

earthUI: earth (MARS) パッケージのための対話型インターフェース

日本語訳。人工知能の支援を得て翻訳し、編集部が校閲した。英語版が正本である。

William Bert Craytor

第 1 巻第 1 号 · 2026 年夏号

発行日: 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.

本論文はオープンアクセス論文であり、[クリエイティブ・コモンズ表示 4.0 国際ライセンス \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/) の条件に従って配布される。

要旨. earthUI は、多変量適応回帰スプライン (MARS, multivariate adaptive regression splines) を実装した R `earth` パッケージのためのグラフィカルユーザインターフェース (graphical user interface) である。一般的な予測モデリング、不動産の鑑定評価、市場地域分析という三つの用途モードを提供し、データの取り込み、モデルの構成、当てはめ、そして `earth` の診断情報とグラフィック出力の解釈に至るまで利用者を導く。本論文は、earthUI のデータ形式要件、モデリングの作業手順、出力画面、および全機能のリファレンスを文書化する。

キーワード: MARS, 多変量適応回帰スプライン, `earth`, R, グラフィカルユーザインターフェース, 不動産評価

発行者注記 (2026 年 7 月): 本論文は earthUI 0.10.0 を対象としている。印刷直前に、earthUI 0.11.0 が姉妹アプリケーションである `glmnetUI` と `mgcvUI` (続く二編の論文で解説する) を統合し、MARS、エラスティックネット、GAM の作業手順を単一のアプリケーションにまとめた。ここで解説するインターフェースと作業手順は、この統合による変更を受けない。

目次

1	序論	4
1.1	earthUI とは何か	4
1.2	一つのパッケージ、三つのルーチン	4
1.3	三つの用途モード	4
1.4	不動産に固有の機能	5
1.5	earthUI と vProlog のインストール	5
1.6	使い始め方	6
2	MLS 入力データの要件	7
2.1	ファイル形式と構造	7
2.2	鑑定評価モードで必要となる列	7
2.3	特殊列の命名規約	8
2.4	データの品質と完全性	8
2.5	対象不動産の行の配置	9
3	一般用途モード	10
3.1	概要	10
3.2	サイドバーの作業手順	10
3.3	メインパネルのタブ	10
3.4	設定の永続化	11

3.5	ダークモード	11
3.6	ロケールと地域設定	11
3.6.1	対応国	12
3.6.2	用紙の上書き	12
3.6.3	既定値の保存	12
4	鑑定評価モード	14
4.1	対象不動産の行の取扱い	14
4.2	基準日と経過日数	14
4.3	特殊列の指定	14
4.4	残差寄与（希少特性の推定）	15
4.5	RCA 補正の概要	15
5	市場地域分析モード	17
5.1	鑑定評価モードとの相違	17
5.2	市場モードを用いるべき場面	17
6	変数の選択	18
6.1	目的変数	18
6.2	予測変数表	18
6.3	データ型の検出と上書き	18
6.4	Factor フラグと Linear フラグ	19
6.5	特殊列型リファレンス	19
6.6	複数の目的変数	20
7	パラメータの選択	21
7.1	基本パラメータ	21
7.2	前進パス	21
7.3	許容する交互作用	22
7.4	枝刈り	22
7.5	交差検証	22
7.6	部分集合フィルタ	23
7.7	パラメータの推奨値	23
7.7.1	前進パスのパラメータ	23
7.7.2	枝刈りのパラメータ	24
7.7.3	交差検証のパラメータ	24
7.7.4	分散モデル	25
7.7.5	Earth の出力ファイル	25
7.8	設定の既定値	25
8	Earth モデルの当てはめ	27
8.1	Fit ボタン	27
8.2	当てはめのモーダルとトレース出力	27
8.3	当てはめログ	27
9	データのダウンロード	28
9.1	出力の列	28
9.2	列の順序と Excel の書式	28
9.3	CQA の評点	29
9.4	並べ替えと対象不動産の行	29
9.5	mgcvUI への自動書き出し	29
9.6	glmnetUI への取り込み	30

9.6.1	主な利点	30
9.6.2	重要: 重みの列が一致していなければならない	30
9.6.3	作業手順	30
10	RCA の計算とダウンロード	31
10.1	RCA ダイアログを開く	31
10.2	CQA の評点の内挿	31
10.3	出力の列	31
10.4	RCA Adjustments タブ	32
11	取引事例比較表	33
11.1	概要と目的	33
11.2	取引事例選択のモジュール	33
11.3	表の配置	33
11.4	売買価格、譲歩額、純売買価格	34
11.5	まとめられた行 (位置、敷地、築年数)	34
11.6	モデル変数の行	34
11.7	CQA Residual の行と残差特性の行	34
11.8	補正後売買価格の数式	35
11.9	シートの保護と編集可能なセル	35
11.10	Excel での表の扱い方	35
12	レポートのダウンロード	36
12.1	Quarto レポートの生成	36
12.2	Quarto レポートの変換	36
12.3	レポートの内容	36
13	後処理モジュール: エラスティックネットと GAM	38
13.1	なぜ補強のための当てはめを行うのか	38
13.2	起動と共有される状態	38
13.3	エラスティックネットのルーチン	38
13.4	GAM のルーチン	38
13.5	実務上の作業手順	39
14	デモデータセット: Appraisal_1.csv	40
14.1	記述	40
14.2	列	41
14.3	推奨する手早い始め方	41
15	実験的機能: Prolog によるテキスト抽出と導出規則	43
15.1	何をするのか	43
15.2	要件: SWI-Prolog ではなく vProlog (Trealla)	43
15.3	有効化と実行	44
15.4	導出規則	44
15.5	版を付した有効データファイル	45
15.6	プログラムからの利用と同梱ファイル	45
15.7	関連する実験: 気候地域の事前情報	45
15.7.1	カリフォルニアを超えて: 法域パックと市場地域モジュール	45

1 序論

1.1 earthUI とは何か

earthUI は、多変量適応回帰スプライン (MARS, multivariate adaptive regression splines) を実装した R `earth` パッケージのためのグラフィカルユーザーインターフェースである (Friedman 1991)。モデリングエンジンは Stephen Milborrow の `earth` パッケージであり (Milborrow 2024)、Milborrow (n.d.[a]) および Milborrow (n.d.[b]) に詳細に文書化されている。earthUI はローカルの Shiny アプリケーションとして動作し、ログインもサーバもアカウントも存在しない。R から起動し、データセット (CSV または Excel) を取り込み、モデルを構成して、対話的に当てはめを行う。

本アプリケーションは、データの取り込み、変数の構成、モデルの当てはめ、診断プロット、変数重要度、モデル方程式、そして HTML、Word、PDF 形式でダウンロードできるレポートに至る、完全な作業手順を提供する。

1.2 一つのパッケージ、三つのルーチン

現行リリースの時点で、earthUI パッケージは二つの後処理モジュール (Post Processing Module) をも内包している。これらは以前のリリースでは `glmnetUI` と `mgcvUI` という別個のパッケージであった。earthUI をインストールすれば三つのルーチンがすべて導入され、それぞれが独自のローカル Shiny アプリケーションとして動作する。

- `earthUI::launch()` — 本論文で解説する `earth` (MARS) アプリケーションであり、ポート 7878 を使用する。`earth()` が中心的な評価エンジンである。
- `earthUI::launch_glmnet()` — エラスティックネットの後処理モジュール (ポート 7879) であり、`earth` モデルの基底関数を取り込み、ラッソ/エラスティックネット正則化のもとで再推定する (後述の “`glmnetUI Import`” を参照)。
- `earthUI::launch_mgcv()` — GAM の後処理モジュール (ポート 7880) であり、`earth` モデルのノット位置を `mgcv` の平滑化項の出発点として用いる (後述の “`mgcvUI Auto-Export`” を参照)。

三つのルーチンは、一つのプロジェクトツリー、一つの設定データベース、そして支援基盤の一つの複製を共有する。したがって、あるアプリケーションで開いたプロジェクトは、ただちに他のアプリケーションでも利用できる。`glmnet` ルーチンと `mgcv` ルーチンは、競合する作業手順ではなく、`earth` モデルの価値結論を裏づける証拠である。意図されている進め方は、まず `earth` モデルを当てはめて洗練させ、その後、裏づけが必要になったときに同じプロジェクトに対してモジュールを実行するというものである。

1.3 三つの用途モード

earthUI は作業をプロジェクトという単位で組織する。プロジェクトとは、入力ファイル、出力物、設定を固定された場所にまとめて保持する、名前を与えられた実体である。すべてのプロジェクトは **Purpose** (用途) をもち、これはプロジェクト作成時に新規プロジェクトウィザードで選択する。モードは三つである。

- **General** (一般) — あらゆる種類の母集団またはデータセットに対する Earth 回帰。既定のモードである。領域固有の追加機能をもたない、完全な MARS モデリングの作業手順を提供する。
- **For Appraisal** (鑑定評価用) — 不動産の鑑定評価に合わせた Earth 回帰。対象不動産の取扱い、特殊列の指定、残差制約アプローチ (RCA, Residual Constraint Approach)、取引事例比較表の生成など、単一不動産の評価に固有の機能を追加する。
- **Market Area Analysis** (市場地域分析) — 市場地域の研究に合わせた Earth 回帰。単一の対象不動産を評価するのではなく、画定された市場内の不動産群を分析するための機能を追加する。

三つのモードのいずれにおいても、中核のモデリングエンジンは同一であり、利用者は常に Earth (MARS) モデルを当てはめている。用途の設定は、どの追加ツールとインターフェース要素が利用可能になるかを制御するにすぎない。

用途はプロジェクトごとに固定される。プロジェクトの用途は、その国、州、郡、市とともに、プロジェクト作成の時点で確定する。別のプロジェクトに切り替えると、earthUI は清浄な既定状態に戻り（取り込んだデータ、モデル結果、タブ、変数構成、earth のパラメータはすべて消去される）、続いてそのプロジェクトの用途について保存された構成を復元する。

1.4 不動産に固有の機能

For Appraisal または **Market Area Analysis** を選択すると、earthUI は不動産分析のために設計されたいくつかの機能を有効にする。

- **特殊列の指定** — 各予測変数には、contract_date、dom、concessions、latitude、longitude、living_area、lot_size、actual_age、effective_age、area、site_dimensions、display_only といった特殊な役割を付与できる。これらの指定は、当てはめ、出力、取引事例比較表の生成において当該列がどのように扱われるかを制御する。完全なリファレンスは第 6 章を参照せよ。
- **緯度と経度の丸め** — 緯度または経度として指定された列は、過適合を防ぐために自動的に小数点以下 3 桁に丸められる。
- **Sale Age 列** — ある列が contract_date として指定され、かつ基準日 (Effective Date) が与えられている場合、earthUI は sale_age 列（売買日と基準日との間の日数）を計算し、これを予測変数として代置する。
- **RCA 計算（鑑定評価モード専用）** — 鑑定評価モードでは、モデルを当てはめた後に earthUI が残差制約アプローチ (RCA) の出力を計算できる。これは取引事例ごとの補正、純補正／総補正の要約、および対象不動産の補正後売買価格を生成する。
- **取引事例比較表（鑑定評価モード専用）** — 取引事例の補正、残差特性の内訳、補正後売買価格に関する数式を備えた複数シートの Excel ワークブックを生成する。第 11 章を参照せよ。

ヒント: *General* 用途モードは、金融、科学、工学、不動産のいずれのデータセットにも通用する。鑑定評価モードと市場モードは、不動産の専門家のための便宜的な機能を加えるにすぎない。

1.5 earthUI と vProlog のインストール

earthUI は CRAN で公開されているので、標準的なインストールは R の一行で済む。

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

本論文は現行リリースを解説している。読者がこれを読む時点で CRAN がなお旧版を配布している場合には、開発版を GitHub からインストールできる（subdir に注意されたい。パッケージはリポジトリの pkg/ フォルダに置かれている）。

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

vProlog (任意)。実験的な備考欄処理機能（本論文の末尾近くの “Prolog Text Extraction” を参照）は、自己完結型の組込み Prolog エンジンである vProlog パッケージを用いる。これは意図的に必須の依存関係とはされていない。earthUI は vProlog なしでもインストールでき、動作もする。Prolog 関連の機能は、vProlog が存在するようになるまで単に無効のままである。vProlog は著者の r-universe リポジトリを通じて配布されており、Windows と macOS 向けのビルド済みバイナリが提供される（コンパイラ環境は不要である）。

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

システムへの Prolog (SWI-Prolog その他) の導入は一切必要ない。Trealla エンジンもパッケージ自体に組み込まれているからである。本論文のそれ以外の内容は、vProlog なしでも成立する。

