waywiser：评估空间模型的人体工学方法

Michael J Mahoney纽约州立大学环境科学与林业学院

arXiv:2303.11312v1 [cs.MS] 20 Mar 2023

摘要

评估预测模型可能颇具挑战。建模者必须在众多评估方法中做出选择，这些方法由多个软件包实现，接口彼此不兼容，所得结果可能不同，甚至相互矛盾；与此同时，还必须确保所选方法能够恰当地估计模型在推广到新观测值时的性能。评估基于空间数据拟合的模型尤其困难，因为模型误差可能呈现空间自相关，最终用户通常会将模型预测值汇总到多个空间尺度，而且模型往往需要推广到初始训练数据边界以外的空间区域。

面向 R 语言的 waywiser 软件包旨在通过提供用于模型评估任务的人体工学工具包，让空间模型的评估更加轻松；其中包含采用统一接口的多种评估方法函数。waywiser 中的函数使用标准化的参数名称和默认值，使面向用户的接口简单易学。这些函数还经过专门设计，能够轻松集成到各种建模工作流中，接受标准类作为输入，并返回大小稳定、类型稳定的结果，确保其结果的数据类型和维度一致且可预测。此外，该软件包还提供了其他功能，使 waywiser 能够轻松配合 tidymodels 生态系统中的软件包和工作流使用。

关键词：空间建模、模型评估、适用域、R。

1. 引言

评估预测模型可能颇具挑战。建模者必须从种类繁多的评估方法中进行选择，而不同方法可能得出不同甚至相互矛盾的结果（Reich 和 Barai 1999），并确保所选方法能够在模型推广到未用于训练模型的观测值时，恰当地估计模型的表现。评估基于空间数据拟合的模型则更加复杂，因为误差在研究区域内的分布可能并非随机（Legendre 和 Fortin 1989）；预测值通常会汇总到更大的单元中，这可能会加剧已有的空间误差模式，因此需要在多个空间尺度上评估模型准确度（Riemann 等 2010年）；此外，模型通常还用于预测初始研究区域空间边界之外的区域（Meyerand Pebesma2021）。

统计软件有助于减轻评估空间数据模型时，因考量因素和方法多种多样而产生的部分复杂性。软件通过为多种成熟的评估程序提供统一的用户界面，让用户能够根据当前任务需要，更轻松地切换评估方法，从而帮助

减轻在不同任务之间切换时所承受的部分认知负担（Roehm等，2012年）。用户界面还应便于遵循科学和统计最佳实践，同时也应让方法学错误难以发生（KuhnandSilge2022）。

许多优秀的 R 软件包致力于解决模型评估中的某些具体问题，从而帮助降低复杂性并推广最佳实践。除此之外，**yardstick**（Kuhn 等人，2023年）、**metrica**（Correndo 等人，2022年）和 **hydroGOF**（Zambrano- Bigiarini，2020）等软件包提供了一系列模型评估指标，并提供标准化接口，用于根据观测值和预测值向量计算模型准确度和一致性。与此同时，**spdep**（Bivand 和 Wong，2018）和 **rgeoda**（Li 和 Anselin，2023年）等软件包提供空间自相关度量，帮助建模者评估模型误差的空间分布。最后，还有一些软件包实现了标准误差评估之外的其他评估方法；例如，**CAST**(Meyer 等人，2023年）和 **applicable**（Gotti andKuhn2022）实现了计算模型适用域的方法。

全新的**waywiser** 软件包实现了上述各种方法中的要素，同时为模型评估的各个方面提供一致且易用的用户界面。waywiserwaywiser 中的函数为多个空间建模文献中常用的评估指标提供了全新实现，并封装了用于计算空间自相关指标的函数。此外，其他函数提供了一种方法，用于评估汇总至多个空间尺度的模型预测值，并重新实现了Meyer 和 Pebesma (2021)提出的不相似性指数和适用区域。这些函数采用一致的接口，并对参数名称和定义进行了标准化，使用户能够轻松掌握软件包的用法，并根据需要在不同评估方法之间切换。

