

mgcvUI: mgcv (广义可加模型) 软件包的交互式界面

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

William Bert Craytor

第 1 卷，第 1 期 · 2026 年夏季

发表日期: 2026 年 7 月 1 日

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

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

本开放获取文章依据 [知识共享署名 4.0 国际许可协议 \(CC BY 4.0\)](#) 的条款发布。

摘要。 mgcvUI 是 R 语言 mgcv 软件包的图形用户界面 (graphical user interface)，该软件包使用带自动光滑度选择 (smoothness selection) 的惩罚回归样条 (penalized regression spline) 来拟合广义可加模型 (generalized additive model, GAM)。它提供三种用途模式——通用预测建模、房地产估价与市场区域分析——并引导用户完成数据导入、光滑项设定、模型拟合、诊断图与效应图查看以及报告下载。本文记录了 mgcvUI 的数据格式要求、建模工作流程、输出显示界面以及完整的功能参考。

关键词： 广义可加模型，GAM，惩罚样条，光滑化，mgcv，R，图形用户界面，房地产估价

出版者按 (2026 年 7 月): 付印前不久，mgcvUI 已并入 earthUI (0.11.0 及以后版本)，本文所述的 GAM 工作流程现与 MARS 和弹性网络工作流程一同由 earthUI 提供。mgcvUI 仍以其最终发布形态 (0.3.0) 保留在 <https://github.com/wcraytor/mgcvUI>；后续开发将在 earthUI 中继续。本文适用于已发布的 mgcvUI。

目录

1	引言	4
1.1	历史与背景	4
1.1.1	广义可加模型	4
1.1.2	mgcv 软件包	4
1.1.3	与 MARS (earth) 和弹性网络 (glmnet) 的比较	4
1.1.4	earthUI-mgcvUI 流水线	5
1.1.5	推荐的工作流程	5
1.2	什么是 mgcvUI?	5
1.3	三种用途模式	6
1.4	房地产专用功能	6
1.5	入门	6
2	MLS 输入数据要求	8
2.1	文件格式与结构	8
2.2	估价模式的必需列	8
2.3	特殊列命名约定	9
2.4	数据质量与完整性	9
2.5	标的物行的位置	9
3	通用模式	10
3.1	概述	10
3.2	侧边栏工作流程	10
3.3	从 earthUI 导入 (可选)	11
3.4	主面板标签页	11
3.5	设置的持久化	12

3.6	暗色模式	12
3.7	区域与地区设置	12
4	估价模式	14
4.1	标的物行的处理	14
4.2	估价基准日与销售时距	14
4.3	特殊列指定	14
4.4	RCA 调整概览	14
5	市场区域分析模式	15
5.1	与估价模式的差异	15
5.2	何时使用市场模式	15
6	变量选择	16
6.1	目标变量	16
6.2	响应变量变换	16
6.3	预测变量表	16
6.4	数据类型检测与覆盖	16
6.5	特殊列类型参考	17
7	参数选择	18
7.1	参数预设	18
7.2	Family (分布族)	18
7.3	Method (方法)	18
7.4	Gamma (光滑度乘子)	18
7.5	交叉验证	19
7.6	Select (收缩)	19
7.7	默认基	19
7.8	默认 k (基维数)	19
7.9	张量类型	19
7.10	允许的交互	20
7.11	因子-光滑交互	20
7.12	高级参数	20
8	拟合模型	21
8.1	Fit 按钮	21
8.2	拟合过程中发生了什么	21
8.3	后台拟合与进度显示	21
9	结果标签页	22
9.1	Data (数据)	22
9.2	Equation (方程)	22
9.3	Summary (摘要)	22
9.4	Variable Importance (变量重要性)	22
9.5	贡献图	22
9.5.1	单变量光滑项 — $s(x)$	23
9.5.2	因子-光滑项 — $s(x, \text{by} = \text{factor})$	23
9.5.3	交互项 — $ti(x_1, x_2)$ 与 $te(x_1, x_2)$	23
9.5.4	参数项 — 线性项与因子项	23
9.6	Correlation (相关性)	23
9.7	Diagnostics (诊断)	23
9.8	RCA Adjustments (RCA 调整)	23

9.9 Anova (方差分析)	24
9.10 Mgcvc Output (mgcv 输出)	24
9.11 Sign Check (符号检查)	24
9.12 Concurvity (共曲性)	24
9.13 Basis Check (基检验)	24
10 下载数据	25
10.1 输出列	25
10.2 CQA 得分	25
11 RCA 计算与下载	26
11.1 打开 RCA 对话框	26
11.2 CQA 得分插值	26
11.3 输出列	26
12 销售比较网络	27
12.1 可比实例选择对话框	27
12.2 工作簿版式	27
12.3 工作表保护与残差单元格	27
13 下载报告	28
13.1 第 1 步: 生成 Quarto 报告	28
13.2 第 2 步: 转换 Quarto 报告	28
13.3 报告内容	28
14 导出模型函数	29
14.1 R 函数	29
14.2 Python 函数	29
14.3 C++ 头文件	29
14.4 SQLite 数据库	29
15 与 earthUI 和 glmnetUI 的比较	30
15.1 主要差异	30
15.2 共有功能	30
15.3 推荐的工作流程	30
16 演示数据集: Appraisal_1.csv	31
16.1 说明	31
16.2 各列	31
16.3 建议的快速上手步骤	32
17 系统要求与故障排除	33
17.1 受支持的平台	33
17.2 R 版本	33
17.3 必需的 R 软件包	33
17.4 可选依赖	33
17.5 与平台相关的说明	33
17.5.1 macOS	33
17.5.2 Windows	34
17.5.3 Linux	34
17.6 优雅降级	34
17.7 故障排除	34
17.7.1 “PDF option is not available” (PDF 选项不可用)	34

17.7.2 “Settings will not persist” (设置将不会保存)	35
17.7.3 “Diagnostics plot unavailable” (诊断图不可用)	35
17.7.4 “Report error: number of items to replace...” (报告错误: 待替换的项数)	35
17.7.5 7880 端口已被占用	35
17.7.6 在 Linux 上安装软件包失败	35

1 引言

1.1 历史与背景

1.1.1 广义可加模型

广义可加模型 (generalized additive model, GAM) 由 Trevor Hastie 与 Robert Tibshirani 在其 1990 年的专著中正式提出。GAM 对线性模型作了推广: 把每一个线性项 $\beta_j x_j$ 替换为一个光滑函数 (smooth function) $f_j(x_j)$, 于是模型成为:

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

每个 f_j 都用惩罚回归样条 (penalized regression spline) 从数据中估计得到。这些光滑函数使模型无需用户事先指定函数形式即可刻画非线性关系。惩罚项控制每个光滑项的摆动程度 (wiggliness), 在保留足够灵活性以追踪真实模式的同时防止过拟合。

1.1.2 mgcv 软件包

R 软件包 mgcv (Mixed GAM Computation Vehicle) 由巴斯大学的 Simon Wood 开发 (Wood 2017)。它于 2001 年首次发布, 如今已是 GAM 在 R 中的标准实现, 并随每一份 R 安装一同发行。其底层方法见于一系列论文, 涵盖稳定的多光滑参数估计 (Wood 2004)、薄板回归样条 (Wood 2003)、快速稳定的 REML 估计 (Wood 2011) 以及一般光滑模型的选择 (Wood, Pya, and Säfken 2016)。主要特性包括:

- 自动光滑度选择 (smoothness selection): 每一项的光滑参数都作为模型拟合的一部分加以估计, 可采用 REML (受限极大似然)、GCV (广义交叉验证) 或 ML (极大似然)。
- 多种基类型: 薄板回归样条 (tp)、三次回归样条 (cr)、P 样条 (ps) 与 B 样条 (bs), 各具不同性质。
- 张量积光滑: `te()` 与 `ti()` 可实现变量间的光滑交互, 而不受维数灾难之困。
- 变量选择: `select = TRUE` 选项会附加一个额外惩罚, 可将光滑项收缩至零, 从而实际完成变量选择。
- 共曲性诊断 (concurvity diagnostics): 度量光滑项中与多重共线性相对应的现象。

1.1.3 与 MARS (earth) 和弹性网络 (glmnet) 的比较

earth 应用程序及其两个后处理模块 (glmnet 与 mgcv) 使用不同的建模引擎。下表概括了主要差异:

特性	mgcv (GAM)	earth (MARS)	glmnet (弹性网络)
模型类型	光滑非线性 (惩罚样条)	分段线性 (自适应样条)	线性 (带正则化)
非线性	通过光滑函数自动实现 (薄板、三次等)	通过带数据驱动节点的铰链函数自动实现	仅能通过交互项或基展开实现
变量选择	可选, 通过双重惩罚收缩实现 (select=TRUE)	内置, 通过前向/后向逐步法配合 GCV 剪枝实现	内置, 通过 L_1 惩罚 (Lasso) 实现
交互项	张量积光滑 (te()、ti())	基于铰链函数的乘积 (可解释, 最高三阶)	两两交叉乘积 (难以解释)
可解释性	光滑的偏效应曲线; 非常直观	g 函数按变量给出分段线性偏效应	系数即线性权重; 易于逐一说明
最适用于	光滑的非线性关系、灵活建模	自动检测非线性与交互	高维数据、变量选择、小样本
RCA 调整	通过偏效应给出光滑调整	通过 g 函数给出分段线性调整	按变量给出线性调整
过拟合控制	REML/GCV 光滑参数估计	基于 GCV 的剪枝	交叉验证选取的 lambda