1.6 使い始め方

earthUI を使うには、次のようにする。

1. 上述のとおり CRAN からパッケージをインストールする。
2. **アプリケーションを起動する**。R で `earthUI::launch()` を実行するか、コマンドラインから `Rscript -e 'earthUI::launch()'` を実行する。アプリはウェブブラウザ上のポート 7878 で開く。ブラウザで `http://localhost:7878` に移動して直接アクセスすることもできる。(後処理モジュールも同じ方法で起動する。`launch_glmnet()` はポート 7879、`launch_mgcv()` はポート 7880 である。)
3. サイドバーのセクション 1 で**プロジェクトを作成または開く**。新規プロジェクトウィザードは、プロジェクト名、Purpose (General、For Appraisal、Market Area Analysis のいずれか)、所在地 (国、州、郡、市)、および任意で初期データファイルを受け取る。データファイルはプロジェクトの `in/` フォルダにコピーされる。
4. セクション 2 で**データを取り込む**。ここには、有効なプロジェクトの `in/` フォルダにある CSV ファイルと Excel ファイルが一覧表示される。Finder や Explorer から手動で追加したファイルは、更新 (Refresh) をクリックすると現れる。
5. **変数を構成する** — 目的変数と予測変数を選び、データ型を設定し、必要に応じて特殊列の役割を割り当てる。
6. **Earth パラメータを設定する** — `degree`、`nprune`、部分集合フィルタ、その他の `earth()` 引数を指定する。
7. **モデルを当てはめる** — “Fit Earth Model” をクリックし、メインパネルで結果を確認する。
8. **書き出す** — 予測値を Excel としてダウンロードし、Quarto レポート一式を生成し、あるいは (鑑定評価モードでは) RCA 補正と取引事例比較表を計算する。

設定はプロジェクトデータベースにサーバ側で保存され、プロジェクトと用途をキーとする。保存はセクション 3 およびセクション 4 の “Save current as default” ボタンによって行う (第 3 章 “Settings Persistence” を参照)。

2 MLS 入力データの要件

不動産の鑑定評価および市場分析の作業手順では、入力データは通常、複数リスティングサービス (MLS, Multiple Listing Service) からの書き出しに由来する。本章では、想定されるファイル構造と、earthUI が利用できる列について解説する。

2.1 ファイル形式と構造

earthUI は CSV と Excel (.xlsx、.xls) のファイルを受け付ける。取り込みの際、列名は自動的に snake_case に変換される。たとえば “Living SqFt” は living_sqft に、“Contract Date” は contract_date に、“Sale Price” は sale_price になる。この正規化により、作業手順の全体を通じて一貫した列参照が保証される。取り込み時に用いる CSV の区切り文字と小数点記号は、ロケール設定によって決まる（第 3 章 “Locale & Regional Settings” を参照）。

データファイルは、1 不動産につき 1 行、1 属性につき 1 列の平坦な表でなければならない。ファイルの最初の行には列見出しが含まれている必要がある。

2.2 鑑定評価モードで必要となる列

earthUI はどのような列の組合せでも動作するが、完全な鑑定評価の作業手順（RCA 補正と取引事例比較表）は、MLS の書き出しに以下の列があると効果を発揮する。

列	特殊型	用途
Sale Price	(目的変数)	モデルの応答変数
Contract Date	contract_date	sale_age（基準日からの日数）の計算に用いる
Listing Date	listing_date	DOM 列がない場合に、契約日と併せて DOM を計算するために用いる
Days on Market	dom	市場滞留日数。取引事例比較表に表示される
Concessions	concessions	売買時の譲歩。取引事例比較表では Net SP = Sale Price – Concessions となる
Living Area (SF)	living_area	平方フィート当たり残差 (residual_sf、cqa_sf) を有効にする
Lot Size	lot_size	取引事例比較表の “Site Size Dimensions” 行にまとめられる
Site Dimensions	site_dimensions	取引事例比較表で敷地面積とまとめられる
Latitude	latitude	小数点以下 3 桁に丸められる。近接性の計算と Location グループに用いる
Longitude	longitude	小数点以下 3 桁に丸められる。近接性の計算と Location グループに用いる
Area ID	area	取引事例比較表の “Loc: Long Lat Area” 行にまとめられる
Actual Age	actual_age	取引事例比較表の “Actual Age Effective Age” 行にまとめられる
Effective Age	effective_age	取引事例比較表で実際の築年数とまとめられる
Parcel Number	(なし)	取引事例比較表の APN 行に表示される
Listing ID	(なし)	取引事例比較表の APN 行に表示される
Address	display_only	取引事例比較表の Address 行に表示される。モデルからは除外される

表計算ソフトの列名は外国語であってもよい。“特殊” な名称が英語であるのは、R のプログラムがそれらを特別に扱えるようにするためである。それ以外の場合、与えられた列名

は回帰モデル、グラフ、そして（鑑定評価を行う場合は）中間取引事例比較表にそのまま現れる。

すべての列が必須というわけではない。earthUI は適応的であり、ある列が欠けていれば、対応する機能が単に省かれるだけである。たとえば `concessions` 列が指定されていなければ、取引事例比較表の Net SP 行には譲歩額を控除しない売買価格が表示される。ただし、不動産の価格モデルにおいては、許容できる当てはまりを得るために、いくつかの列が強く推奨される。

1. **Sale Age** — 契約による売買の日付と、鑑定評価または分析の基準日との間の日数。複数年にわたる売買履歴を、とりわけ 5 年を超える期間について用いる場合、`sale_age` は売買価格の推定において中心的な役割を果たすことが多い。実際、その重要性はしばしば大きく、これがなければ earthUI がモデルをまったく提示できないこともある。
2. **Living Area** — “Living Sqft”、“GLA”（総居住面積）などの名称でも呼ばれる。これも売買価格を左右する主要な決定要因の一つである。
3. **Total Bath Count** — フルバス、1/4 バス、ハーフバス、3/4 バスの合計数。たとえばフルバス 2 つとハーフバス 1 つであれば 2.5 となる。
4. **Garage Bays** または **Garage Area** — ガレージの収容台数、またはガレージの平方フィート数。
5. **Lot Size** — 不動産の土地面積。通常は平方フィートまたはエーカーで表す。
6. **Longitude**、**Latitude**、および利用できる場合は **Area ID**。これらに対する補正は、取引事例比較表では単一の Location 補正としてまとめられる。

2.3 特殊列の命名規約

earthUI は列を、その列名によってではなく、**特殊型の指定**によって識別する。MLS の書き出しにおいて列にどのような名前を付けてもかまわない。重要なのは、変数構成表（第 6 章）で正しい特殊型を割り当てることである。

たとえば、利用している MLS は居住面積を “GLA”、“Living SqFt”、“liv_area”、“gross_living_area” などとして書き出すかもしれない。取り込み後（snake_case に変換される）、Special のドロップダウンで `living_area` と指定すればよい。そうすれば、earthUI は元の名称にかかわらず、その列を平方フィート当たり残差の計算と取引事例比較表のグループ化に用いる。

2.4 データの品質と完全性

- **欠測値 (NA)**: 予測変数または目的変数の列に NA を含む行は、当てはめの前に自動的に除去される。`na.action` は内部的に常に `na.fail` に設定されており、NA の除去は `earth()` の呼出しに先立って処理される。
- **日付列**: 日付は R が解釈できる形式（たとえば 2025-06-15、06/15/2025、June 15, 2025）であるべきである。earthUI は、値の 80% 以上が正しく解釈できる場合に、その列を日付列として自動検出する。2 桁の年（たとえば 06/15/25）は最も近い該当年として読まれるので、近年の MLS の書き出しは正しく解釈される。4 桁の年は常に文字どおりに受け取られる。
- **数値列**: 売買価格、居住面積、敷地面積、譲歩額などの項目は数値でなければならない。MLS が通貨記号や桁区切りを伴って価格を書き出す場合（たとえば “\$350,000” や “350.000,00”）、取り込みの前にこれらを整える必要があるかもしれない。
- **因子列**: 地区 ID、様式、状態といったカテゴリ変数は、扱える範囲の個数の異なる値を含むべきである。earthUI は因子を自動検出するが、その検出結果は上書きできる。

ヒント: 取り込み後は予測変数表の T/F/NA 列を確認せよ。各列の真／偽／NA の件数が示される。欠測値が 30% を超える列は赤で強調される。NA の多い列を除外するか、取り込み前にデータを整えることを検討されたい。

2.5 対象不動産の行の配置

Appraisal（鑑定評価）モードでは、第 1 行が対象不動産でなければならない。残りのすべての行は取引事例である。対象不動産の売買価格は空欄 (NA) のままでもよく、任意の値を入れてもよい。いずれにせよ、earthUI は当てはめの間これを NA として扱う。

Market Area Analysis（市場地域分析）モードでは、第 1 行に対象不動産を置くことは任意であり、自動的に除外される行はない。対象不動産の行（あるいは他の任意の行）を当てはめから外したい場合は、指定した重み列でその行に 0 の値を与えるか、部分集合フィルタを用いる。

General（一般）モードでは、行の特別な取扱いはなく、すべての行が等しく扱われる。

3 一般用途モード

3.1 概要

一般用途 (General Purpose) モードは、earthUI を起動したときの既定である。これは不動産に限らず、あらゆるデータセットに対して完全な MARS モデリングの作業手順を提供する。科学データ、財務分析、工学研究、あるいは変数間の非線形な関係や交互作用を探索したい任意の回帰問題に、earthUI を用いることができる。

General モードでは、インターフェースは不動産に固有の機能（特殊列、sale age、座標の丸め、RCA）を省く。サイドバーは、変数の選択、パラメータの構成、モデルの当てはめ、書き出しに集中できるよう簡素化される。

3.2 サイドバーの作業手順

サイドバーは番号付きの折りたたみ可能なセクションに整理されており、プロジェクトの選択から書き出しまでを案内する。

1. Project — 既存のプロジェクトを選ぶか、新規プロジェクトウィザードで新しいものを作成する。ウィザードは用途、国、行政上の所在地（いずれも作成後は変更不可）を受け取る。有効なプロジェクトによって、入力ファイルをどこから読み、すべての出力をどこへ書くかが決まる。出力フォルダを指定する別個の入力欄は存在しない。

2. Import Data — 有効なプロジェクトの in/ フォルダから CSV または Excel のファイルを選ぶ。Refresh リンクは、手動で追加されたファイルを探すためにフォルダを再走査する。複数のシートをもつ Excel ファイルの場合には、シート選択欄が現れる。列名は自動的に snake_case に変換される。

3. Variable Configuration — 目的変数の選択欄（複数の目的変数に対応）、基準日 (Effective Date)、そして Include、Factor、Linear、Latent のチェックボックスを備えた予測変数表。詳細は第 6 章を参照せよ。

4. Earth Call Parameters — earth() 関数へのすべての引数、すなわち degree、penalty、nk、枝刈りの方法、交差検証などである。パラメータの完全なリファレンスは第 7 章を参照せよ。

Execute Prolog Processing — セクション 4 とセクション 5 の間に置かれた、番号のない任意のセクションであり、設定で実験的な Prolog 機能を有効にしない限り無効である。第 15 章を参照せよ。

5. Fit Earth Model — 乱数シードの入力欄と、モデルを非同期に実行するボタン。当てはめの詳細は第 8 章を参照せよ。

6. Download Estimated Target Variable(s) & Residuals — （鑑定評価モードと市場モードでは “Download Estimated Sale Prices & Residuals” と表示される）予測値、残差、CQA スコア、g 関数ごとの寄与を Excel ファイルとして書き出す。第 9 章を参照せよ。

7–8. RCA Adjustments と Sales Grid — 鑑定評価モード専用。第 10 章および第 11 章を参照せよ。

9. Generate Quarto Report — 自己完結型の Quarto レポート一式（.qmd のソースと付随資源）をプロジェクトの出力フォルダに書き出す。鑑定評価モード以外では番号は 7 となる。第 12 章を参照せよ。

10. Convert Quarto Report — 任意の .qmd ファイルを HTML、Word、PDF に変換する。第 12 章を参照せよ。

3.3 メインパネルのタブ

当てはめの後、メインパネルは 10 個のタブを提供する。

タブ	内容
Data	取り込んだデータのプレビュー。鑑定評価モードでは、対象不動産の表と取引事例の表に分かれる。
Equation	当てはめたモデルの方程式を LaTeX/MathJax で表示し、各基底関数とその係数を示す。
Summary	主要な指標 (R^2 、GRSq、GCV、RSS、 $CV R^2$) と係数の表。
Variable Importance	予測変数の重要度スコアの棒グラフと順位表。
Contribution	g 関数の寄与プロット。ドロップダウンから g 関数を選ぶ。1 変数の予測変数は折れ線プロットを、2 変数の交互作用 ($\text{degree} \geq 2$) は 3 次元の透視図と等高線プロットを表示する。
Correlation	予測変数間の相関のヒートマップ。
Diagnostics	3 つのプロット、すなわち残差対当てはめ値、正規 Q-Q、実測値対予測値。
RCA Adjustments	(鑑定評価モード専用) 取引事例全体にわたる Residual Adj %、Net Adj %、Gross Adj % のヒストグラム。ステップ 7 (Calculate RCA Adjustments) の完了後に表示される。
Residual Contribution	モデルの残差を、モデルが一度も用いなかった列に対してまとめて回帰することにより、希少なまたは除外された特性の市場寄与を推定する (subsection 4.4 を参照)。
ANOVA	各基底関数の RSS への寄与を示す ANOVA 分解表。
Earth Output	'earth::summary()' からの生のテキスト出力であり、枝刈りパスの詳細を含む。