waywiserwaywiser 的输出具有类型稳定和大小稳定的特性，因此其函数的行为可预测，也易于编程使用。waywiserwaywiser 中的函数还接受标准类的输入并返回标准类的输出：空间数据使用常用的sf (Pebesma 2018)软件包中的对象，其他情况则使用简单的数据框和向量。这种可预测性以及对成熟类的依赖，使得 waywiser 易于与 R 中的大多数建模工具配合使用。此外，该软件包的多项功能使其特别容易与 tidymodels 建模框架中的软件包结合使用（**waywiser** 与R中的大多数建模工具配合使用。此外，该软件包的多项功能使其特别容易与 tidymodels 建模框架中的软件包结合使用（**waywiser** （Kuhn and Silge 2022）。例如，尽管**waywiser** 本身不提供执行交叉验证或超参数选择的函数，但其函数能够与**tune** 软件包（Kuhn2022b）自然集成，借助**rsample**（Frick等人2022年）和**spatialsample**（Mahoney 和 Silge，2023)提供的数据划分进行交叉验证模型评估，并使用**dials** (Kuhn 和 Frick 2022)实现自动化超参数选择。

本文其余部分将逐一介绍**waywiser**的各项功能：首先介绍实现（或封装实现）模型评估指标的函数（第3节），然后介绍在跨多个空间尺度汇总预测值时评估模型性能的方法（第4节），接着介绍计算模型适用域的函数（第5节）。另有一节讨论**waywiser** 如何与 tidymodels 建模框架集成（第6节）。

2. 示例数据

用于演示　为进行说明，本文将评估一个拟合于 `worldclim_simulation` 的线性模型

包含在 **waywiser** 中的数据，取自 WorldClim 生物气候变量数据集中的 10,000 个随机样本点（Fick 和 Hijmans 2017年）。变量“bio2”记录了月平均昼夜温差，“bio10”记录了一年中最暖的 3 个月的平均温度，“bio13”记录了一年中降水量最多的月份的降水量，而“bio19”记录了一年中最冷的 3 个月的降水量。最后一个变量“response”则参照 **virtualspecies**软件包的示例，使用 **CAST** (Leroy 等人 2015年；Meyer 等人2023年）进行模拟。

为了创建此模型，我们首先将数据拆分为训练集和测试集，这与标准的预测建模工作流程类似。为简单起见，我们将观测值随机分配到这些集合中；在实际操作中，最好采用空间交叉验证方法，以处理响应变量中的任何空间自相关（Mahoney 和 Silge，2023年）。

```
R> set.seed(1107)
R> data("worldclim_simulation", package = "waywiser")
R> worldclim_training <- sample(nrow(worldclim_simulation),
+                               nrow(worldclim_simulation) * 0.8)
R> worldclim_testing <- worldclim_simulation[-worldclim_training, ]
R> worldclim_training <- worldclim_simulation[worldclim_training, ]
```

随后，我们使用基础 R 的 `lm()` function 拟合线性模型，并使用所得模型为测试集生成预测值：

```
R> worldclim_model <- lm(response ~ bio2 + bio10 + bio13 + bio19,
+                        data = worldclim_training)
R> worldclim_testing$predictions <- predict(worldclim_model,
+                                           worldclim_testing)
```

3. 模型评估指标

3.1. 一致性指标

waywiser 中的多个函数都围绕计算模型一致性指标展开：这些数值指数用于衡量模型预测值（我们记作 $\hat{y}$）与另一个数据集（y）的一致程度，其中 y 通常（但并非必然）代表“真实”的测量值。这组指标通常源自空间建模文献，最常用于评估针对空间数据拟合的模型，但其计算并未纳入任何地理信息。

例如，在一系列论文中，Willmott（1980；1981年；1982）提出了一致性指标，d。该指标以预测值与观测均值之差的绝对值之和（$\hat{y}$）与观测值和观测均值之差的绝对值之和（y）的平方差和之比，表示模型预测值（$\bar{y}$）与观测值的一致性；即：

$$d = 1 - \frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{n} \left(|\hat{y}_i - \bar{y}| + |y_i - \bar{y}|\right)^2} \tag{1}$$