MARS (Multivariate Adaptive Regression Splines, 多元自适应回归样条) 由 Jerome Friedman 于 1991 年提出。它通过从数据中自适应地选择铰链函数及其节点 (knot) 位置来构建分段线性模型。其 R 实现是 Stephen Milborrow 编写的 **earth** 软件包。

弹性网络 (elastic net) 回归由 Friedman、Hastie 与 Tibshirani (2010) 实现于 **glmnet** 软件包中, 它把 Lasso (L_1) 惩罚与岭回归 (L_2) 惩罚结合起来, 同时完成变量选择与系数收缩。

1.1.4 earthUI-mgcvUI 流水线

一套强有力的工作流程把 earthUI 的自动节点发现与 mgcvUI 的光滑估计结合起来。当 earthUI 导出一个 .rds 结果文件而 mgcvUI 将其导入时:

1. **earth** 发现节点 ——MARS 为每个变量找出由数据驱动的铰链位置 (变点)。
2. **mgcvUI** 以其为锚点 ——earth 的节点成为 GAM 中三次回归样条 (cr 基) 的基础, 使光滑项一开始就拥有反映数据真实结构的节点位置。
3. **mgcv** 精化拟合 ——惩罚样条估计把 earth 的分段线性拟合加以光滑化, 在保留 earth 所发现的总体形状的同时, 给出光滑的偏效应曲线。

这条流水线把探索性建模 (earth) 与验证性建模 (GAM) 衔接起来, 兼取两种方法之长。

1.1.5 推荐的工作流程

对于房地产估价及其他要求模型可解释、可辩护的类似应用:

1. 先用 **earthUI** 发现数据中的重要变量、非线性关系与交互作用。earth 的自动基选择与 g 函数能提供极好的初步洞见。
2. 再用 **mgcvUI** 精化, 如果你希望得到更光滑的偏效应。把 earthUI 的节点导入 mgcvUI, 即可从分段线性模型平滑过渡到光滑模型。
3. 使用 **glmnetUI**, 当你需要激进的变量选择 (Lasso)、预测变量数目接近或超过观测数目, 或者你希望得到一个纯线性且带正则化、便于辩护的模型时。

1.2 什么是 mgcvUI?

mgcvUI 是 R 语言 mgcv 软件包的图形用户界面 (graphical user interface)。它以本地 Shiny 应用程序的形式运行——无需登录、无需服务器、也没有账户。你从 R 中启动它, 导入数据集 (CSV 或 Excel), 配置模型, 然后交互式地拟合 GAM。

自当前版本起, mgcvUI 不再是单独安装的软件包: 它随 earthUI 软件包一同发行, 作为其中的 GAM 后处理模块 (Post Processing Module), 用 `earthUI::launch_mgcv()` 启动。应用程序本身——以及本文所述的一切——并无改变; 改变的只是安装与启动命令。三个例程 (`earth`、`glmnet`、`mgcv`) 共用一棵项目树、一个设置数据库以及同一份支撑基础设施。

该应用程序提供完整的工作流程: 数据导入、含光滑项设定的变量配置、带后台处理的模型拟合、诊断图、光滑偏效应曲线, 以及可下载的 Word、PDF 或 HTML 格式报告。

广义可加模型把每个预测变量的线性关系 βx 替换为光滑函数 $f(x)$ 。这意味着:

- 无需指定函数形式——每种关系的形状由数据本身决定。
- 光滑的偏效应——每个预测变量的贡献是一条光滑曲线而非直线, 便于看清某变量在不同取值处如何影响响应变量。
- 自动光滑度选择——mgcv 通过 REML 或 GCV 确定每条曲线应有多摆动, 在拟合优度与过拟合之间取得平衡。

1.3 三种用途模式

每个项目都有一个用途 (purpose), 在侧边栏第 1 节 (**Project**, 项目) 中创建项目时选定, 此后即被锁定。当前活动项目的用途决定哪些工具与界面元素出现; 若要在别的模式下工作, 请创建或打开一个具有该用途的项目。三种模式为:

- **General** (通用)——适用于任何类型总体或数据集的 GAM 回归。这是默认模式, 提供完整的 GAM 工作流程, 不含任何领域专用的附加功能。
- **For Appraisal** (用于估价)——为房地产估价定制的 GAM 回归。增加了单一物业估价所特有的功能, 包括标的物业处理、特殊列指定以及残差约束法 (Residual Constraint Approach, RCA)。
- **Market Area Analysis** (市场区域分析)——为市场区域研究定制的 GAM 回归。增加了在既定市场中分析成组物业的功能, 并可选择处理标的物行。

在三种模式下, 核心建模引擎完全相同——你始终是在通过 `mgcv::gam()` 拟合 GAM。用途设置只控制哪些附加工具与界面元素可用。

1.4 房地产专用功能

当选定 **For Appraisal** 或 **Market Area Analysis** 时, mgcvUI 会激活若干为房地产分析而设计的功能:

- 特殊列指定——每个预测变量都可以被标记上某种特殊角色, 例如 `contract_date`、`dom`、`concessions`、`latitude`、`longitude`、`living_area`、`lot_size`、`actual_age`、`effective_age`、`area`、`site_dimensions` 或 `display_only`。这些指定决定该列在拟合与输出过程中的处理方式。
- **Sale Age** 列 (销售时距)——当某列被指定为 `contract_date` 且提供了 Effective Date (估价基准日) 时, mgcvUI 会计算一个 `sale_age` 列 (成交日期与基准日之间的天数), 并以它作为预测变量代入模型。
- **RCA** 计算 (仅估价模式)——在估价模式下, 模型拟合完成后 mgcvUI 可计算残差约束法 (RCA) 输出, 给出逐个可比实例的调整额、净调整与总调整汇总, 以及标的物业的调整后销售价格。

提示: *General* 用途模式适用于任何数据集——金融、科学、工程或房地产。估价模式与市场模式只是为房地产专业人士增添了一些便利功能。

1.5 入门

使用 mgcvUI 的步骤:

1. 在 R 中安装 **earthUI**: `install.packages("earthUI")` ——mgcv 例程已包含在内。(开发版与可选组件请参见本期 earthUI 一文中的“Installing earthUI and vProlog”一节。)
2. 启动应用程序: 在 R 中运行 `earthUI::launch_mgcv()`。应用程序会在你的网页浏览器中通过 7880 端口打开。
3. 在侧边栏第 1 节创建或打开项目。项目设定用途 (General、For Appraisal 或 Market Area Analysis) 与工作地点 (国家、州、县、市) 及其输入/输出文件夹。项目存放在共享的 `regProj` 根目录之下, 并与同族应用 earthUI 和 glmnetUI 共享。打开一个用途不同的项目, 正是切换模式的方式。
4. 在第 2 节导入数据: 从当前活动项目的输入文件夹中选择一个 CSV 或 Excel 文件。把文件放入该文件夹, 然后点击 **Refresh** (刷新) 以列出它们。
5. 从 **earthUI** 导入 (可选) ——第 3 节允许你导入一个 earthUI 结果 `.rds` 文件, 用 earth 发现的节点为 GAM 提供初值。若你没有 earthUI 结果, 跳过此步即可。
6. 配置变量 ——选择目标变量与预测变量, 设定数据类型, 指派特殊列角色, 并将变量标记为因子型或线性。
7. 设定 **mgcv** 参数 ——family、method、gamma、基类型、k、张量类型、交互项以及其他选项。
8. 拟合模型 ——点击“Fit Mgcvm GAM Model”(拟合 mgcv GAM 模型), 并在主面板中查看结果。
9. 导出 ——把预测结果下载为 Excel, 生成并转换 Quarto 报告, 或 (在估价模式下) 计算 RCA 调整额与销售比较网格。

设置会自动保存在一个 SQLite 数据库中, 并在你重新加载同一输入文件时恢复。

2 MLS 输入数据要求

对于房地产估价与市场分析工作流程而言，你的输入数据通常来自多重上市服务（Multiple Listing Service, MLS）的导出文件。本章说明预期的文件结构以及 mgcvUI 可以使用的各列。

2.1 文件格式与结构

mgcvUI 接受 CSV 与 Excel (.xlsx、.xls) 文件。导入时，列名会自动转换为 snake_case——例如，“Living SqFt”变为 living_sqft，“Contract Date”变为 contract_date，“Sale Price”变为 sale_price。这一规范化确保整个工作流程中列引用的一致性。导入时使用的 CSV 分隔符与小数点符号由区域设置决定（参见第 3 章“Locale & Regional Settings”，区域与地区设置）。

你的数据文件应是一张扁平表格，每个物业占一行，每个属性占一列。文件的第一行必须是列标题。

2.2 估价模式的必需列

虽然 mgcvUI 可以处理任意一组列，但完整的估价工作流程若具备以下各列将获益更多：

列	特殊类型	用途
Sale Price（销售价格）	（目标变量）	模型的响应变量
Contract Date（成交日期）	contract_date	用于计算 sale_age（距估价基准日的天数）
Listing Date（挂牌日期）	listing_date	在缺少 DOM 列时，与成交日期一起用于计算 DOM
Days on Market（市场停留天数）	dom	市场停留天数；在导出文件中显示
Concessions（成交让价）	concessions	销售让价；净售价 = 销售价格 - 让价
Living Area (SF)（居住面积）	living_area	启用按平方英尺计的残差（residual_sf、cqa_sf）
Lot Size（地块面积）	lot_size	宗地面积列
Site Dimensions（宗地尺寸）	site_dimensions	与地块面积归为一组
Latitude（纬度）	latitude	四舍五入至 3 位小数
Longitude（经度）	longitude	四舍五入至 3 位小数
Area ID（区域标识）	area	市场区域 / 街区标识符
Actual Age（实际年龄）	actual_age	物业年龄
Effective Age（有效年龄）	effective_age	物业的有效年龄
Address（地址）	display_only	在导出文件中显示；不进入模型