3.4 設定の永続化

earthUI は構成をサーバ側のプロジェクトデータベース (regProj のルートにある projects.sqlite) に、プロジェクトと用途をキーとして保存する。保存はボタン駆動である。セクション 3 の “Save current as default” ボタンは目的変数、基準日、予測変数の構成を保存し、セクション 4 のボタンは earth の呼出しパラメータと Allowed Interactions 行列を保存する。各ボタンは自らの担当分だけを更新するので、互いを壊し合うことがなく、**当てはめによる自動保存は行われない**。したがって、保存済みの既定値を上書きすることなく自由に試行できる。一つのプロジェクトに属する複数の入力ファイル (小さな検証用抽出、完全なデータセット) は一つの構成を共有する。設定はプロジェクトツリーの中に置かれているので、regProj フォルダを同期、バックアップ、共有する際には設定も一緒に移動する。

3.5 ダークモード

右上隅の月／太陽のアイコンをクリックすると、明るいテーマと暗いテーマを切り替えられる。テーマの選好はローカルストレージに保存され、セッションをまたいで維持される。

3.6 ロケールと地域設定

earthUI は、国を基礎とするロケール体系を通じて、国際的な数値、日付、CSV の書式規約に対応する。上部メニューバーの **Settings** (歯車) メニューにある **Country** ドロップダウンで、対応する 31 か国のプリセットを選択する。各プリセットは以下を構成する。

- **CSV の区切り文字** — 米国／英国／日本ではコンマ (,)、コンマを小数点記号として用いる欧州の大半ではセミコロン (;)。
- **小数点記号** — ピリオド (.) またはコンマ (,)。
- **桁区切り** — コンマ (米国／英国／日本)、ピリオド (ドイツ／イタリア／スペイン)、空白 (フィンランド／フランス／ポーランド／バルト三国／ウクライナ／ロシア)、ま

たはアポストロフィ（スイス）。

- **日付書式** — MM/DD/YYYY（米国）、DD/MM/YYYY（欧州の大半）、または YYYY-MM-DD（スウェーデン／リトアニア／日本／カナダ）。
- **用紙サイズ** — レター（米国／カナダ／メキシコ）または A4（それ以外のすべて）。

3.6.1 対応国

国	CSV 区切り	小数点	桁区切り	日付	用紙	
米国、カナダ (csv)	,	.	,	MDY/YMD	レター	
英国、アイルランド、豪州、NZ	,	.	,	DMY	A4	
ドイツ、オーストリア、ベルギー、オランダ、イタリア、スペイン、ポルトガル、デンマーク、ブラジル	;	,	.	DMY	A4	
フランス、ポーランド、チェコ、ノルウェー、ラトビア、エストニア	;	,	(空白)	DMY	A4	
フィンランド、スウェーデン	;	,	(空白)	DMY/YMD	A4	
スイス	;	,	,	DMY	A4	
ウクライナ、ロシア	;	,	(空白)	DMY	A4	
トルコ	;	,	.	DMY	A4	
日本、韓国	,	.	,	YMD	A4	
メキシコ	,	.	,	DMY	レター	
リトアニア	;	,	(空白)	YMD	A4	

3.6.2 用紙の上書き

国の選択欄の下にある **Paper** ドロップダウン（レターまたは A4）により、国を切り替えることなくレポートの用紙サイズを上書きできる。CSV の区切り文字、小数点記号、日付書式の順序は、選択した国のプリセットに従う。国を変更すると、用紙の上書きはその国の既定値に戻る。

3.6.3 既定値の保存

設定メニューで **Save** をクリックすると、ロケールの選好が全体設定として保存される。これらの既定値は、どのプロジェクトを開くかにかかわらず、以後のすべてのセッションに適用される。プロジェクトごとの設定（目的変数、予測変数、パラメータ）はプロジェクトデータベースに別途保存される。ロケールの既定値は、利用者単位の SQLite データベースを介してすべてのプロジェクトにわたって保持される。この 2 階層の仕組みは、複数国のデータを扱う監査法人のような組織のために設計されている。最も多く扱う国を既定として設定し、必要に応じて調整すればよい。

ヒント: プロットの軸と傾きのラベルにおける数値の書式は、ロケールに自動的に適応する。ドイツのロケールは桁区切りにピリオドを用い (200.000)、フィンランドは空白を (200 000)、スイスはアポストロフィを用いる (200'000)。通貨記号は表示されない。earthUI は通貨に依存しないからである。

4 鑑定評価モード

Purpose として **For Appraisal** を選択すると、earthUI は単一不動産の評価に向けて自らを構成する。第 3 章で解説した機能はすべて引き続き利用できる。本章では、鑑定評価に固有の追加事項のみを扱う。

4.1 対象不動産の行の取扱い

鑑定評価モードでは、データセットの第 1 行が対象不動産であり、残りのすべての行は取引事例である。入力ファイルはそれに従って整理されていなければならない（第 2 章を参照）。対象不動産の売買価格は空欄のままでよく、任意の値を入れてもよい。earthUI は当てはめの間、これを自動的に NA として扱う。

取り込みの後、Data タブは二つの部分に分かれる。**Subject Property**（第 1 行）と **Comparable Sales**（第 2 行以降）である。第 1 行は常にモデルの当てはめから除外され、“Skipping row 1 (subject). Fitting on N rows.” という通知がこれを確認する。当てはめの後も、モデルは対象不動産の行に対する予測値を生成し、出力では `est_<target>` として示される。

4.2 基準日と経過日数

Effective Date（基準日）の入力欄は、どの用途モードでも変数構成のセクションに現れる（既定値は当日の日付である）。鑑定評価に限らず、一般分析や市場分析にも基準日はある。ただしこれが実際に用いられるのは鑑定評価モードと市場モードだけである。そこでは、Special 列のドロップダウンでいずれかの列を `contract_date` と指定すると、earthUI は `sale_age` 列、すなわち各売買の契約日と基準日との間の整数日数を計算する。この列は、予測変数として元の日付列に取って代わる。

契約日を指定した後、初めて Fit をクリックしたとき、earthUI は `sale_age` 列を作成し、それを組み込むためにもう一度 Fit をクリックするよう通知する。

4.3 特殊列の指定

鑑定評価モードと市場モードでは、変数構成表の各予測変数に **Special** のドロップダウンが現れる。特殊型の完全な一覧と、その効果は次のとおりである。

特殊型	効果
no	既定。特別な取扱いをしない
contract_date	基準日からの sale_age の自動計算を起動する
listing_date	dom 列がない場合に、契約日と併せて DOM を計算するために用いる
dom	市場滞留日数の列を識別する。取引事例比較表に表示される
concessions	売買時の譲歩を識別する。取引事例比較表の Net SP の算式 (Sale Price – Concessions) に用いる
latitude	値を小数点以下 3 桁に丸める。ハーバーサイン距離による近接性と、取引事例比較表の Location グループに用いる
longitude	値を小数点以下 3 桁に丸める。ハーバーサイン距離による近接性と、取引事例比較表の Location グループに用いる
area	取引事例比較表の "Loc: Long Lat Area" 行にまとめられる
living_area	ダウンロード出力において平方フィート当たり残差 (residual_sf、cqa_sf) を有効にする
lot_size	取引事例比較表の "Site Size Dimensions" 行にまとめられる
site_dimensions	取引事例比較表で敷地面積とまとめられる
actual_age	取引事例比較表の "Actual Age Effective Age" 行にまとめられる
effective_age	取引事例比較表で実際の築年数とまとめられる
display_only	書き出しには含まれるがモデルの当てはめからは除外される列 (複数指定可)
remarks	(実験的) 任意の Prolog 処理段階によって特性列へと解析される自由記述の備考列 (複数指定可)。第 15 章を参照
list	(実験的) 任意の Prolog 処理段階によってワンホットの TRUE/FALSE 列に分割される、コンマ区切りの列 (複数指定可)。第 15 章を参照

一つの特特殊型につき指定できる列は一つだけである (display_only、remarks、list を除く)。ある特殊型を第 2 の列に割り当てると、第 1 の列からは自動的に解除される。変数名の隣には、割り当てられた特殊型を示す小さな青いバッジが現れる。

4.4 残差寄与 (希少特性の推定)

回帰の経路に乗せるにはあまりに希少な特性、たとえばごく少数の売買にしか現れない工房のようなものは、RCA の残差層に属する。**Residual Contribution** タブは、当てはめたモデルの残差を、モデルが一度も用いなかった列に対して最小二乗法で回帰することにより、それらの市場寄与を推定する。複数の列を選ぶとそれらは同時に推定されるので、関連のある付帯設備 (たとえば outbuilding_sqft と pr_has_workshop) は、寄与を二重に数えるのではなく分け合うことになる。二値のフラグは一括額の推定値をもたらす、フラグの立った各売買の残差を証拠として列挙する。数値列は単位当たりの寄与をもたらす。報告される標準誤差は 1 標準誤差であり、通常の有意性の目安は $t \geq 2$ である。支持する観測が 10 件に満たない項には、生成される作業ファイルの注記に自動的に CAUTION が付され、その注記は方法、推定値、留保事項をレポート用に一段落で文書化する。候補の一覧から除外されるのは、モデルが実際に用いた予測変数だけである。当てはめに供されながら前進パスで一度も選ばれなかった希少なフラグは、依然として候補たりうる。モデルはその価値をまったく吸収していないからである。

4.5 RCA 補正の概要

Calculate RCA Adjustments & Download ボタン (サイドバーのセクション 7。鑑定評価モードで当てはめの後にのみ表示される) は、対象不動産に対する各取引事例の、市場から導かれた補正を計算する。RCA の完全な作業手順は第 10 章で解説する。

RCA 補正を計算した後、**Generate Sales Grid & Download** ボタン（サイドバーのセクション 8）が利用可能になる。取引事例比較表の作業手順は第 11 章で解説する。

ヒント: 対象不動産に与える *CQA* スコアは、残差分布のどれだけを対象不動産に帰属させるかを制御する。*5.00* のスコアは、対象不動産を取引事例の中央値に置く。

5 市場地域分析モード

Purpose として **Market Area Analysis** を選択すると、earthUI は鑑定評価モードと同じ不動産固有の機能（特殊列、経過日数、座標の丸め）を提供するが、単一の対象不動産を評価するのではなく、不動産の集団を分析することに向けられる。

5.1 鑑定評価モードとの相違

- **対象不動産の自動的な除外がない** — 鑑定評価モードとは異なり、市場モードは第 1 行を当てはめから自動的に除外しない。すべての行が含まれる。対象不動産の行を当てはめから外したい場合は、指定した重み列でその行に 0 の値を与えるか（デモデータセットのように）、部分集合フィルタを用いる。
- **RCA と取引事例比較表のセクションがない** — “Calculate RCA Adjustments & Download” と “Generate Sales Grid & Download” の各段階は利用できない。市場モードは、取引事例ごとの補正ではなく、モデルの当てはめと出力に焦点を置く。
- **サイドバーの番号付け** — RCA と取引事例比較表の段階がないため、Generate Quarto Report のセクションは 9 ではなく 7 の番号を与えられる（Convert Quarto Report のセクションは 10 のままである）。

5.2 市場モードを用いるべき場面

市場地域分析モードが適しているのは、次のような場合である。

- 価値の決定要因を理解するために、近隣地域や市場地域についての回帰モデルを構築するとき
- 床面積、築年数、敷地面積、位置といった変数が、不動産の集団にわたって売買価格にどう影響するかを分析するとき
- 市場条件の分析や近隣地域の画定のための裏づけを準備するとき
- 第 1 行に対象不動産を含むデータセットを扱いつつ、それを当てはめに含めるか除外するかを選べるようにしておきたいとき

ヒント：市場モードは、特殊列の機能（経過日数、座標の丸め）は使いたいけど RCA 補正の作業手順は必要ないという、一般的な不動産回帰にも有用である。

6 変数の選択

サイドバーのセクション 3、すなわち **Variable Configuration**（変数構成）は、どの列がモデルに参加し、それらがどう扱われるかを選ぶ場所である。

6.1 目的変数

セクション 3 の冒頭にある **Target (response) variable(s)** のドロップダウンには、データセットのすべての列が並ぶ。標準的な単一応答のモデルであれば 1 列を、多重応答のモデルであれば複数列を選択する。複数の目的変数を選択すると、モデルは `earth(cbind(y1, y2, ...) ~ .)` を用いてすべての応答を同時に当てはめる。

目的変数として選択された列は、予測変数の一覧から自動的に除外される。

6.2 予測変数表

目的変数の選択欄の下に、残りのすべての列を次の項目とともに列挙する表が置かれる。

列	説明
Variable	列名（切り詰められている場合は完全な名称がツールチップに表示される）。その前に $\times$ / $+$ の切替が置かれる（後述）。鑑定評価モードと市場モードでは、青いバッジが割り当てられた特殊型を示す。
Type	データ型のドロップダウン。numeric、integer、character、logical、factor、Date、POSIXct
Include	チェックボックス。この列を予測変数としてモデルに含める
Factor	チェックボックス。この列をカテゴリ変数として扱う
Linear	チェックボックス。線形の項としてのみ入れることを強制する（ヒンジ関数を作らない）
Latent	チェックボックス。この列をモデルから完全に外す（earth からも、glmnetUI / mgcvUI へ書き出されるデータからも除外される）。潜在変数の作業のために取り置くためのものである
Special	ドロップダウン（鑑定評価モードと市場モードのみ）。後述の特殊列型リファレンスを参照
T/F/NA	列ごとの三つの件数。論理型では TRUE/FALSE/NA、数値型では $\neq 0$ / $= 0$ / NA である。欠測値が 30% を超える場合は赤で表示される

表の上のヒント行が略号を説明する。“Type = column data type, Inc = include as predictor, Factor = treat as categorical, Linear = linear-only (no hinges), Special = column role (e.g. contract_date).”

