listing ID	(無)	顯示在銷售網格的 APN 列中
Address	display_only	顯示在銷售網格的 Address 列中；不參與模型

試算表的欄名可以使用外語——「特殊」名稱使用英語，以便 R 程式能對它們進行特殊處理。除此之外，給定的欄名會出現在迴歸模型、圖表，以及（在進行估價時）中間銷售網格 (Intermediate Sales Grid) 中。

並非所有欄都是必需的。earthUI 會自行適應——如果某欄缺失，相應的功能會被直接省略。例如，如果沒有指定 concessions 欄，銷售網格中的 Net SP 列將顯示未扣除讓步的銷售價格。不過，對於房地產定價模型，強烈建議提供某些欄以獲得可接受的擬合效果：

1. **Sale Age**——合約銷售日期與估價或分析的生效日期之間的天數。如果使用多年的銷售歷史資料，尤其是超過 5 年的時段，sale_age 往往在銷售價格估計中扮演核心角色。事實上，它常常重要到沒有它 earthUI 根本無法給出任何模型。
2. **Living Area**——也稱作「Living Sqft」、「GLA」（gross living area，總居住面積）等。這是銷售價格的另一個主要決定因素。

3. **Total Bath Count**——全衛、四分之一衛、半衛和四分之三衛浴的總數。例如，兩個全衛加一個半衛的值為 2.5。
4. **Garage Bays** 或 **Garage Area**——車庫車位數量或車庫面積（平方英尺）。
5. **Lot Size**——房產的土地面積，通常以平方英尺或英畝為單位。
6. **Longitude**、**Latitude**，如果有的話再加上 **Area ID**。這些變數的調整將在銷售網格中合併為單一的 Location 調整項。

2.3 特殊欄命名約定

earthUI 通過欄的特殊類型指定而非欄名來識別欄。您可以在 MLS 匯出檔案中為欄取任何名稱——關鍵在於您在變數配置表（第 6 章）中指定了正確的特殊類型。

例如，您的 MLS 可能將居住面積匯出為「GLA」、「Living SqFt」、「liv_area」或「gross_living_area」。匯入之後（此時它已變為 snake_case），您只需在 Special dropdown（特殊類型下拉式選單）中將其指定為 living_area。此後，無論其原始名稱如何，earthUI 都會將其用於每平方英尺殘差計算和銷售網格分組。

2.4 資料品質與完整性

- **遺漏值 (NA)**：任何預測變數或目標變數欄中含有 NA 值的列會在擬合前自動刪除。na.action 在內部始終設定為 na.fail，NA 的刪除在呼叫 earth() 之前完成。
- **日期欄**：日期應採用 R 可解析的格式（例如 2025-06-15、06/15/2025、June 15, 2025）。當至少 80% 的值能夠成功解析時，earthUI 會自動檢測日期欄。兩位數年份（例如 06/15/25）會被讀取為最接近的匹配年份，因此近期的 MLS 匯出檔案可以正確解析；四位數年份始終按字面值處理。
- **數值欄**：銷售價格、居住面積、地塊面積、讓步及類似欄位必須是數值型。如果您的 MLS 匯出的價格帶有貨幣符號或千分位分隔符號（例如「\$350,000」或「350.000,00」），您可能需要在匯入前清理這些資料。
- **因子欄**：區域 ID、風格或狀況等分類變數應包含數量適中的唯一值。earthUI 會自動檢測因子，但您可以覆蓋該檢測結果。

提示：匯入後請查看預測變數表中的 T/F/NA 欄——它顯示每欄的 true/false/NA 計數。遺漏值超過 30% 的欄會以紅色標出。請考慮排除高遺漏率的欄，或在匯入前清理資料。

2.5 估價對象列的位置

在 **Appraisal**（估價）模式下，第 1 列必須是估價對象 (subject)。其餘所有列均為可比銷售案例。估價對象的銷售價格可以留空 (NA) 或設為任意值——無論如何，earthUI 在擬合時都將其視為 NA。

在 **Market Area Analysis** 模式下，將估價對象放在第 1 列是可選的，且不會自動排除任何列。若要將估價對象所在列（或任何其他列）排除在擬合之外，可在指定的權重欄中將其值設為零，或使用子集過濾器。

在 **General** 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

通用模式 (General Purpose mode) 是啟動 earthUI 時的預設模式。它為任何資料集——而不僅僅是房地產資料——提供完整的多元自適應迴歸樣條 (MARS, multivariate adaptive regression splines) 建模工作流。您可以將 earthUI 用於科學資料、金融分析、工程研究，或任何希望探索變數間非線性關係與交互作用的迴歸問題。

在通用模式下，介面會省略房地產專用功能（特殊欄、銷售日齡、座標捨入、RCA）。側邊欄經過精簡，聚焦於變數選擇、參數配置、模型擬合與匯出。

3.2 側邊欄工作流

側邊欄按編號組織為可折疊的分區，引導您從專案選擇一路走到匯出：

1. Project (專案) ——選擇一個現有專案，或通過 New Project (新增專案) 精靈建立新專案；該精靈會記錄用途、國家和行政區位置（建立後全部鎖定）。當前活動專案決定了輸入檔案的讀取位置以及所有輸出的寫入位置——不存在單獨的輸出資料夾欄位。

2. Import Data (匯入資料) ——從活動專案的 in/ 資料夾中選擇 CSV 或 Excel 檔案；Refresh (重新整理) 連結會重新掃描該資料夾，以發現手動新增的檔案。對於包含多個工作表的 Excel 檔案，會出現工作表選擇器。欄名會被自動轉換為 snake_case。

3. Variable Configuration (變數配置) ——目標變數選擇器（支援多個目標變數）、Effective Date (生效日期)，以及帶有 Include、Factor、Linear 和 Latent 核取方塊的預測變數表。詳見第 6 章。

4. Earth Call Parameters (Earth 呼叫參數) ——earth() 函數的全部參數：degree、penalty、nk、剪枝方法、交叉驗證等。完整的參數參考見第 7 章。

Execute Prolog Processing (執行 Prolog 處理) ——位於第 4 與第 5 分區之間的一個可選、無編號分區，除非在 Settings (設定) 中啟用了實驗性的 Prolog 功能，否則處於停用狀態。見第 15 章。

5. Fit Earth Model (擬合 Earth 模型) ——一個隨機種子欄位和非同步執行模型的按鈕。擬合的詳細說明見第 8 章。

6. Download Estimated Target Variable(s) & Residuals (下載估計的目標變數與殘差) ——（在估價模式和市場模式下標題為「Download Estimated Sale Prices & Residuals」，即下載估計售價與殘差）將預測值、殘差、CQA 評分以及各 g 函數的貢獻匯出為一個 Excel 檔案。見第 9 章。

7–8. RCA Adjustments and Sales Grid (RCA 調整與銷售網格) ——僅限估價模式。見第 10 章和第 11 章。

9. Generate Quarto Report (生成 Quarto 報告) ——將一個自包含的 Quarto 報告包 (.qmd 原始檔及資源檔案) 寫入專案的輸出資料夾。在非估價模式下編號為 7。見第 12 章。

10. Convert Quarto Report (轉換 Quarto 報告) ——將任意 .qmd 檔案渲染為 HTML、Word 或 PDF。見第 12 章。

3.3 主面板頁籤

擬合完成後，主面板提供十個頁籤：

頁籤	內容
Data (資料)	匯入資料的預覽。在估價模式下，拆分為「Subject Property」(估價對象)和「Comparable Sales」(可比銷售案例)兩個表格。
Equation (方程)	以 LaTeX/MathJax 顯示的擬合模型方程，展示每個基函數及其係數。
Summary (摘要)	關鍵指標 (R^2 、GRSq、GCV、RSS、CV R^2) 和係數表。
Variable Importance (變數重要性)	預測變數重要性得分的長條圖和排名表。
Contribution (貢獻)	g 函數貢獻圖。從下拉式選單中選擇一個 g 函數；單變數預測變數顯示折線圖，雙變數交互項 ($\text{degree} \geq 2$) 顯示 3D 透視圖和等高線圖。
Correlation (相關性)	預測變數相關性的熱圖。
Diagnostics (診斷)	三幅圖：殘差 vs 擬合值、常態 Q-Q 圖，以及實際值 vs 預測值。
RCA Adjustments (RCA 調整)	(僅估價模式) 各可比銷售案例的殘差調整百分比 (Residual Adj %)、淨調整百分比 (Net Adj %) 和總調整百分比 (Gross Adj %) 的直方圖，在第 7 步 (Calculate RCA Adjustments) 完成後填充。
Residual Contribution (殘差貢獻)	通過將模型殘差對模型從未使用過的欄進行聯合迴歸，估計稀有特徵或被排除特徵的市場貢獻 (見 subsection 4.4)。
ANOVA	ANOVA 分解表，顯示每個基函數對 RSS 的貢獻。
Earth Output (Earth 輸出)	來自 'earth::summary()' 的原始文字輸出，包括剪枝過程的詳細資訊。

3.4 設定持久化

earthUI 將您的配置儲存在伺服器端的專案資料庫中 (位於 regProj 根目錄下的 `projects.sqlite`)，以專案和用途為鍵。儲存是按鈕驅動的：第 3 分區中的「Save current as default」(將當前設定儲存為預設) 按鈕儲存目標變數、生效日期和預測變數配置，而第 4 分區中的按鈕儲存 earth 呼叫參數和 Allowed Interactions (允許的交互項) 矩陣。每個按鈕只更新自己負責的那一半，因此它們絕不會相互覆蓋，並且擬合不會自動儲存——您可以自由試驗而不必擔心覆蓋已儲存的預設設定。一個專案的多個輸入檔案 (一個小型測試抽樣、完整資料集) 共享同一份配置；而且由於設定存放在專案目錄樹中，當您同步、備份或共享 regProj 資料夾時，設定會隨之遷移。