电子表格的列名可以使用外语——而“特殊”名称使用英文，以便 R 程序对其作特殊处理。除此之外，你所给定的列名会出现在回归模型、图形中，若是做估价，还会出现在输出报告中。

并非所有列都是必需的。mgcvUI 会自行适应——若某列缺失，相应功能便自动略去。不过，对于房地产定价模型，以下几列强烈推荐：

1. **Sale Age**（销售时距）——成交日期与估价基准日之间的天数。若使用跨越数年的销售历史，sale_age 往往起着核心作用。

2. **Living Area**（居住面积）——也常被称作“Living Sqft”“GLA”（gross living area，总居住面积）。
3. **Total Bath Count**（卫生间总数）——全卫、四分之一卫、半卫与四分之三卫的总数。
4. **Garage Bays**（车库车位）或 **Garage Area**（车库面积）——车库泊位数或建筑面积。
5. **Lot Size**（地块面积）——物业的土地面积。
6. **Longitude, Latitude**（经度、纬度），以及若有的话 **Area ID**（区域标识），用于刻画地理上的价格变化。

2.3 特殊列命名约定

mgcvUI 依据列的特殊类型指定而非列名来识别各列。你可以在 MLS 导出文件中随意命名列——关键在于你要在变量配置表（第 6 章）中指派正确的特殊类型。

2.4 数据质量与完整性

- **缺失值 (NA)**：任一预测变量列或目标列中含 NA 的行会在拟合前自动删除。mgcvUI 会把数据子集化为模型实际使用的那些列，因此未被使用的列中的 NA 不会导致丢行。
- **数值列**：销售价格、居住面积、地块面积等类似字段必须为数值型。
- **因子列**：文本列、TRUE/FALSE 列以及 R 因子列会被自动视为分类变量——它们只能以因子形式进入模型。一个数值列（例如数值型的区域码或风格码）只有在你勾选其 **Factor**（因子）复选框时才会成为因子；mgcvUI 绝不会根据数值的取值范围推断其为分类变量，因此像 `bath_count` 这样的离散数值预测变量仍保持连续（并被光滑化）。Special（特殊）角色与因子的判定彼此独立。

提示：导入后请查看预测变量表中的 NA 列。缺失值很多的列可能导致丢行。NA 占比 $\geq 50\%$ 的列会被标红并自动排除。

2.5 标的物行的位置

在 **Appraisal**（估价）模式下，第 1 行必须是标的物。其余各行均为可比销售实例。标的物行不参与模型拟合。拟合完成后，模型仍会为标的物行生成预测值。

在 **Market Area Analysis**（市场区域分析）模式下，是否把标的物置于第 1 行可自行决定。

在 **General**（通用）模式下没有特殊的行处理——所有行一视同仁。

3 通用模式

3.1 概述

启动 mgcvUI 时, 通用 (General Purpose) 模式是默认模式。它为任何数据集——而不仅仅是房地产数据——提供完整的 GAM 工作流程。你可以把 mgcvUI 用于科学数据、金融分析、工程研究, 或任何预期存在光滑非线性关系的回归问题。

在 General 模式下, 界面略去房地产专用功能 (特殊列、销售时距、RCA)。侧边栏被精简, 专注于变量选择、参数配置、模型拟合与导出。

3.2 侧边栏工作流程

侧边栏按编号组织为若干可折叠小节, 引导你从项目选择一路走到报告导出。多数小节只有在打开一个活动项目后才出现, 其中若干还与项目的用途相关。输出文件夹由活动项目推导得出——并没有单独的输出文件夹字段。

1. Project (项目) ——创建或打开一个项目。项目设定用途与工作地点 (国家、州、县市), 以及它的输入 (`<os>_in/`) 与输出文件夹。项目存放在共享的 `regProj` 根目录下 (默认为 `~/regProj`, 可通过 `REGPROJ_ROOT` 环境变量或 Settings 下拉菜单覆盖), 并与同族应用 `earthUI` 和 `glmnetUI` 共享, 因此三者的模型可以并列存放。点击 **Close Project** (关闭项目) 即可切换。

2. Import Data (导入数据) ——从活动项目的输入文件夹中选择一个 CSV 或 Excel 文件。先通过 Finder/资源管理器把文件放入该文件夹, 然后点击 **Refresh** (刷新) 以列出它们。对于含多个工作表的 Excel 文件, 会出现一个工作表选择器。列名会自动转换为 `snake_case`。

3. Import from earthUI (optional) (从 earthUI 导入, 可选) ——导入一个 `earthUI` 结果 `.rds` 文件, 用 `earth` 发现的节点位置为 GAM 提供初值 (同时也为变量选择提供初值)。此步骤是可选的——若你没有可导入的 `earthUI` 结果, 跳过即可。

4. Variable Configuration (变量配置) ——目标变量选择器、响应变量变换 (无、`log`、`log10`), 以及带 Include (纳入)、Factor (因子) 和 Linear (线性) 复选框的预测变量表。Special (特殊) 列只在估价模式与市场模式下出现。详见第 6 章。

5. Mgcvcall Parameters (mgcv 调用参数) ——全部模型配置: 参数预设、`family`、`method`、`gamma`、交叉验证、`select`、基类型、`k`、张量类型、交互矩阵以及高级参数。完整的参数参考见第 7 章。

6. Fit Mgcvcall GAM Model (拟合 mgcv GAM 模型) ——运行模型的按钮。结果会显示在右侧的各个标签页中。

7. Download Output (下载输出) ——把预测值、残差、CQA 得分以及各变量的贡献导出为 Excel 文件, 存入项目的输出文件夹。模型拟合之后, 在所有用途模式下均可使用。

8. Calculate RCA Adjustments & Download (计算 RCA 调整额并下载) ——为每个可比实例计算逐变量的 RCA 调整额, 并通过 CQA 对标的物残差作插值。仅在估价模式与市场模式下、且已完成拟合后出现。

9. Generate Sales Grid & Download (生成销售网格并下载) ——生成销售比较网格 (Sales Comparison Grid), 按总调整百分比推荐可比实例。仅在估价模式下出现。

10. Generate Quarto Report (生成 Quarto 报告) ——把一个自包含的 Quarto (`.qmd`) 报告包——含全部图形与表格——写入项目的输出文件夹。模型拟合之后即可使用。

11. Convert Quarto Report (转换 Quarto 报告) ——把生成的 `.qmd` 包渲染为 HTML、Word 或 PDF。PDF 需要安装 LaTeX; 若未检测到 LaTeX, PDF 选项会被隐藏。参见第 13 章。

在其他模式下, 与用途相关的步骤会被略去 (其余步骤重新编号): 市场模式下没有销售网格步骤, 通用模式下 RCA 与销售网格两个步骤都没有。

3.3 从 earthUI 导入 (可选)

侧边栏的第 3 节——**Import from earthUI**——允许你导入一个 earthUI 结果 .rds 文件。这就启用了 earth-mgcv 流水线：earth 由数据驱动得出的节点位置成为 GAM 光滑项的锚点。

导入 earthUI 结果后，mgcvUI 会：

- 显示一份摘要，列出目标变量、R 平方、项数以及每个变量的节点位置。
- 自动选中“Earth Pipeline” (earth 流水线) 参数预设 (三次回归样条基, $\gamma = 1.4$)。
- 把目标变量设为与 earth 的目标变量一致。
- 为 earth 的预测变量预先勾选 Include。
- 为 earth 的分类变量预先勾选 Factor，为 earth 的线性预测变量预先勾选 Linear。
- 预先填入 earth 检测到的交互项 (在交互矩阵中以黄色高亮显示，并锁定不允许手工更改)。
- 提供一个 **Export Knots CSV** (导出节点 CSV) 按钮，可把 earth 的节点下载为 CSV 文件。

随后，变量配置表会显示一条说明，指出变量选择 (Include、Linear、Factor) 与样条节点是由 earth 模型提供初值的。与 glmnetUI 不同——在 glmnetUI 中 glmnet 的拟合被绑定到 earth 的基上——mgcvUI 是依据你当前的选择来拟合 GAM 的，earth 只是提供了一个起步优势，而每一个控件都仍可编辑。

一个 **Clear** (清除) 按钮可移除已导入的 earth 数据，并重置为独立运行模式。

3.4 主面板标签页

数据导入之后，主面板提供以下标签页 (依赖模型的标签页在拟合完成后才有内容)：

标签页	内容
Data (数据)	已导入数据的交互式 DataTable。在估价模式与市场模式下，分为标的物业表与可比销售表。
Equation (方程)	用 MathJax 渲染的完整 GAM 公式，以及一张光滑函数定义表。
Summary (摘要)	关键指标 (R^2 、CV R^2 、已解释离差、AIC、n、光滑项个数) 以及光滑项 / 参数项表格。
Variable Importance (变量重要性)	F 统计量 (光滑项) 与 $ t $ 值 (参数项) 的水平条形图，以及排序表。
Contribution (贡献)	交互式 plotly 光滑偏效应曲线，带置信带与 earth 节点标记。
Correlation (相关性)	数值型预测变量相关系数的热力图 (拟合前即可查看)。
Diagnostics (诊断)	残差对拟合值、Q-Q 图、直方图、响应值对拟合值 (通过 gratia 生成)、实际值对预测值散点图。
RCA Adjustments (RCA 调整)	RCA 分析直方图 (估价 / 市场模式，计算完成后)。
Anova (方差分析)	参数项与光滑项合并的方差分析表。
Mgcv Output (mgcv 输出)	原始文本：计时、公式、模型摘要、family、共曲性。
Sign Check (符号检查)	earth 与 GAM 的方向一致性检查 (当已导入 earth 结果时)。
Concurvity (共曲性)	整体与两两共曲性矩阵，带颜色编码。
Basis Check (基检验)	<code>gam.check()</code> 的基维数充分性检验。