削除ではなく緩やかな無効化。各変数名の隣にある $\times$ の切替は、その行を灰色にし、データから列を取り除くことなくモデルから外す。切替は緑の $+$ に変わり、列を再び有効にする。列が破壊されることは決してない。

6.3 データ型の検出と上書き

earthUI は取り込み時にデータ型を自動的に検出する。数値、整数、論理、因子、日付の各列が認識される。日付のように見える文字列の列（値の 80% 以上が一般的な日付書式として解釈できるもの）は Date に分類される。

Type のドロップダウンを変更すれば、いかなる検出結果も上書きできる。列を **character** または **factor** に変更すると、Factor のチェックボックスが自動的に入る。型の変更は、その列が `earth()` 関数にどう渡されるかに影響する。

6.4 Factor フラグと Linear フラグ

- **Factor** — チェックすると、その列はカテゴリ変数として扱われる。これは、連続的な測定値ではなく離散的な群を表す `style`、`area_id`、`grade` のような列に適している。因子の列は指示（ダミー）変数としてモデルに入る。
- **Linear** — チェックすると、その列は線形にモデルに入ることを強制され、ヒンジ関数（区分線形関数）は作られない。ある変数が目的変数と厳密に線形の関係にあると分かっている場合に有用である。

6.5 特殊列型リファレンス

鑑定評価モードと市場モードでは、**Special** のドロップダウンが以下の選択肢を提供する。各型は多くとも一つの列に割り当てられる（複数を許す `display_only`、`remarks`、`list` を除く）。

日付と時刻の型:

- `contract_date` — `sale_age` の自動計算を起動する。元の日付列は、売買日と基準日との間の日数を測る整数列に置き換えられる。
- `listing_date` — 明示的な `dom` 列が指定されていない場合に、市場滞留日数を計算するための代替として用いられる（DOM = 契約日 - 掲載日）。
- `dom` — 市場滞留日数の列を識別する。取引事例比較表の APN 行および Date of Sale 行に表示される。

金額の型:

- `concessions` — 売買時の譲歩（売主による負担、買主への誘因など）を識別する。取引事例比較表で純売買価格を計算するために用いる。Net SP = Sale Price - Concessions である。

規模と位置の型:

- `latitude` — 過適合を防ぐため、値は自動的に小数点以下 3 桁に丸められる。ハーバーサイン距離による近接性の計算（対象不動産から各取引事例までの距離）に用いられ、取引事例比較表の Location 行にまとめられる。
- `longitude` — 緯度と同じ丸めの扱いを受ける。近接性と Location グループに用いられる。
- `area` — 通常は近隣地域または地区の識別子である。取引事例比較表の “Loc: Long | Lat | Area” 行で緯度・経度とまとめられる。
- `living_area` — ダウンロード出力において平方フィート当たり残差の計算（`residual_sf` と `cqa_sf`）を有効にする。
- `lot_size` — 取引事例比較表の “Site Size | Dimensions” 行にまとめられる。
- `site_dimensions` — 取引事例比較表で敷地面積とまとめられる（たとえば “75x120”）。

築年数の型:

- `actual_age` — 取引事例比較表の “Actual Age | Effective Age” 行にまとめられる。
- `effective_age` — 取引事例比較表で実際の築年数とまとめられる。

表示の型:

- `display_only` — その列は Excel への書き出しには含まれるが、モデルの当てはめからは完全に除外される。住所欄、MLS 番号、地番 ID、その他、予測変数とすべきでない参照用データに用いる。複数の列にこの指定を与えることができる。

テキスト抽出の型（実験的）:

- **remarks** — 任意の Prolog 処理段階によって特性列へと解析される自由記述の列（たとえば `public_remarks`）。その機能を設定で有効にするまでは何もしない。複数の列を指定できる。第 15 章を参照。
- **list** — 同じ処理段階によってワンホットの `<col>_<value>` TRUE/FALSE 列に分割される、コンマ区切りの列（たとえば `existing_structures`）。複数の列を指定できる。第 15 章を参照。

6.6 複数の目的変数

二つ以上の目的変数を選択すると、多重応答の Earth モデルが当てはめられる。複数の目的変数を選択した場合には、次のようになる。

- モデルは基底関数を目的変数間で共有しながら、すべての応答を同時に予測する
- **wp (response weights)** ボタンが有効になり、各目的変数に数値の重みを割り当てられるようになる
- 分散モデル (`varmod.method`) は無効になる（多重応答では対応していない）
- 結果のタブには、応答ごとの指標、方程式、診断プロットが表示される

ヒント: `display_only` と指定された列はデータセットに残り、*Excel* への書き出しにも現れるが、予測変数表からは完全に除外される。*ID* 列、住所、その他の参照用の項目に用いるとよい。

7 パラメータの選択

サイドバーのセクション 4、すなわち **Earth Call Parameters** は、`earth()` 関数が受け取るすべての引数への入口を提供する。各パラメータには青いヘルプアイコン (?) が付いており、ツールチップで説明が表示される。パラメータは折りたたみ可能な小節に整理されている。

7.1 基本パラメータ

パラメータ	既定値	説明
subset	(空)	行のフィルタ式。後述の “Subset Filtering” を参照。
weights	NULL	事例（行）の重みを与える列の選択欄。数値列のみが列挙される。
wp	NULL	多目的変数モデルにおける応答の重み。ボタンを押すと、目的変数ごとに一つの数値入力欄をもつダイアログが開く（既定はいずれも 1.0）。
keepxy	off	x、y、subset、weights をモデルオブジェクトに保持する。
trace	0	当てはめ出力のトレース水準（0 から 5 まで）。
glm	none	任意の GLM ファミリー。 gaussian、binomial、poisson のいずれか。
degree	1	交互作用の最大次数。degree を ≥ 2 に設定すると交差検証が自動的に有効になり、Allowed Interactions 行列が現れる。
penalty	2	ノット当たりの GCV ペナルティ。値を大きくするとより単純なモデルになる。

7.2 前進パス

パラメータ	既定値	説明
nk	自動	枝刈り前の最大項数。読み込んだデータに対する推奨値が自動的に入力される（後述の “Recommended Parameter Values” を参照）。
thresh	0.001	前進段階の閾値。値を小さくするとより多くの項が許される。
minspan	自動	ノット間の最小間隔。推奨値が自動的に入力される。負の値は予測変数当たりの最大ノット数を設定する。

パラメータ	既定値	説明
endspan	自動	端の間隔、すなわちノットからデータの端までの最小距離。推奨値が自動的に入力される。
newvar.penalty	0	新しい変数を導入することへのペナルティ（既存の予測変数の再利用を促す）。
fast.k	20	高速 MARS アルゴリズムにおいて考慮する親項の数。
fast.beta	1	高速 MARS の経年係数を制御する。

7.3 許容する交互作用

degree を 2 以上に設定すると、基本パラメータの下に交互作用行列が現れる。これは予測変数の対ごとに一つのチェックボックスを配した上三角の格子である。チェックの入った枠は、その二つの予測変数が交互作用してよいことを意味し、チェックを外すとその特定の交互作用が禁じられる。

予測変数名（行または列のラベル）をクリックすると、その変数に関するすべての交互作用が一括して切り替わる。上部の **Allow All** と **Clear All** のチェックボックスが一括操作を提供する。行列は固定見出しを用いるので、スクロールしても列と行のラベルは見えたとままである。

情報の通知が次のように注意を促す。“Interaction terms increase the risk of overfitting. Cross-validation has been enabled (10-fold).”

7.4 枝刈り

パラメータ	既定値	説明
pmethod	backward	枝刈りの方法。backward、none、exhaustive、forward、seqrep、cv のいずれか。
nprune	NULL	枝刈り後の最大項数。自動選択させるには空欄のままにする。

7.5 交差検証

パラメータ	既定値	説明
nfold	10	交差検証の分割数。0 に設定すると CV を無効にする。
ncross	20	交差検証の反復回数。
stratify	on	各分割が同様の応答の分布をもつよう、CV の標本を層化する。

degree が ≥ 2 のとき、交互作用による過適合を防ぐ助けとして、交差検証が自動的に有効になる（**nfold** が 10 に設定される）。

7.6 部分集合フィルタ

subset のテキスト入力欄は、モデルの当てはめにどの行を用いるかを絞り込む R 式を受け取る。式を直接入力してもよいし（たとえば `sale_age < 365 & area_id == 460`）、**Build filter...** ボタンで視覚的に組み立ててもよい。

Build Filter ダイアログ — “Build filter...” をクリックすると、案内付きのダイアログが開く。条件の各行は、列のドロップダウン、演算子（`<`、`>`、`<=`、`>=`、`==`、`!=`）、および列の型に応じて変わる値の入力欄をもつ。数値には数値入力欄、日付には日付選択欄、文字列／因子の列には一意な値のドロップダウンが与えられる。条件は **AND(&)** または **OR(|)** の連結子で結ばれる。下部のプレビューが式と、該当する行数を示す。**Apply** をクリックすると、その式がテキスト入力欄に挿入される。

日付の列は、手書きの式では `as.Date("...")` または `as.POSIXct("...")` で包まなければならない。Build Filter のダイアログはこれを自動的に処理する。

部分集合フィルタは非破壊的である。除外された行もデータセットに残り、Excel への書き出しでは予測値を受け取る。

ヒント: **AND** と組み合わせる際には、**OR** 条件を括弧でまとめよ。括弧がなければ $A \& B / C$ は $(A \& B) / C$ と評価され、意図したものとは異なるかもしれない。

7.7 パラメータの推奨値

earthUI は、セクション 4 の主要なパラメータの下に推奨値を表示する。これらの推奨値は、当てはめに用いる行数 n と選択された予測変数の数 p に応じて反動的に更新される。算式は Friedman の MARS 論文、earth の内部アルゴリズム、および経験的な検証から導かれている。**nk**、**minspan**、**endspan** の各欄は、データの読み込みとともに推奨値があらかじめ入力される。手作業による編集はそのまま保持される。

7.7.1 前進パスのパラメータ

nk (枝刈り前の最大項数) :

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

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 の既定値 $\min(200, \max(20, 2p)) + 1$ はデータセットの大きさを考慮しておらず、大きなデータセットには小さすぎることがあり、前進パスがすべての予測変数を十分に探索する前にこれを制約してしまう。 $\lfloor n/10 \rfloor$ の項は、観測 10 件当たりおよそ 1 項を許すもので、過剰なパラメータ化を避けるための標準的な経験則である。上限の 100 は、収穫逨減を防ぐ。

minspan (ノット間の最小観測数) :

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

endspan (データの境界からの最小観測数) :