3.5 深色模式

點選右上角的月亮/太陽圖示可在淺色與深色主題之間切換。主題偏好儲存在本機儲存中，並在工作階段之間持續生效。

3.6 語言環境與區域設定

earthUI 通過基於國家的語言環境系統支援國際化的數字、日期和 CSV 格式約定。頂部選單欄 **Settings** (設定，齒輪圖示) 選單中的 **Country** (國家) 下拉式選單可為 31 個受支援的國家選擇預設。每個預設配置以下內容：

- **CSV 分隔符號**——美國/英國/日本使用逗號 (,)，歐洲大部分國家使用分號 (;)，因為那裡逗號被用作小數點。
- **小數點符號**——句點 (.) 或逗號 (,)。
- **千分位分隔符號**——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/烏克蘭/俄羅斯) 或撇號 (瑞士)。

- 日期格式——MM/DD/YYYY（美國）、DD/MM/YYYY（歐洲大部分國家）或 YYYY-MM-DD（瑞典/立陶宛/日本/加拿大）。
- 紙張尺寸——Letter（美國/加拿大/墨西哥）或 A4（其他所有國家）。

3.6.1 支援的國家

國家	CSV 分隔符號	小數點	千分位分隔符號	日期	紙張
美國、加拿大 (csv)	,	.	,	MDY/YMD	Letter
英國、愛爾蘭、澳洲、紐西蘭	,	.	,	DMY	A4
德國、奧地利、比利時、荷蘭、義大利、西班牙、葡萄牙、丹麥、巴西	;	,	.	DMY	A4
法國、波蘭、捷克、挪威、拉脫維亞、愛沙尼亞	;	,	(空格)	DMY	A4
芬蘭、瑞典	;	,	(空格)	DMY/YMD	A4
瑞士	;	,	,	DMY	A4
烏克蘭、俄羅斯	;	,	(空格)	DMY	A4
土耳其	;	,	.	DMY	A4
日本、韓國	,	.	,	YMD	A4
墨西哥	,	.	,	DMY	Letter
立陶宛	;	,	(空格)	YMD	A4

3.6.2 紙張覆蓋設定

在國家選擇器下方，**Paper**（紙張）下拉式選單（Letter 或 A4）允許您在不切換國家的情況下覆蓋報告頁面尺寸。CSV 分隔符號、小數點符號和日期格式順序仍遵循所選國家的預設；當您更改國家時，紙張覆蓋設定會重置為該國家的預設值。

3.6.3 儲存預設設定

點選 Settings 選單中的 **Save**（儲存）可全域儲存您的語言環境偏好。這些預設設定適用於今後所有工作階段，無論您開啟哪個專案。專案級設定（目標變數、預測變數、參數）單獨儲存在專案資料庫中；語言環境預設設定則通過使用者級 SQLite 資料庫在所有專案間持續生效。這種兩級設計面向審計事務所等需要處理來自多個國家資料的機構——將您最常用的國家設為預設，需要時再進行調整。

提示：圖表座標軸和斜率標籤上的數字格式會自動適配您的語言環境。德國語言環境使用句點作為千分位分隔符號（200.000），芬蘭使用空格（200 000），瑞士使用撇號（200'000）。不顯示任何貨幣符號——*earthUI* 與貨幣無關。

4 估價模式

當您選擇 **For Appraisal**（用於估價）作為用途（Purpose）時，earthUI 會將自身配置為面向單一物業估值。第 3 章描述的所有功能仍然可用；本章僅涵蓋估價模式特有的附加功能。

4.1 估價對象列的處理

在估價模式下，資料集的第 1 列是估價對象（**subject**）物業，其餘所有列均為可比銷售案例。您的輸入檔案必須按此方式組織（見第 2 章）。估價對象的售價可以留空或設為任意值——earthUI 在擬合時會自動將其視為 NA。

匯入後，Data 頁籤會拆分為兩個部分：**Subject Property**（估價對象物業，第 1 列）和 **Comparable Sales**（可比銷售案例，第 2 列及之後）。第 1 列始終被排除在模型擬合之外——通知訊息「Skipping row 1 (subject). Fitting on N rows.」（跳過第 1 列（估價對象）。在 N 列上擬合。）會對此加以確認。擬合後，模型仍會為估價對象列生成預測值，在輸出中顯示為 `est_<target>`。

4.2 生效日期與銷售日齡

在每種用途模式下，Variable Configuration 分區中都會出現一個 **Effective Date**（生效日期）欄位（預設為當天日期）——通用分析和市場分析同樣有生效日期，而不僅僅是估價——但它只在估價模式和市場模式下才會被使用。在這兩種模式下，如果您在 Special（特殊欄）下拉式選單中將某欄指定為 `contract_date`，earthUI 會計算一個 `sale_age` 欄——即每筆銷售的合約日期與生效日期之間的整數天數。該欄會取代原始日期欄作為預測變數。

在指定合約日期後第一次點選 Fit 時，earthUI 會建立 `sale_age` 欄，並通知您再次點選 Fit 以將其納入模型。

4.3 特殊欄指定

在估價模式和市場模式下，Variable Configuration 表中的每個預測變數旁都會出現一個 **Special**（特殊）下拉式選單。特殊類型及其作用的完整列表如下：

特殊類型	作用
no	預設——無特殊處理
contract_date	觸發基於生效日期的 <code>sale_age</code> 自動計算
listing_date	在不存在 <code>dom</code> 欄時，與合約日期一起用於計算 DOM
dom	標示「在售天數」(Days on Market) 欄；顯示在銷售網格 (Sales Grid) 中
concessions	標示銷售讓步 (concessions)；用於銷售網格中的淨售價公式 (售價 - 讓步)
latitude	數值捨入到小數點後 3 位；用於 Haversine 鄰近度計算及銷售網格中的 Location 分組
longitude	數值捨入到小數點後 3 位；用於 Haversine 鄰近度計算及銷售網格中的 Location 分組
area	在銷售網格中歸入「Loc: Long Lat Area」列
living_area	在下載輸出中啟用按平方英尺計算的殘差 (<code>residual_sf</code> 、 <code>cqa_sf</code>)
lot_size	在銷售網格中歸入「Site Size Dimensions」列
site_dimensions	在銷售網格中與地塊面積歸為一組
actual_age	在銷售網格中歸入「Actual Age Effective Age」列
effective_age	在銷售網格中與實際屋齡歸為一組
display_only	該欄包含在匯出中，但排除在模型擬合之外（允許多欄）
remarks	（實驗性）自由文字備註欄，由可選的 Prolog 步驟解析為特徵欄（允許多欄）；見第 15 章
list	（實驗性）逗號分隔的欄，由可選的 Prolog 步驟拆分為獨熱 (one-hot) TRUE/FALSE 欄（允許多欄）；見第 15 章

每種特殊類型只允許指定一欄 (`display_only`、`remarks` 和 `list` 除外)。將某個特殊類型指定給第二欄時，會自動從第一欄上清除該指定。變數名旁邊會出現一個藍色小徽章，顯示其被指定的特殊類型。

4.4 殘差貢獻（稀有特徵估計）

對於因過於稀有而無法進入迴歸通道的特徵——例如僅在少數幾筆銷售中出現的工作間 (workshop) ——它們屬於 RCA 殘差層。**Residual Contribution**（殘差貢獻）頁籤通過普通最小二乘法，將擬合模型的殘差對模型從未使用過的欄進行迴歸，以估計這些特徵的市場貢獻。選擇多欄時會對它們進行聯合估計，因此相關的配套設施（例如 `outbuilding_sqft` 加上 `pr_has_workshop`）會分攤貢獻，而不會重複計算。二元標誌欄產生一次性總額估計，並列出每筆帶該標誌的銷售的殘差作為證據；數值欄產生每單位貢獻；報告的標準誤為一個 SE（通常以 $t \geq 2$ 作為顯著性判斷標準）。支持觀測數少於十個的項會在生成的工作底稿註記中自動獲得 CAUTION（警示）標記；該註記以一個段落記錄方法、估計值和注意事項，供報告使用。只有模型實際使用過的預測變數才會被排除在候選列表之外——一個曾提供給擬合但從未被前向過程選中的稀有標誌欄仍然符合條件，因為模型沒有吸收它的任何價值。

4.5 RCA 調整概述

Calculate RCA Adjustments & Download（計算 RCA 調整並下載）按鈕（側邊欄第 7 分區，僅在估價模式下擬合完成後可見）計算每個可比銷售案例相對於估價對象的市場推導調整值。完整的 RCA 工作流在第 10 章中描述。

計算完 RCA 調整後，**Generate Sales Grid & Download**（生成銷售網格並下載）按鈕（側邊欄第 8 分區）變為可用。銷售比較網格 (Sales Comparison Grid) 工作流在第 11 章中描述。

提示：您為估價對象指定的 *CQA* 評分決定了殘差分布中有多大比例歸屬於估價對象。評分 *5.00* 會將估價對象置於可比銷售案例的中位數位置。

5 市場區域分析模式

當您選擇 **Market Area Analysis**（市場區域分析）作為用途時，earthUI 提供與估價模式相同的房地產專用功能（特殊欄、銷售日齡、座標捨入），但其定位是分析一組物業，而不是對單一估價對象進行估值。

5.1 與估價模式的區別

- 不自動排除估價對象——與估價模式不同，市場模式不會自動將第 1 列排除在擬合之外。所有列都會被納入；若要將估價對象列排除在擬合之外，可在指定的權重欄中為其賦零值（如示範資料集所示），或使用子集過濾器。
- 沒有 **RCA** 或銷售網格分區——「Calculate RCA Adjustments & Download」和「Generate Sales Grid & Download」步驟不可用。市場模式專注於模型擬合與輸出，而非逐個可比案例的調整。
- 側邊欄編號——由於沒有 RCA 和銷售網格步驟，Generate Quarto Report 分區的編號為 7 而非 9（Convert Quarto Report 分區仍為 10）。

5.2 何時使用市場模式

以下情形適合使用市場區域分析模式：

- 為某個社區或市場區域構建迴歸模型，以理解價值驅動因素
- 分析建築面積、屋齡、地塊面積和區位等變數如何影響一組物業的售價
- 為市場狀況分析或社區界定準備支持材料
- 使用的資料集第 1 列包含一個估價對象物業，但您希望可以自由選擇將其納入或排除在擬合之外