这一公式意味着 d 的取值范围有限， $[0,1]$, 且 d 值越高，表示 y 与$\hat{}y$之间的一致性越高。作为无量纲指标， d 是比较模型的有用工具；然而，分子中使用差值总和意味着 d 对$\hat{}$ $y_i - y_i$ 中的离群值过于敏感（Legates 和 McCabe 1999）。为解决这一问题，Willmott 等人（1985）提出了一种名为 d_1的修订指标，以差值绝对值之和取代平方差和，并且不再对分母平方：

$$d_1 = 1 - \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{\sum_{i=1}^{n} \left(|\hat{y}_i - \bar{y}| + |y_i - \bar{y}| \right)} \tag{2}$$

与 d一样， d_1 的取值范围为 $[0,1]$ ，数值越高表示一致性越好；Willmott（2011）指出， d_1 趋近于 1 的速度比 d慢，因此能够对表现良好的模型进行更精细的比较。

二十五年后，Willmott 等人重新审视了这些指数（2011），指出 d 和 d_1 的解释之所以困难，既是因为该指标的取值范围有限，也因为分母中包含了 $\hat{}y$ ，导致一致性度量的缩放因子取决于模型本身。为解决这一问题，他们提出了一个新指标 d_r，其定义如下：

$$d_r = \begin{cases} 1 - \dfrac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{c \sum_{i=1}^{n} |y_i - \bar{y}|}, \text{ when} \\ \\ \sum_{i=1}^{n} |\hat{y}_i - y_i| \leq c \sum_{i=1}^{n} |y_i - \bar{y}| \\ \\ \dfrac{c \sum_{i=1}^{n} |y_i - \bar{y}|}{\sum_{i=1}^{n} |\hat{y}_i - y_i|} - 1 \text{ otherwise} \end{cases} \tag{3}$$

其中， c 是一个取值为 2 的缩放常数。完整推导见 Willmott 等人（2011）。
与 d 和 d_1相比， d_r 的指标取值范围更大（其范围为 $[-1,1]$，其中 1 表示完全一致），且更易于直接解释； d_r 与平均绝对误差除以 c 乘以平均绝对偏差的结果成正比。

这些一致性指标都是非对称的，假设 y 值比 $\hat{}y$ 更准确。因此，这组指标适用于将模型预测值与测量值进行比较，因为通常认为用于训练模型的测量值比模型预测值更可靠。不过，在某些模型评估场景中，可能希望将 y 和 $\hat{}y$ 都视为可能包含误差。例如，比较两组不同的模型预测值时，通常不希望仅仅因为其中一组被标记为 y ，就认为它天生更准确。

因此，Ji 和 Gallo （2006年）提出了一种一致性系数，一致性系数，它是对称的，并允许 y 和$\hat{}y$ 中都存在误差：

$$\text{AC} = 1 - \frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i \right)^2}{\sum_{i=1}^{n} \left(\left| \bar{\hat{y}} - \bar{y} \right| + \left| \hat{y}_i - \bar{\hat{y}} \right| \right) \left(\left| \bar{\hat{y}} - \bar{y} \right| + |y_i - \bar{y}| \right)} \tag{4}$$

与 d 及相关指标不同，一致性系数 具有对称性：交换 y 和 $\hat{}y$ 后，得到的值相同。因此，一致性系数 是比较不同模型预测值的更合适指标，因为它不假定其中一组预测值比另一组更准确。 Ji 和 Gallo¯(2006年) 将 一致性系数 描述为有界的 $[0,1]$, ；当 $\sum_{i=1}^{n} y_i = \bar{y}$ 和 $\sum_{i=1}^{n} \hat{}y_i = \hat{}y$ （例如在零模型下）

分数项会简化为 1，因此一致性系数为 0。然而在实践中，零模型并非两个数据集彼此一致程度的真正下限；差异很大的模型可能表现不佳，使得 y_i 与 $\hat{}y_i$ 产生负的一致性系数估计值，其实际边界为（$-\infty$，1]。 y_i与 y 的一致性越差，得到的值就越小。