3.5 设置的持久化

mgcvUI 会自动把你的配置保存到一个 SQLite 数据库中，并以输入文件名为键。当你重新加载同一文件时，所有设置都会恢复：目标变量选择、预测变量复选框、数据类型、mgcv 参数以及交互矩阵。用途是活动项目的属性而非全局设置，最近使用的输入文件按项目分别记住，并在项目重新打开时自动加载。

3.6 暗色模式

点击右上角的月亮 / 太阳图标，可在 Nord Light 与 Nord Dark 两种主题之间切换。主题偏好保存在 localStorage 中，跨会话保持。所有界面元素都会随所选主题相应变化。

3.7 区域与地区设置

mgcvUI 通过一套基于国家的区域设置 (locale) 系统支持国际化的数字与 CSV 格式约定。标题栏中的 **Settings** (设置) 下拉菜单为 30 多个受支持国家提供 **Country** (国家) 与 **Paper** (纸张) 选择器。每种预设配置以下各项：

- **CSV 分隔符** —— 美国 / 英国 / 日本用逗号 (,)，欧洲大部分地区用分号 (;)。
- **小数点符号** —— 句点 (.) 或逗号 (,)。
- **千位分隔符** —— 逗号 (美国 / 英国 / 日本)、句点 (德国 / 意大利 / 西班牙)、空格 (芬兰 / 法国 / 波兰 / 波罗的海国家)，或撇号 (瑞士)。
- **纸张尺寸** —— Letter (美国 / 加拿大 / 墨西哥) 或 A4 (其他所有地区)。

点击 **Save** (保存) 即可把你的区域偏好全局存储。Import Data 小节还提供一个按文件生效的 **Locale** (区域) 选择器, 供采用不同约定的一次性导入使用。同一个 Settings 下拉菜单中还有 **regProj root folder** (regProj 根文件夹) 字段, 用于重新指定 mgcvUI、earthUI 与 glmnetUI 共用的项目树位置。

提示: 图形坐标轴与斜率标签上的数字格式会自动适应你的区域设置。德语区域用句点作千位分隔 (200.000), 芬兰用空格 (200 000), 瑞士用撇号 (200'000)。界面不显示任何货币符号——*mgcvUI* 与货币无关。

4 估价模式

当活动项目的用途为 **For Appraisal**（用于估价）时，mgcvUI 会把自身配置为面向单一物业估价。第 3 章所述的全部功能依然可用；本章只讲估价模式特有的附加内容。

4.1 标的物行的处理

在估价模式下，数据集的第 1 行即标的物，其余各行为可比销售实例。你的输入文件必须据此组织（参见第 2 章）。标的物的销售价格可以留空，也可以填任意值——mgcvUI 在拟合时会自动把它当作 NA 处理。

拟合完成后，模型会为标的物行生成预测值，这是 RCA 调整工作流程的基础。

4.2 估价基准日与销售时距

在估价模式与市场模式下，变量配置小节中会出现一个 **Effective Date**（估价基准日）字段（默认为当天日期）。若你在 Special 下拉菜单中把某列指定为 `contract_date`，mgcvUI 就会计算一个 `sale_age` 列——即每笔交易的成交日期与基准日之间的整天数。该列会作为预测变量加入模型。

当估价基准日改变时，`sale_age` 会自动重新计算。

4.3 特殊列指定

在估价模式与市场模式下，变量配置表中的每个预测变量都会出现一个 **Special**（特殊）下拉菜单。特殊类型及其作用的完整参考见第 6 章。

4.4 RCA 调整概览

Calculate RCA Adjustments & Download（计算 RCA 调整额并下载）按钮（在估价模式与市场模式下、拟合完成后可见）会为每个可比实例相对于标的物计算由市场推导出的调整额。完整的 RCA 工作流程见第 11 章。

提示：你为标的物指定的 *CQA* 得分决定了残差分布中有多大比例被归属于标的物。得分 5.00 会把标的物置于各可比实例的中位数位置。

5 市场区域分析模式

当活动项目的用途为 **Market Area Analysis**（市场区域分析）时，mgcvUI 提供与估价模式相同的房地产专用功能（特殊列、销售时距），但其取向是分析一组物业，而非评估单一标的物的价值。

5.1 与估价模式的差异

- 标的物行的处理更为灵活 —— 在市场模式下，第 1 行不会被自动当作标的物。
- 没有销售网格 —— “Generate Sales Grid & Download” 步骤不可用。

5.2 何时使用市场模式

在以下情形下适合使用市场区域分析模式：

- 为某个街区或市场区域建立 GAM 模型，以理解非线性的价值驱动因素
- 分析建筑面积、年龄、地块面积与区位等变量如何影响一组物业的销售价格
- 为市场状况分析或街区界定准备支撑材料
- 处理这样一份数据集：你想要特殊列功能（销售时距），但不需要销售网格工作流程

提示：市场模式对一般性的房地产回归也很有用——当你想要特殊列功能，但不需要单一物业估价工作流程时。

6 变量选择

侧边栏的第 4 节——**Variable Configuration**（变量配置）——是你选定哪些列参与模型、以及它们如何被处理的地方。

6.1 目标变量

第 4 节顶部的 **Target (response) variable**（目标 / 响应变量）下拉菜单列出数据集中的每一个数值列。选择其中一列作为响应变量（例如销售价格）。目标列会自动从预测变量列表中排除。

6.2 响应变量变换

在目标变量选择器下方，**Response Transform**（响应变量变换）下拉菜单提供三个选项：

- **None (raw)**（无，原值）——直接使用响应变量。当预测变量与响应变量在原始尺度上近似可加时，此选项合适。
- **Log (natural)**（自然对数）——在拟合前施加自然对数变换。反变换之后，这会使调整额成为比例式（%）而非绝对额（\$）。这对销售价格尤其有用，因为一笔 \$10,000 的调整在不同价格水平上含义大不相同。
- **Log10**（以 10 为底的对数）——施加以 10 为底的对数变换。

提示：对数变换会把时间调整变为比例式（%）而非绝对额（\$）。对于价格跨度很大的房地产模型，对数变换往往能产生性态更好的残差。

选择对数变换后，响应变量中 ≤ 0 的取值会被自动滤除。所有输出（预测值、贡献、残差）都会自动反变换回原始尺度。

6.3 预测变量表

在目标变量选择器下方，一张可滚动的表格列出其余所有列。复选框列使用旋转的竖排表头以求紧凑。列的顺序为：

列	说明
Variable	列名（等宽字体，过长时以省略号截断）。带边框单元格。
Type	数据类型下拉菜单：numeric、integer、character、factor、logical、Date、POSIXct。带边框下拉菜单。
Include	复选框——把该列作为预测变量纳入模型。竖排表头。
Factor	复选框——按分类因子处理（为每个水平生成一个单独的系数）。竖排表头。
Linear	复选框——强制使用线性关系（不用光滑项，只用 βx ）。竖排表头。
Special	下拉菜单（仅估价 / 市场模式）——参见下文的特殊列类型参考。带边框下拉菜单。
NAs	缺失值的个数与百分比，带颜色编码：红色（ $\geq 50\%$ ）、橙色（ $\geq 20\%$ ）、灰色（ > 0 ）。带边框单元格。

因子、光滑与线性之别：默认情况下，被纳入的数值变量会得到一个光滑项 $f(x)$ 。勾选 **Factor** 会生成一个分类项（每个水平一个系数）。勾选 **Linear** 会生成一个简单线性项 βx 以取代光滑项。同时被标记为 Factor 与 Linear 的变量按 Factor 处理。

6.4 数据类型检测与覆盖

mgcvUI 在导入时会自动检测数据类型，可识别数值、整数、逻辑、因子与日期列。你可以通过更改 **Type** 下拉菜单来覆盖任何一项检测结果。更改类型会影响该列在模型中的处理方式。

6.5 特殊列类型参考

在估价模式与市场模式下，**Special** 下拉菜单提供以下选项：

加权：

- **weight** ——观测权重列（只允许一列；权重为 0 的行不参与拟合）

日期与时间类型：

- **contract_date** ——触发根据估价基准日自动计算 **sale_age**
- **sale_age** ——标识一个已预先计算好的销售时距列（距基准日的天数），按原样使用
- **listing_date** ——在计算市场停留天数时作为备选来源
- **dom** ——标识市场停留天数列

记录类型：

- **sale_type** ——标识用于区分标的物与可比销售实例的记录类型列

货币类型：

- **concessions** ——标识销售让价（卖方补贴、买方优惠等）

面积与区位类型：

- **latitude** ——取值自动四舍五入至 3 位小数
- **longitude** ——与纬度同样的舍入处理
- **area** ——市场区域或街区标识符
- **living_area** ——启用按平方英尺计的残差计算（**residual_sf** 与 **cqa_sf**）
- **lot_size** ——宗地面积列
- **site_dimensions** ——与地块面积归为一组