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

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 が自動計算する minspan は Friedman の式 43、すなわち $\lfloor (-\ln(-\ln 0.95) + \ln(p \cdot n)) / (2.5 \ln 2) \rfloor$ を用いる。これは $\ln(np)$ に比例して増えるため、大きなデータセットに対しては増加が遅すぎる。 $n=1500$ では ≈ 7 にしかならず、連続的な予測変数一つ当たりおよそ 245 か所のノット候補が生じる。これは多すぎて、前進パスが雑音に当てはまることを許してしまう。推奨の算式は、大きな n に対しては予測変数一つ当たりおよそ 100 か所の候補を目標としつつ、小さな n に対しては（十分に検証されている）earth の自動計算に委ねる。endspan は minspan よりわずかに大きく設定され、境界に対する追加の保護を与える。これは、データの裾が薄い場合（不動産ではよくあることである）に役立つ。

penalty（ノット当たりの GCV ペナルティ）：

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

これは earth 自身の既定であり、加法モデルでは 2、交互作用モデルでは 3 である（大きい方のペナルティは、探索空間が広いことを補償する）。

newvar.penalty（新しい予測変数を導入することへのペナルティ）：

推奨値は、予測変数どうしに相関がある場合（たとえば居住面積と寝室数、浴室数、敷地面積）に **0.1** である。これは前進パスを、相関のある代替変数を新たに導入するよりも、すでにモデルにある予測変数を再利用する方向へ偏らせる。その仕組みは次のとおりである。新しい予測変数の RSS の改善は $1/(1 + \text{penalty})$ 倍される。newvar.penalty = 0.1 であれば、新しい変数は既存の変数にもう一つノットを置く場合よりも、少なくとも 10% 多く RSS を改善しなければならない。これにより、最終的な係数の推定値に影響を与えることなく（選択の後にペナルティは取り除かれる）、より少ない予測変数からなる、より単純なモデルが得られる。予測変数どうしに相関がなければ 0 のままにしておく。なお、earth は 0 でない newvar.penalty と事例の重みとの併用に対応していない。重みの列が指定されている場合、earthUI はこの入力欄を無効にする（0 を強制する）。

7.7.2 枝刈りのパラメータ

pmethod: 推奨値は **backward** である。既定の GCV に基づく後退枝刈りは決定論的であり、同じデータからは常に同じモデルが得られ、乱数シードに依存しない。これは再現可能性にとって重要であり、とりわけ鑑定評価の業務では重要である。cv の方法は枝刈りに交差検証を用いるため、シードへの依存が生じる。

nprune: 推奨は **空欄のままにすること (NULL)** である。最適な項数は GCV に選ばせる。nprune を設定すると、GCV の判断を上書きする厳格な上限を課することになる。

7.7.3 交差検証のパラメータ

nfold（CV の分割数）：標本の大きさの帯域ごとの推奨値は次のとおりである。

n (当てはめに用いる行数)	nfold
0-299	5
300-399	6
400-499	7
500-599	8
600-699	9
700+	10

分割を増やせば、各分割はより多くのデータで訓練される（偏りの小さい推定となる）が、取り置く検証用の分割は小さくなる。上の表は、標本が大きくなるにつれて標準的な 10 分割へと段階的に上げつつ、各検証用分割が分割ごとの指標を安定させるのに十分な大きさ（おおよそ 50 行以上）を保ち、小さな標本に対しては 5 分割に戻す。小標本では CVR^2 の推定はなお雑音を含むので、分散を平均化するために `ncross` を上げるとよい（たとえば 15-20）。上限を 10 としているのは、分割を一つ増やすごとに `earth` の完全な当てはめが一回増える一方で、診断上の利得はごくわずかであることを反映している。`pmethod = "backward"` の場合、交差検証はモデルに影響しない。診断のための CVR^2 を計算し、分散モデルのための残差を提供するだけである。

`ncross` (CV の反復回数) : 標本の大きさの帯域ごとの推奨値は次のとおりである。

n (当てはめに用いる行数)	ncross
0-199	20
200-399	16
400-699	12
700-1199	10
1200-2999	7
3000+	5

反復は分割の当たり外れを平均化して CVR^2 を安定させる。標本が大きければ、一度の交差検証だけですでにより安定しているので、必要な反復は少なく済む。反復された交差検証の残差は、分散モデルにも供給される。

7.7.4 分散モデル

`varmod.method`: 推奨値は `lm` であり、これは Milborrow の既定でもある。予測区間のための頑健な線形分散モデルである（ばらつきは雑音を含む残差から推定されるので、単純なモデルの方がよく一般化する）。手元のデータで `lm` が収束しない場合には、`rlm` または `const`（等分散を仮定し、常に収束する）を試すか、`earth`（非線形なばらつきにも対応する）に戻すとよい。`nfold > 1` と `ncross > 1` を必要とする。

7.7.5 Earth の出力ファイル

当てはめに成功するたびに、`earthUI` は `<filename>_earth_output_<timestamp>.txt` を出力フォルダに書き出す。このファイルには、モデルの項、要約統計量 (R^2 、 $GRSq$ 、 $CVRSq$)、分散モデルの詳細、そしてトレースログが含まれる。当てはめ一回につき一つのファイルが作られるので、パラメータの構成を比較するための累積的な記録となる。

7.8 設定の既定値

セクション 4 の上部にあるラジオボタンが、どの既定値を読み込むかを制御する。

- **Use last settings for input file** — この特定のファイルで前回用いた設定を復元する

- **Use default settings** — 保存した独自の既定値を適用する
- **Earth defaults** — すべてのパラメータを出荷時の値に戻す

セクション 4 の下部にある **Save current as default** をクリックすると、現在の earth パラメータと Allowed Interactions 行列が保存される。セクション 3 の対応するボタンは、目的変数、基準日、予測変数の構成を保存する。いずれもプロジェクトと用途ごとに、プロジェクトデータベースに保存される。

8 Earth モデルの当てはめ

8.1 Fit ボタン

サイドバーのセクション 5 には、**Fit Earth Model** ボタンと **Random seed** (乱数シード) の入力欄が置かれている。シードは交差検証の分割の割当てを再現可能にする。ドロップダウンは直近に用いた 10 個のシードを呼び出し、用いたシードは当てはめの後に Earth Output タブに記録される。当てはめの前に、earthUI は構成を検証する。目的変数が選択されており、少なくとも一つの予測変数が含まれていなければならない。鑑定評価モードと市場モードでは、緯度と経度の列が丸められ、経過日数が基準日と契約日 (指定されている場合) から計算される。

8.2 当てはめのモーダルとトレース出力

Fit をクリックすると、暗いモーダルの覆いが現れ、そこには次のものが含まれる。

- **見出し** — “Fitting Earth Model” の文字と、経過秒数を数えるタイマー、および背景で進行中の当てはめを取り消す **Abort** ボタン
- **トレースログ** — スクロールできる等幅のテキスト領域であり、earth() 関数からの実時間の出力、すなわちデータセットの情報、前進パスの進行、交差検証の分割、完了時刻を表示する
- **色分け** — 既定の行は緑、交差検証の行は黄、エラーは赤

earthUI は `callr::r_bg()` を用いてモデルを非同期に当てはめる。これは計算を別の R プロセスで実行するものである。これにより、長時間を要する当てはめの間もアプリケーションは応答性を保つ。300 ミリ秒ごとの監視処理が背景プロセスからの出力を読み取り、トレースログへ流し込む。

当てはめが完了すると “Done in X.Xs” の表示が現れ、モーダルに閉じるボタン (X) が追加され、背景の覆いが取り除かれるので、結果のタブを必要に応じて描画できるようになる。その後、見えている出力の描画が終わるとモーダルは自動的に消える。それより早く閉じることもできる。

成功すると、Fit ボタンに緑のチェックマークが添えられる。エラーが生じた場合には、エラーメッセージがトレースログに表示され、通知が現れる。

8.3 当てはめログ

当てはめは成功・失敗を問わず、そのつど出力フォルダにログファイルを書き出す。

- **ファイル名:** `<datafile>_earth_log_<YYYYMMDD_HHMMSS>.txt`
- **内容:** タイムスタンプ、当てはめ処理からのすべてのトレース出力、およびエラーメッセージ
- **場所:** 有効なプロジェクトの出力フォルダ

ヒント: `callr` が導入されていない場合、`earthUI` は単純な進捗バーを伴う同期的な当てはめに戻る。トレース出力を完全に活用するには `callr` を導入されたい。

9 データのダウンロード

当てはめの後、サイドバーのセクション 6 は、予測値と診断情報を含む Excel ファイルをダウンロードする。このセクションの表題は、一般モードでは “Download Estimated Target Variable(s) & Residuals”、鑑定評価モードと市場モードでは “Download Estimated Sale Prices & Residuals” である。この出力は、ステップ 7(RCA) において対象不動産に CQA (Condition-Quality-Appeal) の評点を与えるために用いる。出力は `residual_sf` と `cqa_sf` で並べ替えられており、対象不動産が順位のどこに位置するかを判断しやすくしている。モデルの品質が良ければ、回帰に入らなかった残差的な特性に基づいて、不動産は魅力の最も低いものから最も高いものへと順位付けられるはずである。中央の値はおおよそ 0 となり、下半分は負の値、上半分は正の値となるはずである。品質の最も劣る住宅、すなわち “fixers” (要修繕物件) は順位の低位近くに、最も良好な住宅は上位に見いだされるはずである。担保権実行、ショートセール、検認（相続に関わる）売買、転職その他の事情による急ぎの売買といった変則的な事例については、通常は例外が生じる。変則的な事例を調べると、たいていは価格の変則性について適切な理由が明らかになる。

ボタンの表示は用途モードに応じて変わる。

- ・ 一般／市場: **Download Output (Excel)**
- ・ 鑑定評価: **Download Intermediate Output (Excel)**

ファイル名の形式は `<datafile>_modified_<YYYYMMDD_HHMMSS>.xlsx` である。

9.1 出力の列

各目的変数について、次の列が付加される。

列	説明
<code>est_<target></code>	モデルの予測値。たとえば <code>est_sale_price</code> (小数点以下 1 桁)
<code>residual</code>	実測値から予測値を引いたもの (小数点以下 1 桁)
<code>cqa</code>	Condition-Quality-Appeal の評点 (小数点以下 2 桁、範囲は 0-10)
<code>residual_sf</code>	残差を居住面積で割ったもの (居住面積が指定されている場合。小数点以下 1 桁)
<code>cqa_sf</code>	<code>residual_sf</code> による順位から計算した CQA (小数点以下 2 桁)
<code><variable>_contribution</code>	g 関数ごとの寄与の値 (小数点以下 1 桁。g 関数一つにつき 1 列)
<code>basis</code>	切片の寄与の値。すべての不動産で同じ (小数点以下 1 桁)
<code>calc_residual</code>	実測値から (<code>basis</code> + すべての寄与) を引いたもの。検算用の列 (小数点以下 1 桁)

多目的変数のモデルでは、各目的変数の列を区別するために `_<i>` の接尾辞が加えられる。

9.2 列の順序と Excel の書式

順位付けの列は出力ファイルの最も左の位置に、`residual_sf`、`cqa_sf`、`residual`、`cqa` の順で置かれる。これにより、並べ替えられた一覧を見渡して、対象不動産が CQA の分布のどこに落ちるかを容易に評価できる。

これらの列には Excel の数値書式が適用される。

列	書式	例
<code>residual_sf</code>	数値、小数点以下 2 桁	12.34
<code>cqa_sf</code>	数値、小数点以下 2 桁	7.25
<code>residual</code>	数値、小数点以下 0 桁	15,200
<code>cqa</code>	数値、小数点以下 2 桁	6.80

9.3 CQA の評点

CQA (Condition-Quality-Appeal) の評点は、各行の残差を他のすべての残差に対して順位付ける。ある行の CQA は、それより符号付き残差の小さい行の割合を 10 倍したものである。これにより 0-10 の尺度が得られ、その意味は次のとおりである。

- **高い CQA** ($\approx 9-10$) — その不動産はモデルの予測を大きく上回る価格で売れた (大きな正の残差)
- **低い CQA** ($\approx 0-1$) — その不動産は予測を大きく下回る価格で売れた (大きな負の残差)
- **CQA ≈ 5** — 残差がすべての残差の中央値の近くにある

`living_area` 列が指定されている場合、`cqa_sf` は平方フィート当たり残差に基づく同様の順位付けを提供する。

9.4 並べ替えと対象不動産の行

鑑定評価モードと市場モードでは、取引事例の行は `residual_sf` の降順 (居住面積が指定されていないならば `residual` の降順) で並べ替えられる。対象不動産の行 (第 1 行) は 1 番目の位置にとどまり、並べ替えられない。一般モードでは、行は元の順序のまま書き出される。

この並べ替えは、最も左に置かれた順位付けの列と相まって、鑑定評価者が取引事例を、最も過大に予測されたものから最も過小に予測されたものへと素早く見渡し、対象不動産に与えた CQA の評点とその分布のどこに落ちるかを評価することを可能にする。

予測用データの因子の水準は、`predict()` を呼ぶ前に訓練データのそれに揃えられる。未知の因子水準をもつ行は NA の予測値となる。

ヒント: `calc_residual` の列は常に (丸めの範囲内で) `residual` と一致するはずである。両者が大きく異なる場合は、調べる価値のある計算上の問題を示している。