提示：市場模式也適用於一般的房地產迴歸場景——您需要特殊欄功能（銷售日齡、座標捨入），但不需要 **RCA** 調整工作流。

6 變數選擇

側邊欄第 3 節——**Variable Configuration**（變數配置）——用於選擇哪些欄參與模型，以及它們如何被處理。

6.1 目標變數

第 3 節頂部的 **Target (response) variable(s)**（目標（反應）變數）下拉式選單列出資料集中的每一欄。選擇一欄可擬合標準的單反應模型；選擇多欄則擬合多反應模型。選擇多個目標變數時，模型使用 `earth(cbind(y1, y2, ...) ~ .)` 同時擬合所有反應。

被選為目標變數的欄會自動從預測變數列表中排除。

6.2 預測變數表

在目標變數選擇器下方，一個表格列出其餘每一欄，包含以下欄位：

欄	說明
Variable （變數）	欄名（若被截斷，完整名稱顯示在工具提示中），前面帶有一個 ×/+ 切換按鈕（見下文）。在估價/市場模式下，藍色徽章顯示所分配的特殊類型。
Type （類型）	資料類型下拉式選單： <code>numeric</code> 、 <code>integer</code> 、 <code>character</code> 、 <code>logical</code> 、 <code>factor</code> 、 <code>Date</code> 、 <code>POSIXct</code>
Include （納入）	核取方塊——將此欄作為預測變數納入模型
Factor （因子）	核取方塊——將此欄視為分類變數
Linear （線性）	核取方塊——強制僅以線性方式進入模型（不生成鉸鏈函數）
Latent （潛變數）	核取方塊——將此欄完全排除在模型之外（既不進入 <code>earth</code> ，也不包含在匯出到 <code>glmnetUI</code> / <code>mgcvUI</code> 的資料中），保留用於潛變數分析
Special （特殊類型）	下拉式選單（僅估價/市場模式）——見下文「特殊欄類型參考」
T/F/NA	每欄三個計數：對邏輯欄為 TRUE/FALSE/NA；對數值欄為 ≠ 0/= 0/NA。當遺漏值超過 30% 時以紅色顯示

表格上方的提示列解釋了這些縮寫：「Type = 欄資料類型，Inc = 作為預測變數納入，Factor = 視為分類變數，Linear = 僅線性（無鉸鏈），Special = 欄角色（例如 `contract_date`）。」

軟性停用而非刪除。每個變數名旁的 × 切換按鈕會將該列變灰，並使該欄不參與模型，但不會將其從資料中刪除；按鈕隨之變為綠色的 +，可重新啟用該欄。欄永遠不會被銷毀。

6.3 資料類型檢測與覆蓋

`earthUI` 在匯入時自動檢測資料類型。數值、整數、邏輯、因子和日期欄都會被識別。看起來像日期的字元欄（至少 80% 的值能按常見日期格式解析）會被歸類為 `Date`。

你可以通過更改 **Type** 下拉式選單來覆蓋任何檢測結果。當你將某欄改為 `character` 或 `factor` 時，Factor 核取方塊會自動勾選。更改類型會影響該欄傳遞給 `earth()` 函數的方式。

6.4 Factor 與 Linear 標誌

- **Factor**——勾選後，該欄被視為分類變數。這適用於 `style`、`area_id` 或 `grade` 等表示離散組別而非連續測量值的欄。因子欄以指示（虛擬）變數的形式進入模型。
- **Linear**——勾選後，該欄被強制以線性方式進入模型——不會為其建立鉸鏈（分段線性）函數。當你確定某個變數與目標變數之間存在嚴格的線性關係時，此選項很有用。

6.5 特殊欄類型參考

在估價和市場模式下，**Special** 下拉式選單提供以下選項。每種類型最多只能分配給一欄（`display_only`、`remarks` 和 `list` 除外，它們允許多欄）：

日期與時間類型：

- `contract_date`——觸發自動計算 `sale_age`。原始日期欄被替換為一個整數欄，度量銷售日期與 Effective Date（評估基準日）之間的天數。
- `listing_date`——當未指定明確的 `dom` 欄時，用作計算在市天數（DOM = 合約日期 - 掛牌日期）的後備來源。
- `dom`——標示在市天數（Days on Market）欄。顯示在 Sales Grid（銷售比較網格）的 APN 列和 Date of Sale 列中。

貨幣類型：

- `concessions`——標示銷售讓步（賣方信貸、買方激勵等）。在 Sales Grid 中用於計算淨售價： $\text{Net SP} = \text{售價} - \text{讓步}$ 。

面積與位置類型：

- `latitude`——值自動四捨五入到小數點後 3 位以防止過擬合。用於 Haversine 鄰近度計算（估價對象（subject）到每個可比銷售案例的距離），並歸入 Sales Grid 的 Location 列。
- `longitude`——與 `latitude` 相同的捨入處理。用於鄰近度計算和 Location 組。
- `area`——通常為社區或區域識別碼。與 `latitude`、`longitude` 一起歸入 Sales Grid 的 “Loc: Long | Lat | Area” 列。
- `living_area`——啟用下載輸出中的每平方英尺殘差計算（`residual_sf` 和 `cqa_sf`）。
- `lot_size`——歸入 Sales Grid 的 “Site Size | Dimensions” 列。
- `site_dimensions`——與 `lot_size` 一起歸入 Sales Grid（例如 “75x120”）。

屋齡類型：

- `actual_age`——歸入 Sales Grid 的 “Actual Age | Effective Age” 列。
- `effective_age`——與 `actual_age` 一起歸入 Sales Grid。

顯示類型：

- `display_only`——該欄包含在 Excel 匯出中，但完全不參與模型擬合。適用於地址欄位、MLS 編號、宗地編號或其他不應作為預測變數的參考資料。允許多欄使用此指定。

文字提取類型（實驗性）：

- `remarks`——一個自由文字欄（例如 `public_remarks`），由可選的 Prolog 步驟解析為特徵欄。在 Settings（設定）中啟用該功能之前不起任何作用。允許多欄。參見第 15 章。
- `list`——一個逗號分隔的欄（例如 `existing_structures`），由同一步驟拆分為獨熱編碼的 `<col>_<value>` TRUE/FALSE 欄。允許多欄。參見第 15 章。

6.6 多個目標變數

選擇多個目標變數會擬合一個多反應 Earth 模型。選擇多個目標變數時：

- 模型同時預測所有反應，各目標變數共享基函數
- **wp (response weights)**（反應權重）按鈕被啟用，允許為每個目標變數分配數值權重
- 變異數模型（`varmod.method`）被停用（多反應模型不支援）
- 結果頁籤會顯示每個反應的指標、方程和診斷圖

提示：指定為 `display_only` 的欄保留在資料集中並出現在 Excel 匯出中，但完全不出現在預測變數表中。適用於 ID 欄、地址或其他參考欄位。

7 參數選擇

側邊欄第 4 節——**Earth Call Parameters** (Earth 呼叫參數)——提供對 `earth()` 函數所接受的全部參數的存取。每個參數都帶有一個藍色說明圖示 (?) 及工具提示說明。參數按可折疊的小節組織。

7.1 基本參數

參數	預設值	說明
subset	(空)	列過濾表達式。見下文「子集過濾」。
weights	NULL	用於個案(列)權重的欄選擇器。僅列出數值欄。
wp	NULL	多目標模型的反應權重。按鈕會開啟一個對話方塊，為每個目標變數提供一個數值輸入(預設各為 1.0)。
keepxy	關閉	在模型對象中保留 <code>x</code> 、 <code>y</code> 、 <code>subset</code> 和 <code>weights</code> 。
trace glm	0 無	擬合輸出的追蹤級別 (0 到 5)。可選的 GLM 族： <code>gaussian</code> 、 <code>binomial</code> 或 <code>poisson</code> 。
degree	1	最大交互階數。將 <code>degree</code> 設為 ≥ 2 會自動啟用交叉驗證，並顯示 Allowed Interactions (允許的交互) 矩陣。
penalty	2	每個節點的 GCV 懲罰。值越高，模型越簡單。

7.2 前向過程

參數	預設值	說明
nk	自動	剪枝前的最大模型項數。自動填充為所載入資料的推薦值(見下文「推薦參數值」)。
thresh	0.001	前向步驟閾值。值越小，允許的模型項越多。
minspan	自動	節點之間的最小間隔。自動填充為推薦值。負值設定每個預測變數的最大節點數。
endspan	自動	端部間隔——節點到資料邊緣的最小距離。自動填充為推薦值。
newvar.penalty	0	引入新變數的懲罰(鼓勵重用已有的預測變數)。
fast.k	20	快速 MARS 演算法中考慮的父項數量。
fast.beta	1	控制快速 MARS 的老化因子。

7.3 允許的交互

當 **degree** 設為 2 或更高時，基本參數下方會出現一個交互矩陣。它是一個上三角形的核取方塊網格，每個預測變數對對應一個核取方塊。勾選表示允許這兩個預測變數交互；取消勾選則禁止該特定交互。

點選預測變數名稱（列或欄標籤）可切換該變數的所有交互。頂部的 **Allow All**（全部允許）和 **Clear All**（全部清除）核取方塊提供批量控制。矩陣使用固定表頭，滾動時列、欄標籤保持可見。

一條資訊提示會提醒你：「交互項會增加過擬合的風險。交叉驗證已啟用（10 折）。」

7.4 剪枝

參數	預設值	說明
pmethod	backward	剪枝方法：backward、none、exhaustive、forward、seqrep 或 cv。
nprune	NULL	剪枝後的最大模型項數。留空則自動選擇。

7.5 交叉驗證

參數	預設值	說明
nfold	10	交叉驗證折數。設為 0 可停用交叉驗證。
ncross	20	交叉驗證重複次數。
stratify	開啟	對交叉驗證樣本進行分層，使每一折具有相似的反應分布。