年龄类型：

- **actual_age** ——物业年龄列
- **effective_age** ——物业的有效年龄

显示类型：

- **display_only** ——该列会包含在 Excel 导出中，但完全不参与模型拟合。可用于地址字段、MLS 编号、宗地编号或其他参考性数据。

提示：被指定为 *display_only* 的列仍保留在数据集中并出现在 Excel 导出里，但会完全从预测变量表中排除。可用于 ID 列、地址或其他参考字段。

7 参数选择

侧边栏的第 5 节——**Mgcv Call Parameters**（mgcv 调用参数）——提供 GAM 的全部配置选项。每个参数都带有一个蓝色帮助图标（?），其提示框给出说明。

7.1 参数预设

顶部的一个下拉菜单提供两种预设，为常见工作流程配置合理的默认值：

预设	基	设置
Standalone (discovery)（独立运行，探索）	tp（薄板）	k = 自动（10），gamma = 1.2，select = TRUE
Earth Pipeline (refinement)（earth 流水线，精化）	cr（三次回归）	k = 自动（10），gamma = 1.4，select = TRUE

导入 earthUI 节点时会自动选中 **Earth Pipeline** 预设。earth 节点的整合必须使用三次回归样条（cr）基，因为只有它允许指定确切的节点位置。

7.2 Family（分布族）

为你的响应变量选择分布族：

Family	适用情形
gaussian	连续型响应变量（例如销售价格）。最为常用。
Gamma	方差与均值成比例的正值连续型响应变量。
poisson	计数数据（例如成交笔数）。
binomial	二元结果。
inverse.gaussian	严重右偏的正值数据。

7.3 Method（方法）

光滑参数的估计方法：

Method	说明
REML	受限极大似然（默认，推荐）。对过拟合最为稳健。
GCV.Cp	广义交叉验证。倾向于选出略为复杂的模型。
ML	极大似然。与 REML 相似，但可能欠光滑。
P-REML	Pearson REML 变体。
P-ML	Pearson ML 变体。
GACV.Cp	广义近似交叉验证。

7.4 Gamma（光滑度乘子）

gamma 参数会对光滑准则中的有效自由度作乘法放大，从而促成更光滑（摆动更少）的拟合。默认值为 1.2。

取值	效果
1.0	标准光滑度
1.2–1.4	略比默认更光滑（推荐）
2.0+	光滑得多——适用于噪声较大的数据或小样本

较大的 `gamma` 值通过更重地惩罚复杂度来防范过拟合。Earth Pipeline 预设采用 1.4，以便在精化 `earth` 的节点时施加额外的光滑。

7.5 交叉验证

勾选时（默认勾选），`mgcvUI` 会在拟合后运行 10 折交叉验证（cross-validation）以计算 CV R 平方。这给出了样本外预测能力的一个诚实估计。

7.6 Select（收缩）

勾选时（默认勾选），会为每个光滑项附加一个额外惩罚，可将其收缩至零。这就实现了自动变量选择——不重要的光滑项实际上被移出模型。其实现方式是在 `gam()` 中设定 `select = TRUE`。

7.7 默认基

光滑项所用的样条基函数（basis function）类型：

基	全称	说明
tp	薄板回归样条	默认。各向同性（无需安放节点）。是 很好的通用选择。
cr	三次回归样条	允许显式安放节点。整合 <code>earth</code> 节点 时必须使用。
ps	P 样条	节点等距的 B 样条基，配以差分惩 罚。计算效率高。
bs	B 样条	灵活的 B 样条基。

7.8 默认 k（基维数）

基维数 `k` 控制每个光滑项摆动程度的上限。取值为 0（默认）表示“自动”——`mgcvUI` 会视情形取 `k = 10` 或 `earth` 节点的个数。

- 较小的 `k`（3–5）：曲线非常光滑，接近直线。
- 中等的 `k`（8–15）：灵活度适中（通常已足够）。
- 较大的 `k`（20 以上）：非常灵活——若无强正则化则有过拟合之虞。

提示： *Summary* 标签页中的有效自由度（EDF）告诉你每个光滑项实际用掉了多少自由度。若 *EDF* 接近 `k - 1`，说明基可能太小——应加大 `k`。若 *EDF* 远小于 `k`，说明该光滑项受到了良好的惩罚。

7.9 张量类型

对于连续变量之间的交互，有两种张量积类型可用：

类型	函数	说明
ti	ti(x1, x2)	张量交互——只对交互效应建模，主效应由单变量 s() 项分别处理。就可解释性而言更受推荐。
te	te(x1, x2)	张量积——对包含主效应在内的整个联合效应建模。在 RCA 调整中较难分解。

7.10 允许的交互

一个可折叠的 **Allowed Interactions**（允许的交互）小节会为所有被纳入的光滑（非线性、非因子）预测变量显示一个上三角复选框矩阵。每个复选框启用相应变量对之间的张量积光滑。

- **Allow All**（全部允许） / **Clear All**（全部清除）——批量切换按钮。
- 点击变量名 ——切换该变量的全部交互。
- 右键点击变量名 ——“Block from main effect”（从主效应中屏蔽）——把该变量标记为仅用于交互（不生成单变量光滑项）。

导入 earthUI 节点时，earth 检测到的交互会被预先勾选并锁定（以黄色背景高亮）。若 earth 模型使用了 degree = 1（无交互），则会显示一条提示信息。

警告

启用张量交互会大幅提高模型复杂度。请在你拥有领域知识、或有 earth 检测到的证据支持时才使用交互项。务必核实所得调整额是合理且可以解释的。

7.11 因子-光滑交互

当因子变量与光滑变量同时被纳入时，会出现一个单独的 **Factor-by-Smooth Interactions**（因子-光滑交互）矩阵。每个复选框会生成一个 s(x, by = factor) 项——即为每个因子水平生成一条单独的光滑曲线。

7.12 高级参数

一个可折叠的“Advanced”（高级）小节展示若干附加设置：

- **Optimizer**（优化器）——光滑参数估计所用的算法：outer/newton（默认）、outer/bfgs 或 efs（扩展 Fellner-Schall 法）。
- **Scale**（尺度参数）——尺度参数。0（默认）表示由数据估计。对于尺度已知的模型可设为一个确定值。
- **Discrete covariate method**（离散化协变量方法）——针对大型数据集的快速离散化拟合（discrete = TRUE）。
- **Threads**（线程数）——并行线程的数目（1-32）。仅在 discrete = TRUE 时有效。

8 拟合模型

8.1 Fit 按钮

侧边栏第 6 节中有 **Fit Mgcgv GAM Model** (拟合 mgcv GAM 模型) 按钮。点击它即以你当前的配置运行模型。

8.2 拟合过程中发生了什么

当你点击 Fit 时, mgcvUI 会:

1. 校验配置 —— 检查是否至少有一个预测变量、以及至少 10 个完整行。若完整行不足 50%, 会给出警告 (并列出造成大部分 NA 的那些列)。
2. 准备数据 —— 只保留与模型相关的列, 删除含 NA 的行, 应用权重 (若已指定), 排除标的物行 (在估价模式下), 并施加响应变量变换。
3. 构建 GAM 公式 —— 为数值预测变量构造光滑项 $s(x, bs="cr", k=10)$, 为分类变量构造因子项, 为被标记为 Linear 的变量构造线性项, 并为所选交互构造张量积。earth 的节点会作为显式的节点位置嵌入到 cr 基的项中。
4. 处理边界情形 —— 剔除常数变量, 把二值预测变量转换为因子, 把 k 限制在不同取值个数减 1 以内, 并在需要更多节点时用数据分位数补充 earth 的节点。
5. 拟合 GAM —— 以构造好的公式、family、method、gamma、select 及其他参数调用 `mgcv::gam()`。
6. 运行交叉验证 (若已启用) —— 10 折交叉验证, 计算样本外 R 平方。
7. 更新所有结果标签页 —— Summary、Equation、Contribution 图、Diagnostics 以及其他所有标签页都会被填充。

8.3 后台拟合与进度显示

当 callr 软件包可用时, 模型拟合会在一个后台进程中运行, 以保持应用程序的响应性。一个暗色主题的终端式进度浮层会显示:

- 计时器, 显示已耗费的秒数 / 分钟数。
- 按颜色编码的跟踪日志: 蓝色表示状态, 黄色表示交叉验证折的进度, 红色表示错误, 绿色表示结果。
- 一个 **Abort** (中止) 按钮, 可在需要时终止拟合进程。
- 一个 **Close** (关闭) 按钮, 在拟合完成后出现。

若 callr 不可用, 拟合将同步运行 (应用程序会冻结, 直到拟合完成)。

拟合成功后, Fit 按钮上会出现一个白色对勾, 同时状态行显示: “R-sq = X, CV R-sq = X, Dev = X%, AIC = X, n = X.”