9.5 mgcvUI への自動書き出し

`degree` が ≤ 2 で当てはめに成功するたびに、`earthUI` は結果オブジェクトの全体を `.rds` ファイルとして自動的にプロジェクトの出力フォルダに保存する。ファイル名は `<datafile>_earthUI_result_<YYYYMMDD_HHMMSS>.rds` の形式に従う。

このファイルは、`earthUI` パッケージに含まれ `earthUI::launch_mgcv()` で起動する `mgcv` の後処理モジュールが、`readRDS()` によって読み込むことができる。GAM のルーチンは、`earth` モデルのノット位置と基底関数を GAM の平滑化項の出発点として用い、MARS から GAM へのモデリングの滑らかな移行を可能にする。

`degree` が 2 を超えるモデルは飛ばされる。`mgcvUI` は 2 変数の交互作用にしか対応していないからである。`earthUI` を Trilogy モード (Trilogy のインターフェースから) で起動した場合には、追加の **Lock Model Output** セクションが現れ、特定の当てはめの `.rds` を、その予測変数の基底と共有の CQA 設定とともに、`glmnetUI` と `mgcvUI` が取り込む正典の `earth` モデルとして固定できる。

9.6 glmnetUI への取り込み

GAM のルーチンのために保存されたのと同じ `.rds` ファイルは、エラスティックネットの後処理モジュールにも取り込むことができる。これも同じく `earthUI` パッケージに含まれ、`earthUI::launch_glmnet()` で起動する。`glmnet` のアプリケーションでは、**Section 2: Import from earthUI** に進み、Browse ボタンで `.rds` ファイルを選択する。

`glmnetUI` は標準的な `earth::model.matrix()` の方式を用いる。すなわち、`earth` モデルの基底関数（ヒンジ、交互作用、線形項）が、`glmnet` の計画行列の列となる。次に `glmnet` がこれらの基底の列に対して正則化回帰を行い、ラッソ／エラスティックネットによってそれらを選択し縮小する。これにより、`earth` の適応的な基底の構成と `glmnet` の正則化とが組み合わせられる。

9.6.1 主な利点

- **解釈できる交互作用:** `earth` のヒンジに基づく交互作用（3 変数の項を含む）は明確な幾何学的意味をもつ。これは、法廷や監査の場で説明することが難しい `glmnet` 本来の交差積による交互作用とは対照的である。
- **自動的な非線形性:** `earth` のヒンジ関数は、生の予測変数を用いた標準的な `glmnet` モデルでは捉え損なう非線形の関係性を捉える。
- **変数の選択:** `glmnet` は `earth` の基底をさらに刈り込み、寄与しないヒンジ項を落とすことができる。

9.6.2 重要: 重みの列が一致していなければならない

警告

`earthUI` が重みの列を用いている場合（たとえば重みを 0 として対象不動産の行を除外している場合）、`glmnetUI` の変数構成でも同じ重みの列を指定しなければならない。重みの設定が異なると、行数が揃わず、`earth` の基底を `glmnet` のモデル行列に付加できない。

9.6.3 作業手順

1. `earthUI` で `earth` モデルを当てはめる（degree は任意）。
2. `.rds` の結果ファイルがプロジェクトの出力フォルダに自動的に保存される。
3. `glmnetUI` で **Section 2: Import from earthUI** を開き、`.rds` ファイルを選ぶ。
4. 両方のアプリケーションで同じデータファイルと同じ重みの列が用いられていることを確認する。
5. セクション 5 で `glmnet` のパラメータを構成する（`earthUI` からの取り込みが有効な場合、交互作用行列は自動的に無効になる）。
6. **Fit Model** をクリックする。`earth` の基底の列が、Equation、Coefficients、Contributions の各タブに現れる。

10 RCA の計算とダウンロード

Calculate RCA Adjustments & Download ボタンはサイドバーのセクション 7 に現れ、鑑定評価モードでモデルを当てはめた後にのみ表示される。RCA (Residual Constraint Approach、残差制約アプローチ) は、対象不動産に対する取引事例ごとの、市場から導かれた補正を生成する。

10.1 RCA ダイアログを開く

このボタンをクリックすると、次のものを備えた小さなモーダルダイアログが開く。

- **Score type** — **CQA** または **CQA per SF** を選ぶラジオボタン (SF の選択肢は `living_area` 列が指定されている場合にのみ利用できる)。“CQA per SF” を選んだ場合、対象不動産に与えた CQA の評点に基づき、その残差の評点は、当該 CQA_SF の評点に対応する `residual_sf` に居住面積を乗じたものとなる。
- **CQA value** — 対象不動産の CQA の評点 (0.00–10.00) の入力欄。この評点は、取引事例の品質の分布のどこに対象不動産が落ちると考えるかを表す。
- **Generate** ボタン — RCA の出力を計算し、Excel としてダウンロードする。有効な評点が入力されるまでボタンは無効のままである。

10.2 CQA の評点の内挿

Generate をクリックすると、earthUI は取引事例の CQA と残差の対から、対象不動産の残差を内挿する。

1. 取引事例の CQA の評点と残差が、CQA の昇順に並べ替えられる
2. 線形内挿 (`stats::approx()`) が、入力された CQA の値を残差へと写像する
3. CQA per SF を用いている場合には、平方フィート当たり残差に対象不動産の居住面積を乗じて、総残差へと戻される
4. 対象不動産の推定価値が、モデルの予測値 + 内挿された残差として計算される

10.3 出力の列

RCA の Excel ファイル (`<datafile>_adjusted_<YYYYMMDD_HHMMSS>.xlsx`) には、中間出力のすべての列に加えて、次の列が含まれる。

列	説明
<code>subject_value</code>	モデルの予測値 + 内挿された残差 (第 1 行のみ)
<code>subject_cqa</code>	入力した CQA の評点 (第 1 行のみ)
<code><variable>_adjustment</code>	対象不動産の寄与から取引事例の寄与を引いたもの (g 関数ごと)
<code>residual_adjustment</code>	対象不動産の残差から取引事例の残差を引いたもの
<code>net_adjustments</code>	すべての補正の合計 (寄与 + 残差)
<code>gross_adjustments</code>	すべての補正の絶対値の合計
<code>residual_pct</code>	取引事例の売買価格に対する残差補正の割合
<code>net_adj_pct</code>	取引事例の売買価格に対する純補正の割合
<code>gross_adj_pct</code>	取引事例の売買価格に対する総補正の割合
<code>adjusted_sale_price</code>	取引事例の売買価格 + 純補正

補正の列は、各取引事例について、モデルがその差異をどの変数にどれだけ帰属させているかを示す。`adjusted_sale_price` は、モデルから導かれたすべての補正を適用した後の取引事例の売買価格であり、これら補正後売買価格の集合は、対象不動産の推定価値の周りに集まるはずである。

10.4 RCA Adjustments タブ

RCA の出力が計算されると、メインパネルの **RCA Adjustments** タブが、取引事例全体にわたる Residual Adj %、Net Adj %、Gross Adj % の分布のヒストグラムを表示し（多目的変数のモデルでは目的変数ごとに一組）、補正がどれほど密に集まっているかを視覚的に手早く点検できるようにする。

ヒント: 5.00 の CQA の評点は、対象不動産を取引事例の中央値に置く。5 を超える評点は対象不動産が平均を上回る品質であることを、5 を下回る評点は平均を下回ることを示す。

11 取引事例比較表

11.1 概要と目的

取引事例比較表 (Sales Comparison Grid) は、鑑定評価モードで RCA 補正を計算した後に生成される Excel ワークブックである（サイドバーのセクション 8）。これは対象不動産と、選択された取引事例とを、構造化された表の形式で並べて示し、各取引事例の補正と補正後売買価格を自動的に計算する Excel の数式を備えている。

この表は鑑定評価者の作業ファイルのために設計されている。Earth モデルから導かれた回帰による補正と、鑑定評価者が CQA の残差を特定の不動産の特性に配分できる編集可能なセルとを組み合わせるものである。出力のファイル名は SalesGrid_<YYYYMMDD_HHMMSS>.xlsx である。

11.2 取引事例選択のモデル

“Generate Sales Grid & Download” をクリックすると、どの取引事例を含めるかを選ぶモーダルダイアログが開く。

- **推奨される取引事例**にはあらかじめチェックが入っている。これは総補正の割合が売買価格の 25% を下回る取引事例であり、経過日数の昇順（直近の売買が先）に並べられ、30 件を上限とする。
- **追加の取引事例**はその下に列挙され、既定ではチェックが入っていない。これらは総補正が 25% 以上のものであるが、なお選択の対象とできる（最大 50 件が表示され、総補正の割合の昇順に並ぶ）。
- 選択できるのは最大 **30 件の取引事例**であり、最大 **10 枚のシート**（1 シートにつき 3 件）が生成される。

選択を確定すると、選ばれた取引事例は総補正の割合の昇順に並べ替えられるので、最も類似した取引事例がシート 1 に現れる。

11.3 表の配置

各シートは、対象不動産と 3 件の取引事例を取める 20 列の配置をもつ。各主体に対する列は次のとおりである。

主体	列（主体ごとに 5 列）
対象不動産	ラベル、事実値 1、事実値 2、事実値 3、価値寄与 (VC)
各取引事例	事実値 1、事実値 2、事実値 3、価値寄与 (VC)、補正

上から下への行は次のとおりである。

1. **表題** — シート名（たとえば “Intermediate Sales Comparable Grid — Sheet 1 of 3”）
2. **見出し** — “Subject” および住所を伴う取引事例の列見出し
3. **Address** — 不動産の完全な住所
4. **APN | MLS# | DOM | Subj.Prox** — 地番、リスティング ID、市場滞留日数、および対象不動産からのハーバーサイン距離（マイル）
5. **Sales Price | Concess. | Net SP** — 売買価格、譲歩額、および純売買価格の数式
6. **Regression Features** の見出し行
7. **BASE VALUE** — モデルの切片
8. **Date of Sale | OffMkt | OnMkt** — 契約日、経過日数、DOM
9. **まとめられた行（条件付き）** — Location、Site Size、Age の各グループ
10. **モデル変数の行** — 予測変数ごとに 1 行（まとめられた変数を除く）
11. **空白の区切り行**

12. **Residual Features** の見出し行
13. **CQA|Residual** — CQA の評点+残りの残差の数式
14. **残差特性の行** — 名称の付いた行と、鑑定評価者が記入するための空白行
15. **Total VC / Net Adjustment**
16. **Net Adjustment %**
17. **Gross Adjustment %**
18. **Adjusted Sale Price** — 数式の行
19. 著作権の脚注

11.4 売買価格、譲歩額、純売買価格

第 5 行は各取引事例について三つの値を示す。

- **Sale Price** — データにおけるその取引事例の売買価格
- **Concessions** — concessions と指定された列の値（指定がなければ 0）
- **Net SP** — Excel の数式。Sale Price – Concessions

対象不動産の列は、売買価格については（それが未知であるため）“N/A”を示し、譲歩額は利用できる場合にその値を示す。

11.5 まとめられた行（位置、敷地、築年数）

特定の特殊型が指定され、かつそれらの変数がモデルに現れる場合、earthUI は関連する変数をまとめた行を作る。

Loc: Long | Lat | Area — longitude、latitude、area のいずれかがモデルにあるときに現れる。構成する各変数の事実値、まとめられた価値寄与（個々の VC の合計）、そして取引事例については補正（対象不動産のまとめられた VC – 取引事例のまとめられた VC）を示す。VC のセルは薄い青の背景で装飾される。

Site Size | Dimensions — lot_size または site_dimensions がモデルにあるときに現れる。Location グループと同じ構造である。

Actual Age | Effective Age — actual_age または effective_age がモデルにあるときに現れる。同じ構造である。

まとめられた行に取り込まれた変数は、下方の個々のモデル変数の行からは除外される。これにより、補正の合計における二重計上が防がれる。

11.6 モデル変数の行

まとめられた行の下（まとめられた行がなければ BASE VALUE のすぐ下）に、残りのモデルの予測変数ごとに 1 行が置かれ、次のものを示す。

- 対象不動産と各取引事例の**事実値**（実際のデータの値）
- **価値寄与 (VC)** — その行における予測値への g 関数の寄与
- **補正**（取引事例のみ） — 対象不動産の VC から取引事例の VC を引いたもの

11.7 CQA|Residual の行と残差特性の行

CQA|Residual の行は、各不動産の CQA の評点を示すとともに、**残りの残差**、すなわち特定の特性にまだ配分されていない残差の部分を求める数式を含む。その数式は次のとおりである。

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

その下には**残差特性の行**がある。名称の付いた 5 行（Location、View/Appeal、Condition、Quality、Other）に加えて、空白の 6 行である。これらは、モデルが捉えていない特性に