當 $\text{degree} \geq 2$ 時，交叉驗證會自動啟用（**nfold** 設為 10），以幫助防範交互項帶來的過擬合。

7.6 子集過濾

subset 文字輸入框接受一個 R 表達式，用於篩選哪些列參與模型擬合。你可以直接輸入表達式（例如 `sale_age < 365 & area_id == 460`），也可以使用 **Build filter...**（構建過濾器…）按鈕以視覺化方式構建。

Build Filter Dialog（構建過濾器對話方塊）——點選“Build filter...”開啟一個引導式對話方塊。每個條件列包含一個欄下拉式選單、運算子（<、>、<=、>=、==、!=），以及一個根據欄類型自適應的值輸入框：數值欄為數字輸入，日期欄為日期選擇器，字元/因子欄為唯一值下拉式選單。各條件通過 **AND**（&）或 **OR**（|）連接運算子相連。底部的預覽會顯示表達式以及匹配的列數。點選 **Apply**（應用）將表達式插入文字輸入框。

日期欄在手動表達式中必須使用 `as.Date("...")` 或 `as.POSIXct("...")` 包裝。Build Filter 對話方塊會自動處理這一點。

子集過濾是非破壞性的——被排除的列仍保留在資料集中，並在 Excel 匯出中獲得預測值。

提示：將 **OR** 條件與 **AND** 組合時，請使用括號進行分組。若不加括號，`A & B | C` 會被解釋為 `(A & B) | C`，這可能不符合你的本意。

7.7 推薦參數值

earthUI 會在第 4 節的關鍵參數下方顯示推薦值。這些推薦值會根據擬合列數 (n) 和所選預測變數數 (p) 動態更新。這些公式來源於 Friedman 的 MARS 論文、earth 的內部演算法以及經驗測試。**nk**、**minspan** 和 **endspan** 欄位會在資料載入時預填充其推薦值；手動編輯的值會被保留。

7.7.1 前向過程參數

nk (剪枝前的最大模型項數)：

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

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

Earth 的預設值 $\min(200, \max(20, 2p)) + 1$ 未考慮資料集規模，對大資料集而言可能過小，會在前向過程充分探索所有預測變數之前就對其加以限制。 $\lfloor n/10 \rfloor$ 一項允許大約每 10 個觀測值對應一個模型項——這是避免過度參數化的標準經驗法則。100 的上限可避免收益遞減。

minspan (節點之間的最小觀測數)：

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

endspan (距資料邊界的最小觀測數)：

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

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

Earth 自動計算的 minspan 使用 Friedman 的公式 43： $\lfloor (-\ln(-\ln 0.95) + \ln(p \cdot n)) / (2.5 \ln 2) \rfloor$ ，它按 $\ln(np)$ 縮放，對大資料集增長過慢。在 $n=1500$ 時它僅得出 ≈ 7 ，使每個連續預測變數約有 245 個候選節點位置——太多了，會讓前向過程去擬合噪聲。推薦公式的目標是在大 n 時每個預測變數約有 100 個候選位置，而在小 n 時（此時 earth 的自動計算經過了充分檢驗）沿用其自動計算結果。endspan 設定得略大於 minspan，以提供額外的邊界保護，這在資料尾部稀薄時（房地產資料中很常見）很有幫助。

penalty (每個節點的 GCV 懲罰)：

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

這正是 earth 自身的預設值——加性模型為 2，交互模型為 3（更高的懲罰補償了更大的搜索空間）。

newvar.penalty（引入新預測變數的懲罰）：

推薦值：當預測變數之間存在相關性時（例如居住面積與臥室數、浴室數、地塊面積相關）取 0.1。這會使前向過程傾向於重用模型中已有的預測變數，而不是引入相關的替代變數。其機制是：每個新預測變數的 RSS 改進量會乘以 $1/(1 + \text{penalty})$ 。當 **newvar.penalty** = 0.1 時，新變數的 RSS 改進必須比在已有變數上再加一個節點至少高出 10%。這會產生預測變數更少、更簡單的模型，且不影響最終的係數估計（懲罰在選擇完成後即被移除）。若預測變數之間不相關，則保持為 0。注意，earth 不支援在使用個案權重的同時設定非零的 **newvar.penalty**——一旦指定了權重欄，earthUI 就會停用該輸入（強制為 0）。

7.7.2 剪枝參數

pmethod：推薦值：**backward**。預設的基於 GCV 的後向剪枝是確定性的——相同的資料總是產生相同的模型，不依賴任何隨機種子。這對可重現性很重要，在估價工作中尤其如此。cv 方法使用交叉驗證進行剪枝，會引入對種子的依賴。

nprune：推薦值：留空（NULL）。讓 GCV 選擇最優的模型項數。設定 **nprune** 會施加一個硬性上限，凌駕於 GCV 的判斷之上。

7.7.3 交叉驗證參數

nfold（交叉驗證折數）：按樣本量區間推薦：

n （擬合列數）	nfold
0–299	5
300–399	6
400–499	7
500–599	8
600–699	9
700+	10

折數越多，每一折的訓練資料越多（估計偏差更小），但留出的測試折會隨之縮小。該表隨著樣本增大逐級升至標準的 10 折，同時保持每個測試折足夠大（約 50 列以上）以獲得穩定的單折指標；對小樣本則回退到 5 折——此時 CVR^2 估計仍有噪聲，因此應提高 **ncross**（例如 15–20）以平均掉變異數。上限 10 反映的是：每增加一折都是一次完整的 earth 擬合，而診斷收益微乎其微。當 **pmethod** = "backward" 時，交叉驗證不影響模型——它只計算診斷性的 CVR^2 ，並為變異數模型提供殘差。

ncross（交叉驗證重複次數）：按樣本量區間推薦：

n （擬合列數）	ncross
0–199	20
200–399	16
400–699	12
700–1199	10
1200–2999	7
3000+	5

重複可以平均掉折劃分中的運氣成分，從而穩定 CVR^2 ；樣本越大所需重複越少，因為單次交叉驗證本身已經更穩定。重複交叉驗證的殘差還會輸入變異數模型。

7.7.4 變異數模型

varmod.method：推薦值：**lm**——Milborrow 的預設值：用於預測區間的穩健線性變異數模型（離散度是從帶噪聲的殘差中估計的，因此簡單模型的泛化能力更好）。如果 **lm** 在你的資料上無法收斂，可嘗試 **rlm** 或 **const**（同變異數，總能收斂），或退回到 **earth**（也能處理非線性的離散度）。它要求 **nfold** > 1 且 **ncross** > 1。

7.7.5 Earth 輸出檔案

每次成功擬合後，**earthUI** 會將 **<filename>_earth_output_<timestamp>.txt** 寫入輸出資料夾。該檔案包含模型項、匯總統計量（ R^2 、 $GRSq$ 、 $CVRSq$ ）、變異數模型詳情和追蹤日誌。每次擬合建立一個檔案，為比較不同參數配置提供累積記錄。

7.8 設定預設值

第 4 節頂部的選項按鈕控制載入哪一組預設值：

- **Use last settings for input file**（使用該輸入檔案的上次設定）——恢復你上次針對這個特定檔案使用的設定
- **Use default settings**（使用預設設定）——應用你儲存的自訂預設值
- **Earth defaults**（Earth 預設值）——將所有參數重置為原廠值

點選第 4 節底部的 **Save current as default**（將當前設定儲存為預設）以儲存當前的 **earth** 參數和 **Allowed Interactions** 矩陣；第 3 節中對應的按鈕儲存目標變數、評估基準日和預測變數配置。兩者都按專案和用途儲存在專案資料庫中。

8 擬合 Earth 模型

8.1 Fit 按鈕

側邊欄第 5 節包含 **Fit Earth Model** (擬合 Earth 模型) 按鈕，以及一個 **Random seed** (隨機種子) 欄位。該種子使交叉驗證折的分配具有可重現性——下拉列表可調出最近使用過的 10 個種子，擬合完成後所用的種子會記錄在 Earth Output (Earth 輸出) 頁籤中。在擬合之前，earthUI 會驗證您的配置——必須選定一個目標變數，並且至少包含一個預測變數。在估價/市場模式下，緯度和經度欄會被四捨五入，銷售時距則根據估價基準日和合約日期 (若已指定) 計算得出。

8.2 擬合模態框與追蹤輸出

當您點選 Fit 時，會出現一個深色模態覆蓋層，其中包含：

- 標題欄——顯示「Fitting Earth Model」，附帶以秒計數的已用時間計時器，以及一個可取消背景擬合的 **Abort** (中止) 按鈕
- 追蹤日誌——一個可滾動的等寬文字區域，即時顯示 `earth()` 函數的輸出：資料集資訊、前向遍歷進度、交叉驗證折以及完成時間
- 顏色編碼——預設行為綠色，交叉驗證行為黃色，錯誤為紅色

earthUI 使用 `callr::r_bg()` 非同步擬合模型，將計算放在單獨的 R 列程中執行。這使應用程式在長時間擬合期間保持回應。一個 300 毫秒的輪詢觀察器從背景列程讀取輸出並將其流式寫入追蹤日誌。

擬合完成後，會出現「Done in X.Xs」訊息，模態框上會新增一個關閉按鈕 (X)，並移除背景遮罩，以便結果頁籤可以按需渲染。待可見輸出渲染完畢後，模態框會自動關閉——您也可以提前關閉它。

擬合成功時，Fit 按鈕上會附加一個綠色勾號。如果發生錯誤，錯誤訊息會顯示在追蹤日誌中，並彈出一條通知。

8.3 擬合日誌

每次擬合 (無論成功或失敗) 都會在您的輸出資料夾中寫入一個日誌檔：

- 檔案名：`<datafile>_earth_log_<YYYYMMDD_HHMMSS>.txt`
- 內容：時間戳記、擬合過程中的全部追蹤輸出，以及所有錯誤訊息
- 位置：當前活動專案的輸出資料夾