9 结果标签页

9.1 Data (数据)

以交互式 DataTable 显示已导入的数据，支持横向滚动，每页 15 行。在估价模式与市场模式下，该标签页分为一张 **Subject Property** (标的物业，第 1 行) 表和一张 **Comparable Sales** (可比销售实例，其余各行) 表；在 General 模式下只显示一张表。

每个单元格都保持单行，并按列宽截断，这样很宽的自由文本字段（例如物业备注）就不会把整行拉得很长。单击可选中一行；双击任一单元格可弹出窗口，显示其完整、未经截断的内容。

9.2 Equation (方程)

分两部分显示已拟合的 GAM 模型：

模型方程（由 MathJax 渲染）：

- 左端显示响应变量及其相应的变换（log、log10）与联系函数。
- 右端显示截距值、以 f_i (变量) 形式给出的光滑项、按变量分组并标明水平数的因子项，以及带系数值的线性项。
- family 与 method 显示在方程上方。

Smooth Function Definitions (光滑函数定义) 表：

- 项的公式字符串（例如 `s(living_sqft, bs="cr", k=8)`）
- 变量名
- 类型（s、te、ti 或 linear）
- 基函数类型
- k 值
- earth 节点位置（若已导入）

9.3 Summary (摘要)

顶部横向排列六张指标卡片：

- **R-squared** —— 已解释离差的比例
- **CV R-squared** —— 交叉验证 R 平方（若已启用交叉验证）
- **Dev. Explained %** —— 已解释离差的百分比
- **AIC** —— 赤池信息准则
- **n** —— 所用观测的个数
- **Smooths** —— 光滑项的个数

卡片下方：

- **Smooth Terms** 表（光滑项）—— 项、EDF（有效自由度）、Ref.df、F 统计量、p 值
- **Parametric Terms** 表（参数项）—— 项、估计值、标准误、t 值、p 值

9.4 Variable Importance (变量重要性)

一张水平条形图，显示各模型项的相对重要性：

- 光滑项：按 F 统计量排序（蓝色条）
- 参数项：按 $|t|$ 值排序（绿色条）

图下方的 DataTable 列出项、类型、EDF、统计量与 p 值，按统计量降序排列。

9.5 贡献图

每一个对预测值有贡献的模型项都有一张交互式 **plotly** 贡献图，显示在一张带边框的卡片中。该标签页会为所有类型的项给出图形：

9.5.1 单变量光滑项 — $s(x)$

- 蓝色曲线，带 95% 置信带
- 灰色散点，显示偏残差（用于未作对数变换的模型）
- 红色竖直虚线，位于 earth 节点处（若已导入 earth 结果）
- 悬停提示：变量取值、贡献（\$）、95% 置信区间、变化率（每单位的斜率）

对于经对数变换的模型，纵轴会用公式 $\bar{y} \times (e^{f(x)} - 1)$ 反变换为以美元计的贡献，使曲线可以直接按美元数额来解读。

对于纬度/经度变量，斜率按每 0.001 度而非每 1 度给出，因为前者是更有意义的地理增量。

9.5.2 因子-光滑项 — $s(x, \text{by} = \text{factor})$

每个因子水平一条彩色曲线，叠绘在同一张图上。这有助于看清某个预测变量的效应如何随组别（例如不同街区）而变化。

9.5.3 交互项 — $ti(x_1, x_2)$ 与 $te(x_1, x_2)$

用热力图显示二维贡献曲面。色标为：红色（负）→ 白色（零）→ 蓝色（正）。悬停时同时显示两个变量的取值与贡献数额。

9.5.4 参数项 — 线性项与因子项

- 数值型线性项：数据点的散点图，叠加拟合直线
- 因子项：条形图，显示每个因子水平的平均贡献

9.6 Correlation（相关性）

一张热力图矩阵，显示所有数值型预测变量之间的 Pearson 相关系数。数据导入后即可查看——无需拟合模型。

- 颜色渐变：蓝色（-1）→ 白色（0）→ 红色（+1）
- 每个单元格中标出相关系数值
- 有助于在拟合之前识别共曲性（concurvity，多重共线性在光滑项中的对应现象）

9.7 Diagnostics（诊断）

两组诊断可视化：

诊断面板（通过 `gratia::appraise` 生成）：四张图——残差对拟合值、Q-Q 图、残差直方图，以及响应值对拟合值。

实际值对预测值散点图：观测值对预测值，附一条 45 度虚线参考线。对于经对数变换的模型，取值会被反变换回原始尺度。

9.8 RCA Adjustments（RCA 调整）

RCA 计算完成后显示三张直方图（仅限估价 / 市场模式）：

- **Residual Adjustment %**（残差调整百分比）——残差调整额占销售价格百分比的分布
- **Net Adjustment %**（净调整百分比）——净调整额的分布
- **Gross Adjustment %**（总调整百分比）——总调整额的分布

每张直方图的副标题中给出均值、中位数与标准差，并以虚线标出均值与中位数的参考位置。

9.9 Anova (方差分析)

一张合并的方差分析表，把 GAM 摘要中的参数项与光滑项汇合在一起。显示项、类型（参数项或光滑项）以及所有相关统计量。

9.10 Mgcvc Output (mgcv 输出)

原始文本输出，内容包括：

- 拟合计时
- GAM 公式
- 模型的 print 输出
- 完整的 `summary(model)` 输出
- family 与联系函数
- 整体共曲性度量

9.11 Sign Check (符号检查)

当已导入 earthUI 节点时，该标签页会把每个变量在 earth 模型中的方向与其在 GAM 中的方向作比较：

- **earth_direction**: 递增、递减或混合（依据铰链函数的符号判定）
- **gam_direction**: 递增、递减或混合（由光滑项在 100 点网格上的导数估计得出）
- **consistent**: TRUE/FALSE
- **warning**: 对任何不一致之处的说明

状态行会显示“All directions consistent”（所有方向一致）或存在不一致的变量个数。

9.12 Concurvity (共曲性)

共曲性是多重共线性在光滑项中的对应现象。该标签页显示：

- 整体共曲性表：各行分别对应 worst / observed / estimate 三种度量，各列对应每个光滑项。带颜色编码：白色 (< 0.5)、黄色 ($0.5-0.8$)、红色 (> 0.8)。
- 两两共曲性表：最坏情形的两两矩阵，采用相同的颜色编码。

“worst”一行中高于约 0.8 的取值表明，该光滑项可以被其他光滑项很好地逼近，提示存在冗余。

9.13 Basis Check (基检验)

`mgcv::gam.check()` 的输出：针对每个光滑项的基维数充分性检验。若某个光滑项的 k 过小，该检验会以一个较小的 p 值将其标示出来，提示你应当加大 k 。

10 下载数据

拟合完成后，**Download Output (Excel)**（下载输出，Excel）按钮（侧边栏第 7 节）会把含预测值与诊断信息的 Excel 文件导出到活动项目的 mgcv 输出文件夹。

10.1 输出列

列	说明
basis	反变换后的截距值（所有行相同）
<var>_contribution	各光滑项对预测值的贡献。对于对数模型，会按比例分配为美元数额。
est_<target>	模型预测值（例如 est_sale_price），若适用则已从对数尺度反变换
residual	实际值 - 预测值
calc_residual	校验用：实际值 - 各项贡献之和（反变换后）
cqa	状况-品质-吸引力（Condition-Quality-Appeal, CQA）得分（0-10 分制）
residual_sf	残差 / 居住面积（若已指定 living_area）
cqa_sf	由 residual_sf 排序计算得到的 CQA

在估价模式下，第 1 行（标的物）的 residual/cqa/cqa_sf 均置为 NA。当指定了 living_area 列时，各行按 residual_sf 降序排列。
成功完成后，下载按钮上会出现一个白色对勾。

10.2 CQA 得分

CQA 把每一行的残差与其余所有行相比较，按 0-10 分制排序打分：

- 高 CQA（约 9-10）：成交价远高于预测值——很可能状况 / 品质 / 吸引力较优
- 低 CQA（约 0-1）：成交价远低于预测值——很可能状况较差或属于困境出售
- CQA 约为 5：接近中位数

11 RCA 计算与下载

RCA (Residual Constraint Approach, 残差约束法) 工作流程在 **Appraisal** (估价) 与 **Market Area Analysis** (市场区域分析) 模式下、模型拟合之后可用。标的物业必须位于第 1 行。

11.1 打开 RCA 对话框

点击侧边栏第 8 节中的 **Calculate RCA Adjustments & Download** 按钮，会弹出一个模态对话框，其中包含：

- **Score type** (得分类型)：CQA 或 CQA per SF (若已指定 living_area)。按平方英尺计的选项会用居住面积对标的物的残差作缩放。
- **Subject CQA Score** (标的物 CQA 得分)：接受 0.00 至 10.00 之间的取值。该字段初始为空；若已导入过 earthUI 模型，你在 earthUI 中使用的得分类型与取值会从共享的项目设置中沿用过来。

点击 **Generate** (生成) 即可计算并下载。Excel 文件会写入活动项目的 mgcv 输出文件夹 (<os>_out_mgcv)。

11.2 CQA 得分插值

1. 先对各可比实例的 CQA 得分与残差排序
2. 用线性插值 (approx()) 把你给定的 CQA 值映射为一个残差
3. 标的物价值 = 模型预测值 + 插值所得残差
4. 权重为 0 的行按与标的物具有相同残差来处理

11.3 输出列

列	说明
subject_value	模型预测值 + 插值所得残差
subject_cqa	用户输入的 CQA 得分
<var>_adjustment	标的物贡献 - 可比实例贡献 (以美元计)
residual_adjustment	标的物残差 - 可比实例残差
net_adjustments	全部调整额之和
gross_adjustments	全部调整额绝对值之和
residual_adj_pct	残差调整额占销售价格的百分比
net_adj_pct	净调整额占销售价格的百分比
gross_adj_pct	总调整额占销售价格的百分比
adjusted_sale_price	实际销售价格 + 净调整额