除了这些无量纲的一致性指标外，这些作者还提出了许多以 y为单位的模型评估指标，这些指标还可进一步分解为系统性成分和非系统性成分。例如，Willmott（1981年）详细介绍了如何将熟悉的均方误差（MSE）分解为：

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{5}$$

其系统性成分和非系统性成分，其中 MSE 的系统性成分为：

$$\mathrm{MSE}_s = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y'_i)^2 \tag{6}$$

MSE 的非系统性成分为：

$$\mathrm{MSE}_u = \frac{1}{n}\sum_{i=1}^{n}(y_i - y'_i)^2 \tag{7}$$

其中， y' 是线性回归模型 $y' = a + b\hat{}y$对 y 的预测值。这两个成分之和为 MSE：

$$\mathrm{MSE} = \mathrm{MSE}_s + \mathrm{MSE}_u \tag{8}$$

由于 MSE 的单位是 y 的平方，通常使用均方根误差（RMSE）及其系统性成分和非系统性成分更为实用；这些值通过对 MSE 及其各成分开平方计算得出。

Ji 和 Gallo (2006年) 为其 一致性系数 指标提出了类似的分解方法，使用几何平均函数关系（GMFR）模型取代线性回归，以允许 y 和 $\hat{}y$ (Draper 和 Yang 1997年) 中两者均存在误差。GMFR 的估计方式如下：

$$\begin{aligned} y' &= a + b\hat{y} \\ b &= \pm\left(\frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{\hat{y}}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}\right)^{\frac{1}{2}} \\ a &= \bar{\hat{y}} - b\bar{y} \end{aligned} \tag{9}$$

其中， b 的符号与 y 和$\hat{}y$之间的相关系数符号相同。该回归方程可以反向使用，以 y为自变量预测$\hat{}y$，记作$\hat{}y'$：

$$\hat{y}' = -\frac{a}{b} + \frac{1}{by} \tag{10}$$

Ji 和 Gallo （2006年）利用这些量将平方差和分解为系统性成分和非系统性成分，分别称为系统性乘积差之和和非系统性乘积差之和（乘积差之和）。非系统性成分乘积差之和 定义如下：

$$\text{SPD}_u = \sum_{i=1}^{n} \left[\left(\left| \hat{y}_i - \hat{y}'_i \right| \right) \left(\left| y_i - y'_i \right| \right) \right] \tag{11}$$

系统性成分可通过从平方差和中减去 SPD_u 得出：

$$\text{SPD}_s = \left(\sum_{i=1}^{n} \left(\hat{y}_i - y_i \right)^2 \right) - \left(\sum_{i=1}^{n} \left[\left(\left| \hat{y}_i - \hat{y}'_i \right| \right) \left(\left| y_i - y'_i \right| \right) \right] \right) \tag{12}$$

对这些项取算术平均值，可得非系统性和系统性平均乘积差（分别为 MPD_u 和 MPD_s）：

$$\text{MPD}_u = \frac{1}{n} \left(\text{SPD}_u \right) \tag{13}$$

$$\text{MPD}_s = \frac{1}{n} \left(\text{SPD}_s \right) \tag{14}$$

这些量可表示为 MSE 的比值，用以反映 y 与 $\hat{}y$ 之间系统性和非系统性不一致所占的比例。MPD_u 和 MPD_s 的平方根（分别为 RMPD_u 和 RMPD_s）以 y 为单位，可用于以绝对单位描述系统性和非系统性不一致。

由于 SPD_u 和 SPD_s 之和等于平方差和，因此可将 一致性系数 （方程（4））的分子分解为系统性和非系统性成分，方法是将分子替换为 SPD的相应成分：

$$\text{AC}_u = 1 - \frac{\text{SPD}_u}{\sum_{i=1}^{n} \left(\left| \bar{\hat{y}} - \bar{y} \right| + \left| \hat{y}_i - \bar{\hat{y}} \right| \right) \left(\left| \bar{\hat{y}} - \bar{y} \right| + \left| y_i - \bar{y} \right| \right)} \tag{15}$$

$$\text{AC}_s = 1 - \frac{\text{SPD}_s}{\sum_{i=1}^{n} \left(\left| \bar{\hat{y}} - \bar{y} \right| + \left| \hat{y}_i - \bar{\hat{y}} \right| \right) \left(\left| \bar{\hat{y}} - \bar{y} \right| + \left| y_i - \bar{y} \right| \right)} \tag{16}$$