提示：如果未安裝 `callr`，earthUI 會回退到帶簡單進度條的同步擬合。安裝 `callr` 可獲得完整的追蹤輸出體驗。

9 下載資料

擬合完成後，側邊欄第 6 節——在通用模式下標題為「Download Estimated Target Variable(s) & Residuals」(下載估計的目標變數與殘差)，在估價和市場模式下標題為「Download Estimated Sale Prices & Residuals」(下載估計的銷售價格與殘差)——可下載一個包含預測值和診斷資訊的 Excel 檔案。該輸出用於第 7 步 (RCA)，為估價對象 (subject) 指定 CQA (狀況—品質—吸引力) 評分。輸出按 `residual_sf` 和 `cqa_sf` 排序，以幫助您評估估價對象在排名中的位置。如果模型品質良好，那麼各房產應按照未納入迴歸的殘差特徵，從吸引力最低到吸引力最高進行排序。中間值應約為 0，下半部分為負值，上半部分為正值。您應當在排名底部附近發現品質最差的住宅，即「待修房」(fixers)，而最好的住宅位於頂部。通常會有一些異常情形的例外，例如法拍、虧本急售 (short sale)、遺囑認證 (與繼承相關的) 銷售，以及因工作變動或其他原因需要的快速出售。對異常案例的調查通常會揭示出價格異常的相關原因。

按鈕標籤會根據用途模式而變化：

- 通用/市場：**Download Output (Excel)** (下載輸出 (Excel))
- 估價：**Download Intermediate Output (Excel)** (下載中間輸出 (Excel))

檔案名格式為 `<datafile>_modified_<YYYYMMDD_HHMMSS>.xlsx`。

9.1 輸出欄

對於每個目標變數，將追加以下各欄：

欄名	說明
<code>est_<target></code>	模型預測值，例如 <code>est_sale_price</code> (保留 1 位小數)
<code>residual</code>	實際值減去預測值 (1 位小數)
<code>cqa</code>	狀況—品質—吸引力評分 (2 位小數，範圍 0–10)
<code>residual_sf</code>	殘差除以居住面積 (若已指定，1 位小數)
<code>cqa_sf</code>	通過 <code>residual_sf</code> 排名計算得出的 CQA (2 位小數)
<code><variable>_contribution</code>	每個 g 函數的貢獻值 (1 位小數，每個 g 函數一欄)
<code>basis</code>	截距貢獻值，對所有房產相同 (1 位小數)
<code>calc_residual</code>	實際值減去 (<code>basis</code> + 所有貢獻值)——驗證欄 (1 位小數)

對於多目標模型，會新增 `_<i>` 後綴以區分各目標對應的欄。

9.2 欄排序與 Excel 格式

排名欄被放置在輸出檔案的最左側位置，順序為：`residual_sf`、`cqa_sf`、`residual`、`cqa`。這樣便於瀏覽排序後的列表，並評估估價對象在 CQA 分布中的位置。

這些欄應用了以下 Excel 數字格式：

欄名	格式	示例
<code>residual_sf</code>	數值，2 位小數	12.34
<code>cqa_sf</code>	數值，2 位小數	7.25
<code>residual</code>	數值，0 位小數	15,200
<code>cqa</code>	數值，2 位小數	6.80

9.3 CQA 評分

CQA（狀況—品質—吸引力）評分將每一列的殘差與所有其他殘差進行排名比較。對於給定的列，CQA 是帶符號殘差小於該列的列所佔的百分比乘以 10。由此得到一個 0–10 的評分尺度，其中：

- 高 CQA ($\approx 9-10$) ——該房產的成交價遠高於模型預測（較大的正殘差）
- 低 CQA ($\approx 0-1$) ——該房產的成交價遠低於預測（較大的負殘差）
- CQA ≈ 5 ——殘差接近所有殘差的中位數

當指定了 `living_area` 欄時，`cqa_sf` 提供基於每平方英尺殘差的同類排名。

9.4 排序與估價對象列

在估價和市場模式下，可比銷售案例行按 `residual_sf` 降序排序（若未指定居住面積則按 `residual` 排序）。估價對象所在列（第 1 列）保持在第 1 位，不參與排序。在通用模式下，各列按其原始順序匯出。

這種排序方式與最左側的排名欄相結合，使估價師能夠快速瀏覽從預測偏高最多到預測偏低最多的可比銷售案例，並評估估價對象被賦予的 CQA 評分在該分布中的位置。

在呼叫 `predict()` 之前，預測資料中的因子水準會與訓練資料對齊。含有未見過的因子水準的列將產生 NA 預測值。

提示：`calc_residual` 欄應始終等於 `residual`（在捨入誤差範圍內）。如果兩者差異顯著，則表明存在值得調查的計算問題。

9.5 mgcvUI 自動匯出

每次成功擬合且 $\text{degree} \leq 2$ 時，`earthUI` 會自動將完整的結果對象儲存為 `.rds` 檔案，存入專案的輸出資料夾。檔案名遵循如下模式：`<datafile>_earthUI_result_<YYYYMMDD_HHMMSS>.rds`。

該檔案可由 `mgcv` 後處理模組載入——該模組包含在 `earthUI` 套件中，通過 `earthUI::launch_mgcv()` 啟動——使用 `readRDS()` 讀取。GAM 常式使用 `earth` 模型的節點位置和基函數作為 GAM 平滑項的起點，從而實現從 MARS 到 GAM 建模的無縫過渡。

$\text{degree} > 2$ 的模型會被跳過，因為 `mgcvUI` 僅支援兩兩交互。當 `earthUI` 以 Trilogy 模式（從 Trilogy UI）啟動時，額外的 **Lock Model Output**（鎖定模型輸出）區塊可讓您將某次擬合的 `.rds`——連同其預測變數基和共享的 CQA 設定——鎖定為供 `glmnetUI` 和 `mgcvUI` 匯入的規範 `earth` 模型。

9.6 glmnetUI 匯入

為 GAM 常式儲存的同一 `.rds` 檔案也可匯入彈性網絡後處理模組——同樣包含在 `earthUI` 套件中，通過 `earthUI::launch_glmnet()` 啟動。在 `glmnet` 應用中，導航至 **Section 2: Import from earthUI**（第 2 節：從 `earthUI` 匯入），使用 Browse（瀏覽）按鈕選擇該 `.rds` 檔案。

`glmnetUI` 採用標準的 `earth::model.matrix()` 方法：`earth` 模型的基函數（鉸鏈、交互項和線性項）成為 `glmnet` 設計矩陣的欄。`glmnet` 隨後對這些基欄執行正則化迴歸，通過 lasso/彈性網絡對其進行選擇和收縮。這將 `earth` 的自適應基構造與 `glmnet` 的正則化相結合。

9.6.1 主要優勢

- 可解釋的交互項：`earth` 基於鉸鏈的交互項（包括三階項）具有清晰的幾何含義，而 `glmnet` 原生的叉積交互項在法庭或審計場合難以解釋。
- 自動非線性：`earth` 的鉸鏈函數可以捕捉使用原始預測變數的標準 `glmnet` 模型會遺漏的非線性關係。
- 變數選擇：`glmnet` 可以進一步修剪 `earth` 基，剔除沒有貢獻的鉸鏈項。

9.6.2 重要提示：權重欄必須一致

警告

如果 earthUI 使用了權重欄（例如，用 `weight=0` 排除估價對象所在列），則必須在 glmnetUI 的變數配置中指定相同的權重欄。如果權重設定不同，列數將無法對齊，earth 基也就無法附加到 glmnet 的模型矩陣中。

9.6.3 工作流程

1. 在 earthUI 中擬合一個 earth 模型（任意階數）。
2. `.rds` 結果檔案會自動儲存到專案的輸出資料夾。
3. 在 glmnetUI 中，開啟 **Section 2: Import from earthUI**，瀏覽至該 `.rds` 檔案。
4. 確認兩個應用中使用的是同一資料檔案和同一權重欄。
5. 在第 5 節中配置 glmnet 參數（當 earthUI 匯入處於活動狀態時，交互矩陣會自動停用）。
6. 點選 **Fit Model**（擬合模型）。earth 基欄將出現在 Equation（方程）、Coefficients（係數）和 Contributions（貢獻）頁籤中。

10 RCA 計算與下載

Calculate RCA Adjustments & Download (計算 RCA 調整並下載) 按鈕位於側邊欄第 7 節，僅在估價模式下且模型擬合完成後可見。RCA (Residual Constraint Approach, 殘差約束法) 生成相對於估價對象 (subject) 的、源自市場的逐個可比案例調整。

10.1 開啟 RCA 對話方塊

點選該按鈕會開啟一個小型模態對話方塊，其中包含：

- **Score type** (評分類型) —— 選項按鈕，用於選擇 **CQA** 或 **CQA per SF** (僅當已指定 **living_area** 欄時，SF 選項才可用)。如果選擇「CQA per SF」，則根據你為估價對象指定的 CQA 評分，其殘差評分將是其居住面積乘以與給定 CQA_SF 評分匹配的 **residual_sf**。
- **CQA value** (CQA 值) —— 用於輸入估價對象的 CQA 評分 (0.00–10.00)。該評分表示你認為估價對象在可比案例品質分布中所處的位置。
- **Generate** (生成) 按鈕——計算 RCA 輸出並以 Excel 格式下載。在輸入有效評分之前，該按鈕保持停用狀態。

10.2 CQA 評分插值

當你點選 **Generate** 時，earthUI 會根據可比案例的 CQA/殘差對來插值估價對象的殘差：

1. 將可比案例的 CQA 評分和殘差按 CQA 升序排序
2. 通過線性插值 (`stats::approx()`) 將你輸入的 CQA 值映射為一個殘差
3. 如果使用 CQA per SF，則將每平方英尺殘差乘以估價對象的居住面積，換算回總殘差
4. 估價對象的估計價值計算為：模型預測值 + 插值殘差

10.3 輸出欄

RCA Excel 檔案 (<datafile>_adjusted_<YYYYMMDD_HHMMSS>.xlsx) 包含所有中間輸出欄，另加：

欄	說明
subject_value	模型預測值 + 插值殘差 (僅第 1 列)
subject_cqa	你輸入的 CQA 評分 (僅第 1 列)
<variable>_adjustment	估價對象貢獻減去可比案例貢獻 (按 g 函數逐項計算)
residual_adjustment	估價對象殘差減去可比案例殘差
net_adjustments	所有調整之和 (貢獻 + 殘差)
gross_adjustments	所有調整絕對值之和
residual_pct	殘差調整佔可比案例售價的比例
net_adj_pct	淨調整佔可比案例售價的比例
gross_adj_pct	總調整佔可比案例售價的比例
adjusted_sale_price	可比案例售價 + 淨調整