计算成功完成后，按钮上会出现一个白色对勾。

12 销售比较网格

销售比较网格 (Sales Comparison Grid) 仅在 **Appraisal** (估价) 模式下、且已计算 RCA 调整额之后可用。它会生成一份格式化的 Excel 工作簿, 适合收入估价报告之中。

12.1 可比实例选择对话框

点击 **Generate Sales Grid & Download** 按钮会打开一个模态对话框, 列出所有可比销售实例。可比实例分为两组:

- **Recommended comps** (推荐的可比实例, 已预先勾选): 总调整额 < 销售价格的 25%, 按销售时距排序 (最近的在前)。最多 30 个。
- **Other comps** (其他可比实例, 默认不勾选): 其余的可比实例, 按总调整百分比排序。

选定你的可比实例后, 点击 **Generate Sales Grid** 即可创建工作簿。

必需的特殊类型: 必须已指定 `contract_date` 与 `living_area`。推荐的特殊类型: `latitude`、`longitude` 与 `lot_size`——没有它们网格也能生成, 但某些字段会留空。

12.2 工作簿版式

生成的工作簿含多个工作表, 每张表并排容纳 3 个可比实例:

- 表头行: 标的物与可比实例的地址、销售价格、成交日期以及市场停留天数
- 分组行: 区位 (纬度、经度、区域)、宗地面积 (地块面积、宗地尺寸) 与年龄 (实际年龄、有效年龄), 带组标题与子条目
- 模型变量行: 每个光滑项一行, 显示标的物贡献、各可比实例贡献以及调整额
- 残差特征行: 空白且未锁定的单元格, 供估价师填写模型未能刻画的特征 (例如状况、品质、景观)
- 汇总行: 净调整百分比、总调整百分比与调整后销售价格, 均带有可实时计算的 Excel 公式

12.3 工作表保护与残差单元格

每张工作表都受到保护以防误改, 但残差特征的取值单元格被显式解锁。这既让估价师能够为模型之外的特征输入取值, 又保留了由公式驱动的调整额计算。

13 下载报告

报告生成采用两步式 Quarto 工作流程，取代了早期版本中的一步式报告。

13.1 第 1 步：生成 Quarto 报告

Generate Quarto Report（生成 Quarto 报告）按钮（模型拟合后可用）会把一个自包含的 .qmd 包写入活动项目的 mgcv 输出文件夹。该包含有 Quarto 源文件、预先渲染好的图形文件（PNG 与 PDF）、一个装载了全部表格与模型元数据的 report_data.rds，以及用于 Word 输出的 reference.docx 样式模板。由于该包是自包含的，你可以在渲染之前编辑 .qmd 源文件，或把它与其他 Quarto 文档组合起来。

图形在生成该包时即已预先渲染；若某张图渲染失败，程序会写入一张占位图，包中其余部分照常生成。

13.2 第 2 步：转换 Quarto 报告

Convert Quarto Report（转换 Quarto 报告）小节通过 quarto::quarto_render() 把一个 .qmd 文件渲染为一种或多种输出格式。它默认指向最近生成的那个包，但 **Browse**（浏览）按钮允许你选择任意 .qmd 文件。可用格式有三种：

- **HTML** ——自包含的 HTML 文档。在所有平台上均可使用，无需额外软件。推荐作为默认选项。
- **Word (.docx)** ——使用随包提供的 reference.docx 样式模板渲染。适合编辑与分发。
- **PDF** ——需要安装 LaTeX（参见下文的系统要求）。若未检测到 LaTeX，PDF 选项会自动从格式复选框中隐去。

早期的一步式渲染器 render_gam_report() 与 export_gam_docx() 仍保留，可供程序化调用。

13.3 报告内容

生成的报告包括：

- **数据集与模型**：数据文件、样本量、family、method 以及拟合信息
- **结果摘要**： R^2 、 $CV R^2$ 、已解释离差、AIC 及相关的拟合指标
- **模型方程**：模型方程，其中光滑项用 f_i 记号表示、线性项给出显式系数，并附光滑函数定义。较长的方程在 PDF 中会跨多行折行。
- **摘要表**：光滑项（EDF、F、p 值）与参数项（估计值、标准误、t、p 值）
- **已启用的交互**：模型中所含张量交互项与因子-光滑交互项的矩阵（仅在存在交互项时显示）
- **变量重要性**：按 F 统计量（光滑项）或 $|t|$ （参数项）排序的条形图
- **光滑效应**（所有类型的项）：
 - 单变量光滑图（交互式图形的 ggplot2 静态版本）
 - 因子-光滑的多线条图
 - 交互热力图（ t_i 与 t_e 项）
 - 参数项图（数值型用散点图，因子型用条形图）
- **预测变量相关性热力图**
- **诊断**：残差诊断以及实际值对预测值的散点图
- **mgcv 摘要**：已拟合模型的完整文本输出
- **记号附录**：一个不编号的结尾小节，说明报告中所用的价值贡献函数记号

成功生成后，Generate Quarto Report 按钮上会出现一个白色对勾。

14 导出模型函数

mgcvUI 可以把已拟合 GAM 的光滑函数导出为多种语言的独立代码。每个光滑项都在一个 200 点网格上求值，并以带线性插值的查找表形式导出，从而使该模型可在 R 之外使用，且不依赖 mgcv。

在当前版本中，函数导出是通过程序化方式进行的。导出用的辅助函数接受一个已拟合的结果对象（由 `fit_gam()` 产生），并写出下述文件：

- `generate_r_code()`、`generate_python_code()` 与 `generate_cpp_code()` ——各语言的源文件
- `export_functions_sqlite()` ——SQLite 查找表数据库
- `export_functions_zip()` ——把这四种格式的任意组合打包成一个 zip 压缩包

14.1 R 函数

文件：gam_functions.R

- 头部注释，含响应变量、 R^2 、CV R^2 与模型公式。
- 一个 INTERCEPT 常量。
- 每个光滑变量对应一个 `approxfun()`，带预先算好的 x/y 向量。
- 线性项系数以具名常量形式给出。
- 用法：把截距、各个 `g_var(value)` 以及各个 `coef_var * value` 相加。

14.2 Python 函数

文件：gam_functions.py

- numpy 导入语句。
- INTERCEPT 常量。
- 每个变量对应：_x_var 与 _fx_var 数组，以及使用 `np.interp()` 的 `g_var(x)` 函数。
- 线性项系数以 `COEF_*` 常量形式给出。

14.3 C++ 头文件

文件：gam_functions.hpp

- 仅含头文件，使用 `#pragma once`。
- `gam_functions` 命名空间。
- 使用 `std::lower_bound` 的 `interp()` 辅助函数。
- 每个变量对应：一个带 `static const` 数组的内联函数。
- 线性项系数以 `constexpr double` 形式给出。

14.4 SQLite 数据库

文件：gam_functions.sqlite

- 表 `smooth_grids`：列为 (variable, x, fx) ——即 200 点查找表。
- 表 `model_info`：response、r_squared、cv_rsqr、dev_explained、aic、n_obs、family、method、formula。
- 表 `linear_terms`：variable、coefficient。
- 表 `intercept`：截距值。

15 与 earthUI 和 glmnetUI 的比较

earth 应用程序及其后处理模块构成一套用于回归建模的工具集。它们共用相同的数据格式、特殊列类型、RCA 工作流程与演示数据集，但使用不同的建模引擎。

15.1 主要差异

特性	mgcvUI	earthUI	glmnetUI
方法	GAM (mgcv)	MARS/Earth (earth)	弹性网络 (glmnet)
关系形态	光滑非线性	分段线性	线性
变量选择	select=TRUE 收缩	逐步剪枝	Lasso 惩罚
交互项	张量积 (te/ti)	铰链函数乘积 (2-3 阶)	两两乘积
偏效应	带置信区间的光滑曲线	g 函数 (分段)	线性斜率
earth 流水线	从 earth 导入节点	节点的来源	导入铰链函数基
代码导出	R、Python、C++、SQLite	R、Python、C++、SQLite	—
报告格式	Word、PDF、HTML	Word、PDF、HTML	Word、PDF、HTML

有一项 earthUI 独有的差异值得一提：earthUI 含有一个实验性的 Prolog/DCG 功能，可从自由文本的 MLS 备注中提取结构化的物业属性。

15.2 共有功能

三个工具都提供：

- 共享的 regProj 项目树（默认为 ~/regProj）——项目、输入文件以及按方法分列的输出在同族应用之间共享
- CSV 与 Excel 导入，并作 snake_case 列名转换
- 支持 30 多个国家的区域设置，涵盖 CSV 分隔符、小数点符号与纸张尺寸
- 特殊列指定 (contract_date、living_area、area、latitude、longitude 等)
- 依据估价基准日自动计算 sale_age
- 暗色 / 亮色模式切换 (Nord 主题)
- 设置的持久化 (SQLite 与 / 或 localStorage)
- 相关性热力图、变量重要性、诊断
- CQA 打分与 RCA 调整工作流程
- 带可比实例选择与格式化 Excel 输出的销售比较网格
- 兼容同一批演示数据集 (Appraisal_1.csv)

15.3 推荐的工作流程

- 先用 earthUI 发现重要变量、非线性关系与交互作用。earth 的自动建模提供了最佳的起点。
- 再用 mgcvUI 精化，如果你希望得到更光滑的偏效应。导入 earthUI 的节点即可无缝过渡。
- 使用 glmnetUI，当你需要一个带激进变量选择或系数符号约束的纯线性模型时。
- 在同一份数据上比较三者，以评估非线性模型所增添的复杂度是否值得。