ついて鑑定評価者が価値寄与を入力するセルである。鑑定評価者が値を記入するにつれて、Remaining Residual の数式の値は自動的に減っていく。

各取引事例について、補正の列には数式が入っている。対象不動産の特性の VC – 取引事例の特性の VC である。

11.8 補正後売買価格の数式

Adjusted Sale Price の行には Excel の数式が入っている。

- **対象不動産:** BASE VALUE から最後の残差特性の行までの、すべての価値寄与のセルの合計。これは、対象不動産に対するモデルの総予測に、鑑定評価者が配分した残差特性を加えたものを表す。
- **取引事例:** Net SP に、最初のまとめられた行または変数の行から最後の残差特性の行までの、すべての補正のセルの合計を加えたもの。これにより、回帰から導かれた補正と鑑定評価者が入力した補正のすべてを適用した後の、その取引事例の売買価格が得られる。

11.9 シートの保護と編集可能なセル

各シートは、数式とデータが誤って変更されるのを防ぐために**保護されている**。保護によって固定されるのは次のものである。

- すべての数式のセル (Net SP、残りの残差、補正、補正後売買価格など)
- すべてのデータのセル (住所、売買価格、事実値、モデルから得た価値寄与)
- ラベルと見出し

固定されていない (編集できる) 唯一のセルは、残差特性の価値寄与の入力欄、すなわち CQA|Residual の行の下にある、鑑定評価者が内訳を入力するセルである。これらは編集可能であることを示すために、薄い黄色の背景で装飾されている。

11.10 Excel での表の扱い方

1. Excel または互換の表計算ソフトで**ファイルを開く**。
2. **回帰による補正を確認する** — モデルから導かれた補正はあらかじめ入力され、固定されている。
3. **残差を配分する** — 残差特性の行 (黄色のセル) に、残りの残差のうちどれだけが各特性 (Location、View、Condition、Quality など) に帰属すると判断するかを入力する。配分するにつれて Remaining Residual の値は減っていく。
4. **補正後売買価格を確認する** — この数式は、残差特性の値を入力するにつれて自動的に更新される。
5. **複数のシート** — 取引事例を 3 件より多く選んだ場合は、シート間を移動する。各シートには左の列に同じ対象不動産があり、それぞれ異なる 3 件の取引事例が並ぶ。

ヒント: 目標は、Remaining Residual がほぼ零になるまで、CQA の残差を特性の各行に配分することである。これにより、各取引事例の回帰による推定値と実際の売買価格との差異を説明できることが示される。

12 レポートのダウンロード

モデルを当てはめた後、レポートの生成は Quarto による二段階の作業手順となる。**Generate Quarto Report**（鑑定評価モードではサイドバーのセクション 9、それ以外では 7）が自己完結型のレポート一式を書き出し、**Convert Quarto Report**（セクション 10）がそれを、あるいは他の任意の Quarto ファイルを、最終的な形式へと描画する。

12.1 Quarto レポートの生成

Generate Quarto Report をクリックすると、プロジェクトの出力フォルダの下に `<base>_qmd/` に、自己完結型の Quarto 一式が書き出される。そこには、値の入った `.qmd` のソース、あらかじめ生成されたすべてのプロットの資源（PNG と PDF）、レポートのデータ、および Word の参照テンプレートが含まれる。この段階では描画は行われない。この一式は手作業で編集することも、Quarto の `include` によって他の Quarto ソースと組み合わせることも、`quarto preview` で単独でプレビューすることもできる。

12.2 Quarto レポートの変換

セクション 10 はファイル選択欄（既定では現在のプロジェクトで最も新しく生成された一式を指すが、任意の `.qmd` ファイルを選ぶこともできる）と、形式のチェックボックスを提供する。

- **HTML** — KaTeX による数式の描画、埋め込まれた画像、目次を備えた `.html` ファイル。Flatly Bootstrap のテーマと Roboto Condensed のフォントを用いる。
- **Word** — 編集と配布に適した `.docx` ファイル。目次とページ番号を含み、体裁を統一するために独自の参照テンプレートを用いる。
- **PDF** — 専門的な品質の出力のために LuaLaTeX で組版される。用紙サイズはロケール設定（レターまたは A4）に従う。大きな交互作用行列のための横向きのページを含み、Roboto Condensed と、数式のための Latin Modern Math を用いる。PDF の選択肢は、LaTeX の環境が検出された場合にのみ提供される。

Convert をクリックすると、選択した形式が描画される。描画は背景で実行される。モールドダイアログが経過時間と Quarto の進行を表示する間も、アプリケーションは応答性を保つ。完了時には通知がファイルの場所を知らせる。

すべての操作（モデルの当てはめ、データのダウンロード、RCA の計算、取引事例比較表の生成、レポートの描画）は、開始・終了のタイムスタンプと経過時間とともに、出力フォルダの `<datafile>_earthui_log.txt` に記録され、問題の切り分けと性能の監視に用いられる。

12.3 レポートの内容

生成されるレポートには、常に次の各節が含まれる。

1. **データセットの記述** — 観測数、目的変数、用いた予測変数、カテゴリ型の予測変数
2. **モデルの仕様** — degree、交差検証の有無、項の数、モデル内の予測変数
3. **許容した交互作用**（degree が ≥ 2 の場合のみ） — チェックマークを付した交互作用行列。出力の形式に合わせて整形される
4. **結果の要約** — R^2 、GRSq、GCV、RSS、および（交差検証を有効にした場合は） $CV R^2$
5. **モデルの方程式** — 完全な Earth モデルの方程式を LaTeX の記法で示したもの
6. **係数と基底関数** — すべての項、その係数、およびそれらを定義する基底関数の表
7. **変数重要度** — 予測変数の重要度スコアの棒グラフと順位表
8. **g 関数の寄与** — 各 g 関数のプロット。1 変数の項には折れ線プロット、2 変数の交互作用には 3 次元の透視図と等高線プロット
9. **相関行列** — 予測変数間の相関のヒートマップ

10. **診断** — 残差対当てはめ値、正規 Q-Q、実測値対予測値の各プロット
11. **ANOVA 分解** — 各基底関数の RSS への寄与を示す表
12. **Earth の出力** — `earth::summary()` からの生のテキスト出力。枝刈りパスの詳細を含む

多重応答のモデルでは、第 4 節から第 8 節までと第 10 節が、目的変数ごとに繰り返される。

ヒント: レポートには *quarto* の R パッケージと *Quarto* の導入が必要である。*PDF* 出力にはさらに *LaTeX* の環境が導入されていなければならない (たとえば `tinytex::install_tinytex()` による)。そうでなければ *PDF* のチェックボックスは表示されない。

13 後処理モジュール: エラスティックネットと GAM

earthUI パッケージは二つの後処理モジュールを同梱している。glmnet の上に構築されたエラスティックネットのルーチンと、mgcv の上に構築された GAM のルーチンである。それぞれについては本号の個別の論文が詳しく扱う。本節では、earthUI の作業手順においてそれぞれが何のためにあるのか、そして三つのルーチンがどう協働するのかを説明する。

13.1 なぜ補強のための当てはめを行うのか

価値の結論を生み出すのは earth モデルである。後処理モジュールは、それを裏づけるために存在する。すなわち、方法論的に異なるエンジンによって、同じプロジェクトのデータに当てはめられた、同一の対象不動産についての二つの追加の推定である。earth の基底のエラスティックネットによる再推定と、独立した GAM の平滑化とが実質的に同じ価値に到達するとき、鑑定評価者の結論はもはや単一のモデルの仮定に依存しない。これは審査や訴訟の場面で意味のある利点である。両者が一致しないとき、その不一致自体が診断的であり、たいいていはデータの薄さ、柔軟すぎる項、あるいは見直す価値のある外れ値の取引事例を指し示している。いずれのモジュールも earth モデルに優越しない。それらは補強の証拠であり、あえて重み付けを与えずに提示される。

13.2 起動と共有される状態

各ルーチンは、それぞれ別個のローカル Shiny アプリケーションである。

- earthUI::launch_glmnet() — エラスティックネットのアプリケーション。ポート 7879。
- earthUI::launch_mgcv() — GAM のアプリケーション。ポート 7880。

三つのアプリケーションはいずれも同じプロジェクトツリーと同じ設定データベースを読むので、earthUI で開いているプロジェクトはただちにモジュールでも利用できる。earthUI がそのプロジェクトのために保存する引継ぎ項目、すなわち応答（目的変数）、基準日、RCA における対象不動産の CQA の種別と値、および居住面積の指定も同様である。これらを入力し直す必要はない。モジュールがそれらを読み取るので、同じ値が三つの当てはめのすべてを貫いて流れる。

13.3 エラスティックネットのルーチン

glmnet のルーチンは、earth モデルの基底関数、すなわちそのヒンジ、交互作用、線形項を計画行列の列として取り込み、ラッソ／エラスティックネット正則化のもとで再推定する（仕組みと重みの列に関する留保については、前述の“glmnetUI Import”を参照）。その結果は、earth の解釈可能で幾何学的に意味のある項を保ちつつ、正則化装置に、値打ちのない項を縮小または削除させる。詳細は本号の glmnetUI の論文を参照されたい。

13.4 GAM のルーチン

mgcv のルーチンは逆の方向をとる。earth の基底を再利用するのではなく、earth モデルのノット位置を出発点として、新たに罰則付き平滑化項を当てはめる（前述の“mgcvUI Auto-Export”を参照）。GAM は異なる平滑化の枠組みのもとで関数形を導き直すので、earth モデルとの一致は、こだまではなく真に独立した裏づけとなる。degree が ≤ 2 の earth モデルに対応する。詳細は本号の mgcvUI の論文を参照されたい。

13.5 実務上の作業手順

1. まず earthUI で **earth** モデルを当てはめて洗練させ、満足のいくものにする。当てはめに成功するたびに (degree が ≤ 2 の場合)、モジュールが取り込む .rds が自動的に書き出される。
2. 同じプロジェクトに対して **glmnet** と **mgcv** のルーチンを開き、earth の結果を取り込んで当てはめる。一度に一つのルーチンで作業せよ。当てはめは長時間を要し、ルーチンどうしは共有のプロジェクトを介して連携するので、理にかなった進め方は並行ではなく逐次である。
3. 自由に反復せよ。各アプリケーションは、行き来する間も自らの結果を保持し、その途中でレポートを生成することを強いられることはない。
4. 三つの結果のすべてに満足したときに初めて、各ルーチンに戻ってその Quarto レポートを生成する。各手法は自らの当てはめと価値の結論をプロジェクトの三部作の記録に登録するので、三つのレポートは一つの比較実行に遡って辿ることができる。

ヒント: モジュールはどの段階でも任意である。earthUI 自体の作業手順、すなわち当てはめ、RCA、取引事例比較表、レポートは、それらがなくとも完結している。結論に裏づけが求められるとき、たとえば難しい対象不動産、データの薄い市場区分、審査を受ける可能性の高い業務の場合に、これらを用いるとよい。

14 デモデータセット: Appraisal_1.csv

earthUI には、鑑定評価の作業手順を試すためのデモ用 MLS データセットが含まれている。プログラムから読み込むには次のようにする。

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

あるいは Shiny アプリでは、これをプロジェクトの in/ フォルダに複製し（またはプロジェクト作成時に初期データファイルとして添付し）、セクション 2(Import Data) で選択する。

14.1 記述

このファイルには、模擬的な MLS の書き出しによる 1,501 件の住宅の売買と、第 1 行の 1 件の対象不動産（合わせて 1,502 行）が含まれている。データは、規模、築年数、位置の幅をもつ複数地区の市場における戸建住宅の売買を表している。

これは実在のデータではないが、北カリフォルニアの現実的な近隣地域に基づいている。識別情報はすべて改変または除去してある。

14.2 列

列	型	特殊型	説明
weight	numeric	—	観測の重み (0 = 当てはめから除外)
id	numeric	display_only	内部レコード ID
property_id	numeric	display_only	MLS の物件識別子
listing_id	character	display_only	MLS のリスティング番号
parcel_number	character	display_only	郡査定官の地番 (APN)
street_address	character	display_only	不動産の住所
city_name	character	display_only	市
postal_code	character	display_only	郵便番号
county_name	character	display_only	郡
contract_date	Date	contract_date	売買の契約日 (sale_age を計算する)
sale_age	numeric	—	契約日から基準日までの日数 (計算済み)
coe_date	Date	display_only	エスクロー完了日
listing_status	character	display_only	リスティングの状態 (たとえば”Sold”)
sale_price	numeric	(目的変数)	売買価格。応答変数
rent	numeric	—	月額賃料 (多目的変数モデル用)
list_price	numeric	display_only	売出価格
original_list_price	numeric	display_only	当初の売出価格
living_sqft	numeric	living_area	平方フィートによる総居住面積
beds_total	integer	—	寝室の数
baths_total	numeric	—	浴室の合計数 (たとえば 2.5 = フル 2 + ハーフ 1)
lot_size	numeric	lot_size	平方フィートによる敷地面積
area_id	integer	area	MLS の地区識別子
area_text	character	display_only	地区名
age	numeric	actual_age	年による築年数
year_built	integer	display_only	建築年
latitude	numeric	latitude	緯度 (モデル用に小数点以下 3 桁に丸められる)
longitude	numeric	longitude	経度 (モデル用に小数点以下 3 桁に丸められる)
latitude6	numeric	display_only	完全精度の緯度
longitude6	numeric	display_only	完全精度の経度
garage_spaces	integer	—	ガレージの収容台数
fp_count	integer	—	暖炉の数
no_of_stories	numeric	—	階数
style	character	—	建築様式
view	character	—	眺望の種別 (たとえば”Neighborhood”、”Hills”)
days_on_market	integer	dom	市場滞留日数
listing_date	Date	listing_date	掲載日
sale_concessions	numeric	concessions	売主による譲歩