各調整欄告訴你，對於每個可比案例，模型將差異歸因於各變數的金額。**adjusted_sale_price** 是應用所有模型推導調整後的可比案例售價——這組調整後售價應聚集在估價對象的估計價值附近。

10.4 RCA Adjustments 頁籤

RCA 輸出計算完成後，主面板中的 **RCA Adjustments** (RCA 調整) 頁籤會顯示各可比案例的 Residual Adj %、Net Adj % 和 Gross Adj % 分布的直方圖（多目標模型則每個目標一組），從而可以快速直觀地檢查調整的聚集程度。

提示：CQA 評分 5.00 表示估價對象處於可比案例的中位數位置。評分高於 5 表示估價對象品質高於平均水準；低於 5 則表示低於平均水準。

11 銷售比較網格

11.1 概述與用途

銷售比較網格 (Sales Comparison Grid) 是在估價模式下 (側邊欄第 8 節)、計算 RCA 調整之後生成的 Excel 活頁簿。它以結構化的網格形式將估價對象與選定的可比案例並排呈現，並附帶自動計算調整和每個可比案例調整後售價的 Excel 公式。

該網格專為估價師的工作底稿而設計——它將 Earth 模型推匯出的迴歸調整與可編輯儲存格相結合，估價師可在其中將 CQA 殘差分配到具體的房產特徵上。輸出檔案名為 SalesGrid_<YYYYMMDD_HHMMSS>.xlsx。

11.2 可比案例選擇模態框

點選「Generate Sales Grid & Download」(生成銷售網格並下載) 會開啟一個模態對話方塊，供你選擇要包含的可比案例：

- 推薦可比案例已預先勾選。這些可比案例的總調整百分比低於售價的 25%，按銷售時間由近及遠排序 (最近的銷售在前)，上限為 30 個。
- 其他可比案例欄在下方，預設不勾選。它們的總調整達到或超過 25%，但仍可供選擇 (最多顯示 50 個，按總調整百分比升序排序)。
- 最多可選擇 30 個可比案例，最多生成 10 張工作表 (每張工作表 3 個可比案例)。

確認選擇後，所選可比案例按總調整百分比升序排序，因此最相似的可比案例會出現在 Sheet 1 上。

11.3 網格佈局

每張工作表採用 20 欄佈局，容納估價對象加 3 個可比案例。各實體的欄為：

實體	欄 (每個實體 5 欄)
估價對象	標籤、事實值 1、事實值 2、事實值 3、價值貢獻 (VC)
每個可比案例	事實值 1、事實值 2、事實值 3、價值貢獻 (VC)、調整

各列從上到下依次為：

1. 標題——工作表名稱 (例如「Intermediate Sales Comparable Grid — Sheet 1 of 3」)
2. 表頭——「Subject」與各可比案例的欄標題及地址
3. 地址——完整的房產地址
4. APN | MLS# | DOM | Subj.Prox (宗地號 | 掛牌編號 | 在售天數 | 與估價對象的距離)——宗地號、掛牌編號、在售天數，以及與估價對象的 Haversine 距離 (英里)
5. Sales Price | Concess. | Net SP (售價 | 讓價優惠 | 淨售價)——售價、讓價優惠，以及 Net Sale Price (淨售價) 公式
6. Regression Features (迴歸特徵) 表頭列
7. BASE VALUE (基準值)——模型截距
8. Date of Sale | OffMkt | OnMkt (成交日期 | 下市 | 上市)——合約日期、銷售距今時間和 DOM
9. 分組列 (視情況而定)——Location (位置)、Site Size (地塊面積) 和/或 Age (屋齡) 組
10. 模型變數列——每個預測變數一列 (不含已分組的變數)
11. 空白分隔列
12. Residual Features (殘差特徵) 表頭列
13. CQA|Residual (CQA | 殘差)——CQA 評分 + 剩餘殘差公式
14. 殘差特徵列——命名列加空白列，供估價師填寫

15. **Total VC / Net Adjustment** (總 VC / 淨調整)
16. **Net Adjustment %** (淨調整百分比)
17. **Gross Adjustment %** (總調整百分比)
18. **Adjusted Sale Price** (調整後售價) ——公式列
19. 版權頁腳

11.4 售價、讓價優惠與 Net SP

第 5 行為每個可比案例顯示三個值：

- **Sale Price**——來自資料的可比案例售價
- **Concessions**——指定為 **concessions** 的欄中的值 (未指定時為 0)
- **Net SP**——一個 Excel 公式： $\text{Sale Price} - \text{Concessions}$

估價對象欄的售價顯示為「N/A」(因為未知)，若有讓價優惠值則一並顯示。

11.5 分組列 (位置、地塊、屋齡)

當指定了某些特殊類型且這些變數出現在模型中時，earthUI 會建立將相關變數合併的分組列：

Loc: Long | Lat | Area (位置：經度 | 緯度 | 區域) ——當 **longitude**、**latitude** 或 **area** 中的任一變數出現在模型中時出現。顯示每個組成變數的事實值、合併的價值貢獻 (各 VC 之和)，以及針對可比案例的調整 (估價對象合併 VC – 可比案例合併 VC)。VC 儲存格以淺藍色背景標示。

Site Size | Dimensions (地塊面積 | 尺寸) ——當 **lot_size** 或 **site_dimensions** 出現在模型中時出現。結構與 **Location** 組相同。

Actual Age | Effective Age (實際屋齡 | 有效屋齡) ——當 **actual_age** 或 **effective_age** 出現在模型中時出現。結構相同。

被分組列佔用的變數將從下方的單個模型變數列中排除，以防止調整合計中的重複計算。

11.6 模型變數列

在分組列下方 (若無分組列，則直接在 **BASE VALUE** 下方)，每個剩餘的模型預測變數佔一列，顯示：

- **事實值**——估價對象和每個可比案例的實際資料值
- **價值貢獻 (VC)** ——該列 **g** 函數對預測值的貢獻
- **調整 (僅可比案例)** ——估價對象 VC 減去可比案例 VC

11.7 CQA|Residual 與殘差特徵列

CQA|Residual 列顯示每個房產的 CQA 評分，並包含剩餘殘差公式——即尚未分配到具體特徵的殘差部分。公式為：

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

其下是殘差特徵列——5 個命名列 (**Location**、**View/Appeal**、**Condition**、**Quality**、**Other**，即位置、景觀/吸引力、狀況、品質、其他) 加 6 個空白列。這些是輸入儲存格，供估價師為模型未捕捉到的特徵填寫價值貢獻。隨著估價師填入數值，**Remaining Residual** 公式的值會自動減少。

對於每個可比案例，調整欄包含一個公式：估價對象特徵 VC – 可比案例特徵 VC。

11.8 調整後售價公式

Adjusted Sale Price 列包含 Excel 公式：

- 估價對象：從 BASE VALUE 到最後一個殘差特徵列的所有價值貢獻儲存格之和。這代表模型對估價對象的總預測值，加上估價師分配的殘差特徵。
- 可比案例：Net SP + 從第一個分組列或變數列到最後一個殘差特徵列的所有調整儲存格之和。這給出應用所有迴歸推導調整和估價師填寫調整後的可比案例售價。

11.9 工作表保護與可編輯儲存格

每張工作表均已受保護，以防止公式和資料被意外修改。保護鎖定了：

- 所有公式儲存格（Net SP、剩餘殘差、調整、調整後售價等）
- 所有資料儲存格（地址、售價、事實值、來自模型的價值貢獻）
- 標籤和表頭

唯一未鎖定（可編輯）的儲存格是殘差特徵價值貢獻輸入格——即 CQA|Residual 列下方供估價師填寫分解值的儲存格。這些儲存格以淺黃色背景標示，表示可編輯。

11.10 在 Excel 中使用網格

1. 開啟檔案——使用 Excel 或相容的試算表應用程式。
2. 審核迴歸調整——模型推導的調整已預先填入並鎖定。
3. 分配殘差——在殘差特徵列（黃色儲存格）中，填寫你對剩餘殘差中歸因於各特徵（位置、景觀、狀況、品質等）的金額的判斷。隨著分配的進行，Remaining Residual 公式的值會相應減少。
4. 檢查 **Adjusted Sale Price**——該公式會隨你填入殘差特徵值而自動更新。
5. 多張工作表——如果選擇了 3 個以上的可比案例，請在工作表之間切換。每張工作表的左欄均為同一估價對象，並配以 3 個不同的可比案例。

提示：目標是將 CQA 殘差分配到各特徵列，直至 *Remaining Residual* 接近於零。這表明你能夠解釋每個可比案例的迴歸估計值與其實際售價之間的差異。

12 下載報告

模型擬合完成後，報告生成是一個兩步式的 Quarto 工作流：**Generate Quarto Report**（生成 Quarto 報告，估價模式下為側邊欄第 9 節，其他模式下為第 7 節）寫出一個自包含的報告包，而 **Convert Quarto Report**（轉換 Quarto 報告，第 10 節）將其——或任何其他 Quarto 檔案——渲染為最終格式。

12.1 生成 Quarto 報告

點選 **Generate Quarto Report** 會在專案輸出資料夾下的 `<base>_qmd/` 中寫出一個自包含的 Quarto 包，其中包含已填充內容的 `.qmd` 原始檔、所有預生成的圖表資源（PNG 和 PDF）、報告資料，以及一個 Word 參考模板。此步驟不進行任何渲染——該報告包可以手工編輯、通過 Quarto includes 與其他 Quarto 原始檔合併，或使用 `quarto preview` 單獨預覽。