16 演示数据集：Appraisal_1.csv

16.1 说明

该演示数据集与 earthUI 和 glmnetUI 共用。它位于 mgcvUI 源代码的 demo_mls/ 文件夹中；若已安装 earthUI，则可用：

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

该文件包含约 1,500 条住宅销售记录，取自一份模拟的 MLS 导出，标的物业位于第 1 行，可比销售实例位于其余各行。数据代表一个多区域市场中的独户住宅销售，物业面积、年龄与区位各不相同。

这并非真实数据，但以北加利福尼亚一个实际存在的街区为蓝本。所有标识信息均已改写或删除。

16.2 各列

列	类型	特殊类型	说明
sale_type	character	sale_type	用于区分标的物与可比销售实例的记录类型码
weight	numeric	weight	观测权重 (0 = 不参与拟合)
id	numeric	display_only	内部记录 ID
property_id	numeric	display_only	MLS 物业标识符
listing_id	character	display_only	MLS 挂牌编号
parcel_number	character	display_only	县估价官宗地编号 (APN)
street_address	character	display_only	物业地址
city_name	character	display_only	市
postal_code	character	display_only	邮政编码
county_name	character	display_only	县
contract_date	Date	contract_date	销售合同日期 (用于计算 sale_age)
sale_age	numeric	—	从合同日期至估价基准日的天数 (已预先计算)
coe_date	Date	display_only	交割日期
listing_status	character	display_only	挂牌状态 (例如“Sold”，已售)
sale_price	numeric	(目标变量)	销售价格——响应变量
rent	numeric	—	月租金 (用于多目标模型)
list_price	numeric	display_only	挂牌价格
original_list_price	numeric	display_only	初始挂牌价格
living_sqft	numeric	living_area	以平方英尺计的总居住面积
ppsf	numeric	display_only	每平方英尺单价
beds_total	integer	—	卧室数
baths_total	numeric	—	卫生间总数 (例如 2.5 = 2 个全卫 + 1 个半卫)
lot_size	numeric	lot_size	以平方英尺计的地块面积
area_id	integer	area	MLS 区域标识符
area_text	character	display_only	区域名称
age	numeric	actual_age	以年计的物业年龄
year_built	integer	display_only	建成年份
latitude	numeric	latitude	纬度 (入模前四舍五入至 3 位小数)
longitude	numeric	longitude	经度 (入模前四舍五入至 3 位小数)
latitude6	numeric	display_only	全精度纬度
longitude6	numeric	display_only	全精度经度
garage_spaces	integer	—	车库泊位数
fp_count	integer	—	壁炉数
no_of_stories	numeric	—	楼层数
style	character	—	建筑风格
view	character	—	景观类型 (例如“Neighborhood”街区景、“Hills”山景)
days_on_market	integer	dom	市场停留天数
listing_date	Date	listing_date	挂牌日期

列	类型	特殊类型	说明
sale_concessions	numeric	concessions	卖方让价

16.3 建议的快速上手步骤

1. 启动 mgcv 例程: `earthUI::launch_mgcv()`
2. 在第 1 节中新建一个用途为 **For Appraisal** 的项目
3. 把 Appraisal_1.csv 放入该项目的输入文件夹, 并在第 2 节中选中它 (必要时点击 **Refresh**)
4. 选择 `sale_price` 作为目标变量
5. 如需比例式调整, 可将 Response Transform 设为 **Log (natural)**
6. 按上表所示指派特殊类型
7. 纳入以下预测变量: `sale_age`、`living_sqft`、`baths_total`、`lot_size`、`area_id` (作为因子)、`age`、`latitude`、`longitude`、`garage_spaces`
8. 保留默认设置: REML、 $\gamma = 1.2$ 、tp 基、`k = 自动`、`select = TRUE`
9. 点击 **Fit Mgcvm GAM Model**
10. 查看 Summary、Contribution (光滑图) 与 Diagnostics 标签页
11. 下载输出 (第 7 步), 查看 CQA 排名
12. 以约 ~5.00 的 CQA 得分计算 RCA 调整额 (第 8 步)
13. 用推荐的可比实例生成销售比较网格 (第 9 步)

17 系统要求与故障排除

17.1 受支持的平台

mgcvUI 可在 **macOS**、**Windows** 与 **Linux** 上运行。本应用程序在 macOS 配合 RStudio 的环境中开发，也主要在该环境中测试。与平台相关的说明见下文。

17.2 R 版本

要求 $R \geq 4.1.0$ 。强烈推荐使用 RStudio Desktop (2023.06 或更高版本) ——它自带 **Quarto** 与 **pandoc** (渲染报告所需)，并为启动本应用程序提供了便利的环境。

17.3 必需的 R 软件包

以下软件包会作为依赖自动安装：

软件包	用途
shiny	Web 应用程序框架
mgcv	GAM 模型拟合
valengrCore	共享的估价内核（特殊角色、earthUI 设置沿用）
ggplot2, plotly	静态与交互式贡献图
gratia	光滑项提取与诊断
DT	交互式数据表
quarto, rmarkdown	Quarto 报告的生成与渲染
readxl	Excel 文件导入
shinyFiles	文件与文件夹浏览器
RSQLite, DBI	设置与项目数据库 (SQLite)
jsonlite	变量状态的序列化

17.4 可选依赖

组件	何时需要
LaTeX (TinyTeX、MiKTeX 或 MacTeX)	仅用于生成 PDF 报告。若未检测到 LaTeX 安装，PDF 选项会自动从格式下拉菜单中隐去。安装命令：tinytex::install_tinytex()
sysfonts + showtext	供 ggplot2 图形使用的 Roboto Condensed 字体。若不可用或处于离线状态，应用程序会自动退回到系统无衬线字体。
earth ($\geq 5.3.0$)	仅在使用 earthUI 导入流水线时需要。独立使用时不需要。
callr	后台模型拟合（保持应用程序的响应性）。没有它则拟合同步运行。
writexl, openxlsx	Excel 下载（模型输出、RCA、销售网格）。缺失时应用程序会提示。
bslib	Bootstrap 5 主题 (Nord Light/Dark)。
officer	早期的程序化 Word 导出 (export_gam_docx())。

17.5 与平台相关的说明

17.5.1 macOS

问题最少。系统库推荐使用 Homebrew 安装。若需要 PDF 报告：

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

17.5.2 Windows

配合 RStudio Desktop 运行良好。要点：

- 企业／受限机器：若临时目录或用户数据目录受到限制，设置的持久化与报告生成可能失败。应用程序会在 R 控制台中显示警告，但仍继续运行，只是不保存设置。
- **LaTeX**：从 R 中安装 TinyTeX（命令与上文 macOS 相同），或从 miktex.org 安装 MiKTeX。
- 过长的文件路径：Windows 有 260 个字符的路径长度上限。请保持工作目录路径简短。

17.5.3 Linux

各发行版之间差异最大。Ubuntu/Debian 用户在编译 R 软件包之前可能需要先装好系统库：

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

若需 PDF 报告：使用 `tinytex::install_tinytex()`，或从系统软件包管理器安装 `texlive-full`。

对于无显示设备的无头服务器，`plotly` 交互式图形在报告中会以静态图形作为退回方案渲染。Shiny 应用程序本身仍需要一个网页浏览器连接。

17.6 优雅降级

`mgcvUI` 的设计使其即便在缺少可选组件时也能工作：

缺失的组件	行为
无 LaTeX	PDF 选项从报告格式下拉菜单中隐去。HTML 与 Word 仍然可用。
无网络 / 字体加载失败	Roboto Condensed 字体由系统无衬线字体替代。控制台记录一条提示信息。
无 RSQLite / 只读文件系统	设置不会跨会话保存。控制台显示警告。除此之外应用程序正常运行。
<code>gratia::appraise</code> 失败	诊断标签页与报告中显示占位信息。可在控制台中改用 <code>mgcv::gam.check()</code> 。
临时目录不可写	报告生成失败，并给出明确的错误信息。可通过把 <code>TMPDIR</code> 环境变量设为可写位置来解决。

17.7 故障排除

17.7.1 “PDF option is not available” (PDF 选项不可用)

未检测到 LaTeX 安装。请在 R 中运行：

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

然后重启应用程序，PDF 选项便会出现。

17.7.2 “Settings will not persist” (设置将不会保存)

SQLite 数据库目录无法创建或无法写入。请检查以下位置的权限：

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

17.7.3 “Diagnostics plot unavailable” (诊断图不可用)

`gratia::appraise()` 函数在复杂模型（含许多张量交互项、非标准分布族）上偶尔会失败。这是一个已知的上游问题。变通办法：在 R 控制台中运行 `mgcv::gam.check(model)` 以直接查看诊断图。

17.7.4 “Report error: number of items to replace...” (报告错误：待替换的项数……)

这是早期版本中一个与 `gratia` 软件包相关的已知问题。请确保你安装的是最新版本的 `mgcvUI`。当前版本会捕获该错误，并在生成报告时为诊断部分放置一个占位内容。

17.7.5 7880 端口已被占用

若应用程序因 7880 端口被占用（来自上一次会话）而无法启动，请运行：

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

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

17.7.6 在 Linux 上安装软件包失败

缺少系统库。请安装上文 Linux 一节中列出的开发头文件，然后重新运行 `install.packages()`。

参考文献

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