14.3 推奨する手早い始め方

1. earthUI を起動する: `earthUI::launch()`
2. 新規プロジェクトウィザードで Purpose を **For Appraisal** とするプロジェクトを作成し、Appraisal_1.csv を初期データファイルとして添付する
3. セクション 2(Import Data) で Appraisal_1.csv を選択する
4. 目的変数として `sale_price` を選択する
5. 上の表のとおり特殊型を割り当てる
6. 予測変数として次を含める。`sale_age`、`living_sqft`、`baths_total`、`lot_size`、`area_id` (因子として)、`age`、`latitude`、`longitude`、`garage_spaces`
7. `degree` を 1 に設定し、**Fit Earth Model** をクリックする
8. 中間出力をダウンロードし (ステップ 6)、CQA の順位を確認する
9. CQA の評点を ~5.00 として RCA 補正を計算する (ステップ 7)
10. 取引事例比較表を生成する (ステップ 8)

ヒント: デモデータセットは、自前の *MLS* データを取り込む前に *earthUI* のすべての機能を試すのに適している。上に挙げた推奨の予測変数からは、通常 R^2 が 0.85 を超えるモデルが得られる。

15 実験的機能: Prolog によるテキスト抽出と導出規則

バージョン 0.10.0 は、MLS の書き出しの記述的なテキスト列から、モデルに用いられる特性を抽出するための任意の、そして真に実験的な機能を導入した。対象は何よりもまず “Public Remarks” (一般向け備考) の列であるが、仲介業者の備考や、室内の設備や既存の構築物といったコンマ区切りの特性一覧の列も含まれる。抽出のエンジンは Prolog であり、とりわけ **確定節文法 (DCG, Definite Clause Grammars)** である。備考のテキストの中に “move in ready”、“3 car garage”、“owned solar” といった語句を認識する規則は、Prolog の文法規則として書かれている。

この機能はなお開発中である。今日でも抽出のために部分的に用いることはできる。同梱の文法はすでに有用な一群の特性を認識する。しかし、それを自らの MLS の語彙に合わせるには、Prolog を、とりわけ DCG を読み書きできなければならない。それが自分に当てはまらないのであれば、本章はまるごと読み飛ばしてよい。この機能は既定では無効であり、earthUI のそれ以外の部分は、これがなくともまったく同じように働く。

警告

これは活発に開発中の実験的な機能である。その抽出文法は完全でもなければ、実務の鑑定評価業務のために検証されてもいない。抽出された列は、モデルにおいて依拠する前に、必ず原文のテキストと突き合わせて確認しなければならない。この機能を有効に用いるには、Prolog と DCG に関する実用的な知識が必要である。この機能が行うことはすべて付加的かつ可逆的である。何かの不都合に働いたなら、設定の切替を無効にして通常どおり作業を続けられればよい。

15.1 何をするのか

MLS のリスティングは、自由記述のテキストと区切り付きの一覧の項目の中に、価値に関わる情報を数多く抱えている。この任意の処理段階は、二つの特殊列の役割 (第 6 章) を介して、そのテキストを回帰が利用できる普通の列へと変える。

- **remarks の列**は、Prolog の DCG 文法によって解析され、抽出された属性ごとに 1 列の特性列となる。列の接頭辞は、元の列の名称の頭文字から作られる。public_remarks からは pr_* の列が、agent_remark からは ar_* が生じる、といった具合である。
- **list の列** (たとえば existing_structures = "Shed, Workshop") は、ワンホットの <col>_<value> TRUE/FALSE 列に分割される。この部分は純粋な R であり、Prolog をまったく必要としない。

同梱の文法 (inst/prolog/mls_remarks.pl) は、現在のところ、とりわけ次のものを抽出する。1-5 の condition_score を伴う段階付けされた状態の等級、眺望のフラグと眺望の種類、台所と浴室の改装のフラグ、硬材その他の床材、改良の一覧とその件数、ガレージの収容台数とガレージ面積、プール、スパ、その他の屋外設備、暖炉の数、“40x50” のような寸法から解析した大きさを伴う付属建物 (物置、納屋、工房、カシタ、来客用住居、犬舎など)、RV の駐車場と収納、太陽光発電の所有か賃借かの別、蓄電池、敷地面積 (エーカーから平方フィートへの換算を含む)、対象と結び付いた売買の文脈における金額 (値下げ、HOA の管理費、太陽光のリースまたはその清算、Mello-Roos)、位置の手がかり、そして [-1, 1] の範囲の長所／懸念の感情スコアである。

15.2 要件: SWI-Prolog ではなく vProlog (Trealla)

備考の解析は **vProlog** の R パッケージ、すなわち自己完結型の Trealla Prolog エンジンの上で動く。これは CRAN にはない。本論文の冒頭の “earthUI と vProlog のインストール” に示したとおり、著者の r-universe リポジトリから導入する (Windows と macOS 向けのビルド済みバイナリがあり、ソースは github.com/wcraytor/vProlog にある)。システムへの

Prolog (SWI-Prolog その他) の導入は必要でもなければ、用いられもしない。vProlog がない場合、備考の解析は注記とともに飛ばされるが、list の分割はなお実行される。補助関数 vProlog_available() は、エンジンと同梱の文法の双方が存在するかどうかを報告する。

15.3 有効化と実行

1. **Settings** (歯車) を開き、**Enable Prolog / list column processing** にチェックを入れる。この切替は既定では無効であり、“advanced feature — for users with Prolog expertise” の注記を伴い、プロジェクトごとに記憶され、Project Information のダイアログに表示される。
2. 予測変数表で、該当する列の **Special** の役割を **remarks** または **list** に設定する（それぞれの役割は複数の列がもちうる）。
3. サイドバーのセクション 4 とセクション 5 の間にある **Execute Prolog Processing** のセクションを用いる。そこには二つの段階がある。

1. **Expand lists & remarks** — 各 **remarks** 列に対して DCG 文法を実行し、各 **list** 列をワンホットに分割する。展開は 漸進的である。前回の実行以降に指定された列だけが処理され、指定されたものがすべて展開済みになるとボタンは無効になる。進捗バーが解析の進行を示す。

2. **Execute derivation rules** — 展開された列に対して、利用者自身の Prolog の **derive/2** 規則を適用する（後述）。規則は実行の前に構文検査と試験実行による検証を受ける。

15.4 導出規則

各データ行について、earthUI はその行の列を **cell(Name, Value)** の事実として主張し、利用者の **derive(Col, Val)** 規則のすべての解を収集する。異なる Col のそれぞれが新しい列となり、列ごとに 最初に見つかった値が採用されるので、代替の節を後に続けることができる。補助述語 (**count_true/2**, **count_false/2**, **count_true_prefix/2**, **cell_or/3**, **is_na/1**) が集計と NA の取扱いを支え、**drop/1** の事実は規則の実行後に列を破棄する。たとえば次のようになる。

```
% 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).
```

規則はプロジェクトごとの <project>_rules.pl ファイルに置かれ、十分に注釈の付いたテンプレートを種として作られる。編集は **Edit derivation rules...** から、浮動して移動できるコードエディタで行い、これはサイドバーに固定することもできる。**Check** ボタンは、規則を実行することなく検証する（構文の走査、読み込みの確認、およびデータの第 1 行に対する試験実行）。

15.5 版を付した有効データファイル

いずれの段階の後も、拡張されたデータフレームはプロジェクトの入力フォルダに `<original_name>_<YYMMDD_HHMMSS>.xlsx` として書き出され、そのプロジェクトの有効なデータファイルとなる。プロジェクトを開き直せば拡張済みのデータが読み込まれるので、展開が再計算されることはなく、`glmnetUI` と `mgcvUI` が取り込むのも同じファイルである。備考の元の列は自動的にモデルから外される。生成された列は予測変数として自動的に含まれることはない。テキストの解析は多くの列を生みうるので、(まだ) 回帰に入れたくないものは **Latent** の印を付けるか、× で無効にせよ (第 6 章)。

15.6 プログラムからの利用と同梱ファイル

この一連の処理はすべて、一括処理のために公開されている `execute_prolog_processing(df, remarks_cols, list_cols, rules_text)` が各パスを統括し、その基礎には `remarks_to_columns_()`、`split_list_column_()`、`apply_user_rules_()`、`validate_rules_()` がある。`inst/prolog/` の下に同梱された Prolog のソースは、`mls_remarks.pl` (DCG 文法と特性の集計)、`mls_analyze.pl` (一括解析のためのファイル入出力の駆動部)、`mls_rules.pl` (導出規則から利用できる補助述語)、および `climate_regions_ca.pl` (後述) である。

15.7 関連する実験: 気候地域の事前情報

これに付随する実験的な機能は、データの薄い特性、すなわちごく少数の取引事例にしか現れないために回帰が推定できない特性 (たとえば気化式冷却機) を扱う。三つの R の補助関数、すなわち `climate_region_for()`、`market_area_profile()`、`climate_feature_priors()` は、プロジェクトの所在地 (郡+市) を、買い手に認識される七つのカリフォルニアの気候地域 (カリフォルニアエネルギー委員会の 16 の建築気候帯を実務的に縮約したもの) のいずれかに分類し、地域ごとに空調と冷房の設備 (暖房、エアコン、気化式冷却機、全館換気扇、濾過装置) についての順序尺度の寄与価値の事前情報を返す。Prolog のモジュール `climate_regions_ca.pl` は、導出規則の内部で用いるために同じ分類を写している。両者は一つの正典となる R のソースから生成されるので、乖離しえない。この事前情報は、特性の評価に示唆を与え、推定された係数の妥当性を点検する役割を果たすことを意図している。地域の割当て (とりわけ複数の地域にまたがる郡について) は、鑑定評価者が確認すべきである。

15.7.1 カリフォルニアを超えて: 法域パックと市場地域モジュール

カリフォルニアは最初の法域であって、境界ではない。`earthUI` は国際的に用いられており、気候の事前情報に関する知識は法域パック (jurisdiction pack) として再構成されつつある。これは自己完結した、データの形をとる事実の表 (帯域の定義、行政単位の割当て、特性の事前情報) であり、プロジェクト体系の国および行政階層のコードをキーとする。したがって一つのパックが、米国の州を、ドイツの州 (Land) を、あるいは小さな国そのものを記述できる。これに付随する市場地域モジュールの層が、名前を与えられた市場地域について、その地域の語彙と精緻化された事前情報を捉える。

いずれも、バリュエーション・エンジニアと鑑定評価者が現場で、自らの言語で作成することを想定して設計されている。そして現実的には、LLM の支援を得ての作成である。世界に Prolog のプログラムは多くないので、現場で作られる成果物はあえてプログラムではなく表とされている。パックの形式は、完全な作例 (カリフォルニアそのもの) を伴う文書化されたスキーマであり、これはまさに、Claude のような助手が、ある市場の平易な言葉による記述から妥当なパックを起草できるようにするためである。`earthUI` はパックから基礎となる Prolog を生成し、パックを有効にできるようになる前には、妨げとなる検証器 (スキーマ、網羅性、参照整合性、試験実行の各検査) を通過しなければならない。現場で作成されたすべてのパックには、記録上の作成者、日付、および引用された典拠という来歴が必須として

付される。これにより、起草の道具ではなく責任ある専門家が知識の背後に立ち、レポートは用いたパックとその版を正確に引用できる。

ヒント: 実務的な進め方は次のとおりである。備考を展開し、細かな `pr_*` のフラグを `derive/2` 規則によって少数の導出された件数や真偽値へとまとめ、まとめた元のフラグを `drop/1` で落とし、導出された列だけを予測変数の候補として残す。こうすれば予測変数の一覧は短く保たれ、多重共線性も扱いやすい程度にとどまる。

参考文献

- 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>.