12.2 轉換 Quarto 報告

第 10 節提供一個檔案選擇器（預設指向當前專案最近生成的報告包——但你可以瀏覽到任意 `.qmd` 檔案）以及格式核取方塊：

- **HTML**——一個 `.html` 檔案，帶有 KaTeX 數學公式渲染、內嵌圖片和目錄。使用 Flatly Bootstrap 主題和 Roboto Condensed 字體。
- **Word**——一個適合編輯和分發的 `.docx` 檔案。包含目錄、頁碼，並使用自訂參考模板以保持樣式一致。
- **PDF**——使用 LuaLaTeX 排版，輸出達到專業品質。紙張尺寸遵循區域設定（Letter 或 A4）。包含用於大型交互矩陣的橫向頁面，並使用 Roboto Condensed 搭配 Latin Modern Math 排版公式。僅當檢測到 LaTeX 發行版時才會提供 PDF 選項。

點選 **Convert**（轉換）以渲染所選格式。渲染在背景執行——一個模態對話方塊顯示已用時間和 Quarto 進度，應用保持可回應狀態。完成後會有通知確認檔案位置。

所有操作（模型擬合、資料下載、RCA 計算、銷售網格生成和報告渲染）都會連同起訖時間戳記和耗時記錄到輸出資料夾中的 `<datafile>_earthui_log.txt`，以便排查問題和監控效能。

12.3 報告內容

每份生成的報告都包含以下各節：

1. 資料集描述——觀測數量、目標變數、所用預測變數、分類預測變數
2. 模型規格——階數（degree）、交叉驗證狀態、項數、模型中的預測變數
3. 允許的交互項（僅當 $\text{degree} \geq 2$ 時）——帶勾選標記的交互矩陣，按輸出類型格式化
4. 結果摘要—— R^2 、GRSq、GCV、RSS，以及 $CV R^2$ （若啟用了交叉驗證）
5. 模型方程——以 LaTeX 記號表示的完整 Earth 模型方程
6. 係數與基函數——所有項、其係數以及定義它們的基函數的表格
7. 變數重要性——預測變數重要性得分的長條圖和排名表
8. g 函數貢獻——每個 g 函數的圖：單變數項為折線圖，雙變數交互項為 3D 透視圖和等高線圖
9. 相關矩陣——預測變數相關性的熱圖
10. 診斷——殘差對擬合值（Residuals vs Fitted）、常態 Q-Q 圖，以及實際值對預測值（Actual vs Predicted）圖
11. ANOVA 分解——顯示每個基函數對 RSS 貢獻的表格
12. Earth 輸出——來自 `earth::summary()` 的原始文字輸出，包括剪枝過程的細節

對於多反應模型，第 4–8 節和第 10 節會針對每個目標變數重複一次。

提示：生成報告需要 *quarto R* 套件和 *Quarto* 安裝。若要輸出 *PDF*，還必須安裝 *LaTeX* 發行版（例如通過 *tinytex::install_tinytex()*）——否則 *PDF* 核取方塊會被隱藏。

13 後處理模組：彈性網絡與 GAM

earthUI 套件附帶兩個後處理模組——一個基於 `glmnet` 的彈性網絡 (elastic net) 常式和一個基於 `mgcv` 的廣義可加模型 (GAM) 常式。本期各有一篇文章對它們分別做了深入論述；本節說明它們在 earthUI 工作流中的用途，以及三個常式如何協作。

13.1 為什麼需要輔助擬合？

earth 模型給出價值結論。後處理模組的存在是為了印證它：由方法學上不同的引擎、針對同一估價對象 (subject)、基於相同專案資料產生的兩個額外估計。當對 earth 基函數進行彈性網絡重估計與獨立的 GAM 平滑擬合得出實質上相同的價值時，估價師的結論便不再僅僅依賴於單一模型的假設——這在複核和訴訟場景中是一項實實在在的優勢。當它們不一致時，這種分歧本身就是診斷資訊，通常指向資料稀薄、某個過度靈活的項，或某個值得重新審視的離群可比銷售案例。兩個模組都不凌駕於 earth 模型之上；它們是佐證證據，並有意不做加權呈現。

13.2 啟動與共享狀態

每個常式都是一個獨立的本機 Shiny 應用：

- `earthUI::launch_glmnet()`——彈性網絡應用，連接埠 7879。
- `earthUI::launch_mgcv()`——GAM 應用，連接埠 7880。

三個應用讀取同一專案樹和同一設定資料庫，因此你在 earthUI 中開啟的專案在模組中立即可用——連同 earthUI 為其儲存的沿用欄位：Response (反應/目標變數)、Effective Date (估價基準日)、RCA 估價對象的 CQA 類型和取值，以及居住面積 (living-area) 指定。這些無需重新輸入；模組會讀取它們，使同樣的取值貫穿全部三次擬合。

13.3 彈性網絡常式

`glmnet` 常式將 earth 模型的基函數——其鉸鏈函數、交互項和線性項——作為設計矩陣的欄匯入，並在套索迴歸 (lasso) / 彈性網絡正則化下對其重新估計 (其機制及權重欄的注意事項參見上文「`glmnetUI Import`」)。其結果保留了 earth 可解釋、具有幾何意義的項，同時讓正則化器收縮或剔除那些不能證明自身價值的項。完整論述參見本期的 `glmnetUI` 文章。

13.4 GAM 常式

`mgcv` 常式走的是另一個方向：它不複用 earth 的基函數，而是重新擬合帶懲罰的平滑項，以 earth 模型的節點位置作為起點 (參見上文「`mgcvUI Auto-Export`」)。由於 GAM 在不同的平滑範式下重新推導函數形式，它與 earth 模型的一致性真正獨立的印證，而非回聲。支援 $\text{degree} \leq 2$ 的 earth 模型。完整論述參見本期的 `mgcvUI` 文章。

13.5 一個實用的工作流

1. 先在 earthUI 中擬合并打磨 earth 模型，直到你對其滿意為止。每次成功的擬合 ($\text{degree} \leq 2$) 都會自動匯出供模組使用的 `.rds` 檔案。
2. 針對同一專案開啟 `glmnet` 和 `mgcv` 常式，匯入 earth 結果並進行擬合。一次只在一個常式中工作——擬合耗時較長，且各常式通過共享專案進行協調，因此合理的節奏是順序進行，而非並行。
3. 自由迭代。在你於各應用之間切換時，每個應用都會保留其結果，也沒有任何環節強制你在過程中生成報告。
4. 只有當三個結果都令你滿意時，再回到各個常式生成其 Quarto 報告。每種方法都會將其擬合和價值結論登記到專案的三部曲 (trilogy) 記錄中，使三份報告可追溯到同一次比較執行。

提示：這些模組在任何層面上都是可選的：*earthUI* 自身的工作流——擬合、*RCA*、銷售網格、報告——沒有它們也是完整的。當某個結論需要印證時再使用它們：棘手的估價對象、資料稀薄的市場細分，或很可能面臨複核的委託任務。

14 示範資料集：Appraisal_1.csv

earthUI 包含一個示範 MLS 資料集，用於探索估價工作流。可通過編程方式載入：

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

或者，在 Shiny 應用中，將其複製到專案的 in/ 資料夾（或在建立專案時將其作為初始資料檔案附加），然後在第 2 節（Import Data，匯入資料）中選擇它。

14.1 描述

該檔案包含來自一份模擬 MLS 匯出的 1,501 條住宅銷售記錄，外加第 1 列的 1 個估價對象 (subject)（共 1,502 列）。資料代表一個多區域市場中的獨棟住宅銷售，涵蓋多種房屋面積、屋齡和位置。

這不是真實資料，但基於北加州一個真實感很強的社區。所有識別資訊均已被更改或刪除。

14.2 資料欄

欄名	類型	特殊類型	描述
weight	numeric	—	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 房產識別碼
listing_id	character	display_only	MLS 掛牌編號
parcel_number	character	display_only	縣稅務評估員宗地號 (APN)
street_address	character	display_only	房產地址
city_name	character	display_only	城市
postal_code	character	display_only	郵遞區號 (ZIP)
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	—	從合約日期到估價基準日的天數 (預先計算)
coe_date	Date	display_only	過戶完成 (close of escrow) 日期
listing_status	character	display_only	掛牌狀態 (例如 "Sold")
sale_price	numeric	(target)	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	原始掛牌價格
living_sqft	numeric	living_area	總居住面積 (平方英尺)
beds_total	integer	—	臥室數量
baths_total	numeric	—	浴室總數 (例如 2.5 = 2 個全衛 + 1 個半衛)
lot_size	numeric	lot_size	地塊面積 (平方英尺)
area_id	integer	area	MLS 區域識別碼
area_text	character	display_only	區域名稱
age	numeric	actual_age	屋齡 (年)
year_built	integer	display_only	建成年份
latitude	numeric	latitude	緯度 (建模時四捨五入到小數點後 3 位)
longitude	numeric	longitude	經度 (建模時四捨五入到小數點後 3 位)
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	—	壁爐數量
no_of_stories	numeric	—	樓層數
style	character	—	建築風格
view	character	—	景觀類型 (例如 "Neighborhood"、"Hills")
days_on_market	integer	dom	在市天數
listing_date	Date	listing_date	掛牌日期
sale_concessions	numeric	concessions	賣方讓步 (concessions) 金額

14.3 建議的快速上手步驟

1. 啟動 earthUI：`earthUI::launch()`
2. 在 New Project（新增專案）精靈中建立一個 Purpose（用途）為 **For Appraisal**（用於估價）的專案，並附加 `Appraisal_1.csv` 作為初始資料檔案
3. 在第 2 節（Import Data）中選擇 `Appraisal_1.csv`
4. 選擇 `sale_price` 作為目標變數
5. 按上表所示指定特殊類型
6. 納入預測變數：`sale_age`、`living_sqft`、`baths_total`、`lot_size`、`area_id`（作為因子）、`age`、`latitude`、`longitude`、`garage_spaces`
7. 將 `degree` 設為 1，點選 **Fit Earth Model**（擬合 Earth 模型）
8. 下載中間輸出（第 6 步），審閱 CQA 排名
9. 計算 RCA 調整（第 7 步），CQA 得分約為 ~5.00
10. 生成銷售比較網格（第 8 步）

提示：在匯入你自己的 *MLS* 資料之前，示範資料集是探索 *earthUI* 全部功能的好方法。上面建議的預測變數通常能得到 R^2 高於 0.85 的模型。

15 實驗性功能：Prolog 文字提取與推導規則

0.10.0 版本引入了一項可選且確屬實驗性的功能，用於從 MLS 匯出資料的描述性文字欄中提取可供模型使用的特徵——首要的是「Public Remarks」（公開備註）欄，此外還包括經紀人備註以及以逗號分隔的特徵列表，例如房間特徵或既有建築物。提取引擎是 Prolog，具體而言是確定子句文法（**Definite Clause Grammars, DCG**）：在備註（remarks）文字中識別諸如「move in ready」「3 car garage」或「owned solar」等短語的規則，均以 Prolog 文法規則的形式編寫。

該功能仍在開發之中。目前它可以部分用於提取——隨附的文法已能識別一組實用的特徵——但要將其適配到您所在 MLS 的用語習慣，您需要能夠閱讀並編寫 Prolog，尤其是 DCG。如果這不適合您，可以完全跳過本章：該功能預設關閉，關閉時 earthUI 的其餘部分工作方式完全不變。

Warning

這是一項處於積極開發中的實驗性功能。其提取文法既不完整，也未經過面向實際估價工作的驗證——在將提取出的欄用於模型之前，必須將其與原始文字核對。要有效使用該功能，需要具備 Prolog 和 DCG 的實際工作知識。它所做的一切都是附加且可逆的；如果出現任何異常，只需關閉 Settings 中的開關並照常繼續即可。

15.1 它的作用

MLS 房源在自由文字和分隔列表欄位中承載著大量與價值相關的資訊。這一可選步驟通過兩種特殊的欄角色（見第 6 章），將這些文字轉化為迴歸可以使用的普通欄：

- **remarks** 欄由 Prolog DCG 文法解析，每個提取出的屬性生成一個特徵欄。欄前綴由源欄名稱各單詞的首字母構成：public_remarks 生成 pr_* 欄，agent_remark 生成 ar_* 欄，依此類推。
- **list** 欄（例如 existing_structures = "Shed, Workshop"）被拆分為獨熱編碼的 <col>_<value> TRUE/FALSE 欄。這一部分是純 R 實作，完全不需要 Prolog。

隨附的文法（inst/prolog/mls_remarks.pl）目前可提取的內容包括：帶有 1–5 級 condition_score 的分級狀況等級；景觀標誌與景觀類型；廚房與衛浴翻新標誌；硬木及其他地板材質；升級項目列表及數量；車庫車位數與車庫面積；泳池、水療池及其他室外設施；壁爐數量；附屬建築（棚屋、穀倉、工作間、casita、客房、犬舍等）及從「40x50」這類尺寸中解析出的面積；房車停放與儲物空間；太陽能自有與租賃之分、電池儲能；地塊面積（含英畝到平方英尺的換算）；與主體相關聯的交易背景價格金額（降價、HOA 會費、太陽能租賃或買斷、Mello-Roos）；區位線索；以及取值範圍為 [-1, 1] 的亮點 / 顧慮情感評分。

15.2 環境要求：vProlog (Trealla)，而非 SWI-Prolog

備註解析執行於 **vProlog** R 套件之上——一個自包含的 Trealla Prolog 引擎。它不在 CRAN 上；請按照本文開頭「Installing earthUI and vProlog」（安裝 earthUI 與 vProlog）一節所示，從作者的 r-universe 儲存庫安裝（提供 Windows 和 macOS 的預編譯二進位套件，原始碼位於 github.com/wcraytor/vProlog）。無需也不會使用任何系統級 Prolog 安裝（無論是 SWI-Prolog 還是其他實作）。如果缺少 vProlog，備註解析會被跳過並給出提示，而 list 欄拆分仍照常執行。輔助函數 vProlog_available() 可報告引擎與隨附文法是否均已就緒。

15.3 啟用與執行

1. 開啟 Settings（設定，齒輪圖示），勾選 **Enable Prolog / list column processing**（啟用 Prolog / 列表欄處理）。該開關預設關閉，附有「進階功能——面向具備 Prolog 專

業知識的使用者」的說明，按專案記憶設定，並顯示在 Project Information（專案資訊）對話方塊中。

2. 在預測變數表中，將相關欄的 **Special**（特殊）角色設為 **remarks** 或 **list**（每種角色都可以指定給多個欄）。
3. 使用位於側邊欄第 4 節與第 5 節之間的 **Execute Prolog Processing**（執行 Prolog 處理）區域。它提供兩個步驟：

1. **Expand lists & remarks**（展開列表與備註）——對每個 **remarks** 欄執行 DCG 文法，並對每個 **list** 欄進行獨熱拆分。展開是增量式的：只處理自上次執行以來新指定的欄，一旦所有已指定的欄都已展開，按鈕即被停用。解析過程由進度條追蹤。

2. **Execute derivation rules**（執行推導規則）——在展開後的欄上應用您自己的 Prolog **derive/2** 規則（見下文）。規則在執行前會經過語法檢查和試執行驗證。

15.4 推導規則

對於每一資料列，earthUI 將該列的各欄斷言為 **cell(Name, Value)** 事實，並收集您的 **derive(Col, Val)** 規則的所有解——每個不同的 **Col** 都成為一個新欄，且每欄以找到的第一個值為準，因此可以在後面附加一條後備子句。輔助謂詞（**count_true/2**、**count_false/2**、**count_true_prefix/2**、**cell_or/3**、**is_na/1**）支援聚合與 NA 處理，而 **drop/1** 事實會在規則執行後丟棄相應欄。例如：

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

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

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

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

規則存放在每個專案獨立的 **<project>_rules.pl** 檔案中，該檔案由一份帶完整註釋的模板初始化，並通過 **Edit derivation rules...**（編輯推導規則……）在一個可浮動、可拖動、也可停靠到側邊欄的程式碼編輯器中編輯。**Check**（檢查）按鈕可在不執行規則的情況下對其進行驗證（語法掃描、載入檢查，以及針對資料第 1 列的試執行）。

15.5 帶版本號的活動資料檔案

任一步驟完成後，增補後的資料框都會以 **<original_name>_<YYMMDD_HHMMSS>.xlsx** 的形式寫入專案的輸入資料夾，並成為該專案的活動資料檔案——重新開啟專案時載入的即是增補後的資料，因此展開結果永遠不會被重複計算，而 **glmnetUI** 和 **mgcvUI** 使用的也正是同一個檔案。原始的備註欄會被自動排除在模型之外；生成的欄則不會被自動納入預測變數。文字解析可能產生大量欄，因此請將您（暫時）不希望進入迴歸的欄標記為 **Latent**（潛在），或用 **×** 將其停用（見第 6 章）。

15.6 編程介面與隨附檔案

整個流程均已匯出，可供批次處理使用：**execute_prolog_processing(df, remarks_cols, list_cols, rules_text)** 負責編排各個處理階段，其底層基於 **remarks_to_columns()**、

`split_list_column()`、`apply_user_rules()` 和 `validate_rules()`。位於 `inst/prolog/` 下隨附的 Prolog 原始檔包括：`mls_remarks.pl`（DCG 文法與特徵聚合）、`mls_analyze.pl`（用於批量解析的檔案 IO 驅動）、`mls_rules.pl`（可供推導規則使用的輔助謂詞），以及 `climate_regions_ca.pl`（見下文）。

15.7 相關實驗：氣候區先驗

一項配套的實驗性功能針對資料稀薄的特徵——即那些因僅出現在極少數可比銷售案例中而無法由迴歸估計的特徵（例如蒸發式冷風機）。三個 R 輔助函數——`climate_region_for()`、`market_area_profile()` 和 `climate_feature_priors()`——將專案所在位置（縣+城市）歸入七個購房者認知層面的加利福尼亞氣候區之一（這是對加州能源委員會 16 個建築氣候區的實用性歸並），並按區域返回 HVAC 與製冷特徵（供暖、空調、蒸發式冷風機、全屋換氣扇、空氣過濾）的序數化貢獻價值先驗。Prolog 模組 `climate_regions_ca.pl` 鏡像了同一套分類，可在推導規則內部使用；兩側均由同一個規範的 R 源生成，因此不會產生偏差。這些先驗旨在為特徵估值提供參考，並作為對估計係數的合理性核查——區域歸屬（尤其是橫跨多個區域的縣）應由估價師複核。

15.7.1 超越加利福尼亞：轄區包與市場區域模組

加利福尼亞是第一個轄區，而非邊界。`earthUI` 在國際範圍內被使用，氣候先驗知識正被重構為轄區包（**jurisdiction packs**）：自包含、資料化的事實表（分區定義、行政單元歸屬、特徵先驗），以專案系統的國家代碼和行政層級代碼為鍵，因此一個包可以描述一個美國的州、一個德國的聯邦州（Land），乃至直接描述一個小國。與之配套的市場區域模組（**market-area modules**）層則捕捉某一具名市場區域的在地用語與更精細的先驗。

二者的設計目標都是由 Valuation Engineer 和估價師在一線現場、用他們自己的語言編寫——並且，從現實出發，可借助 LLM 完成：世界上並沒有多少 Prolog 程式設計師，因此現場產出的成果被有意設計為表格而非程式。包格式是一個有說明文件的模式（schema），並附有完整的成品示例（即加利福尼亞本身），其明確目的就是讓 Claude 這樣的助手能夠根據對某一市場的自然語言描述起草出一個有效的包；`earthUI` 會從包生成底層的 Prolog，並且一個具有阻斷性的驗證器（模式、覆蓋度、引用完整性和試執行檢查）必須通過後，包才能被啟用。每個現場編寫的包都必須攜帶出處資訊——署名作者、日期以及所引用的權威依據——因此為這些知識背書的是負有責任的專業人員而非起草工具，報告也能準確引用所使用的包及其版本。

提示：一個實用的模式：先展開備註，再用 `derive/2` 規則將細粒度的 `pr_*` 標誌整合為少數幾個派生的計數或布林欄，用 `drop/1` 丟棄已整合的原始標誌，只保留派生欄作為候選預測變數。這樣可以保持預測變數列表精簡，並使共線性可控。

參考文獻

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