这些指标均通过 **waywiser** 提供的基础设施实现，使用的是 **yardstick**（Kuhn 等人，2023年）。函数名称以 `ww_`为前缀，便于在代码编辑器中使用自动补全；所有函数的用户界面也完全一致。每个指标函数的主要版本都接受三个主要参数，即 `data` （一个继承自 `data.frame`S3 类的对象）、 `truth` （包含 y的列的名称），以及 `estimate` （包含 $\hat{}y$的列的名称）。 `truth` 和 `estimate` 均遵循整洁求值规则，用户最容易注意到的影响是，这些参数既可以接受带引号的列标识符，也可以接受不带引号的列标识符。这些函数会返回一个“tibble”（Müller 和 Wickham，2022年），包含一行和三列： `.metric` （包含指标名称）、 `.estimator` （包含字符串“standard”，以兼容 **yardstick** 及更广泛的 tidymodels 生态系统），以及 `.estimate` （包含指标估计值）。

```
R> waywiser::ww_willmott_d(data = worldclim_testing,
+                          truth = response,
+                          estimate = predictions)


# A tibble: 1 x 3
  .metric    .estimator .estimate
  <chr>      <chr>          <dbl>
1 willmott_d standard       0.919
```

当 `data` 是一个分组数据框（由 `group_by()` 函数生成）时，**dplyr**(Wickham 等人 2023a)，**waywiser** 会针对每个组分别计算指标。在这种情况下，生成的 tibble 每组占一行，并包含用于对数据分组的列，以及标准的 `.metric`、 `.estimator`和 `.estimate` 列：

```
R> worldclim_testing$group <- sample(1:2, nrow(worldclim_testing),
+                                    replace = TRUE)
R> waywiser::ww_willmott_d(data = dplyr::group_by(worldclim_testing, group),
+                          truth = response,
+                          estimate = predictions)


# A tibble: 2 x 4
  group .metric    .estimator .estimate
  <int> <chr>      <chr>          <dbl>
1     1 willmott_d standard       0.924
2     2 willmott_d standard       0.914
```

这些函数还各自有一个变体，后缀为 `_vec`，可直接接受数值向量，以计算 `truth` 和 `estimate`。这些函数会返回一个包含指标估计值的数值向量。

```
R> waywiser::ww_willmott_d_vec(truth = worldclim_testing$response,
+                               estimate = worldclim_testing$predictions)

[1] 0.9187938
```

在内部，基于数据框的函数会调用其 `_vec` 变体来计算指标，从而确保无论使用哪种接口，执行的计算都相同（因此返回的结果也相同）。

由于这些函数利用了 yardstick 包提供的基础架构，因此可以使用 yardstick 函数 `metric_set()`组合数据框变体。此函数接受任意数量的指标函数作为输入，并返回一个新函数，以便在一次调用中计算所有指标。返回的函数具有与数据框指标函数相同的用户界面，接受参数 `data`、 `truth` 和 `estimate` ，并返回一个包含列 `.metric`、 `.estimator` 和 `.estimate` 的 tibble。

```
R> metrics <- yardstick::metric_set(
+   waywiser::ww_willmott_d, waywiser::ww_willmott_d1,+   waywiser::ww_willmott_dr,
 waywiser::ww_systematic_mse,+   waywiser::ww_unsystematic_mse,
 waywiser::ww_systematic_rmse,+   waywiser::ww_unsystematic_rmse,
 waywiser::ww_agreement_coefficient,
+   waywiser::ww_systematic_agreement_coefficient,
+   waywiser::ww_unsystematic_agreement_coefficient,+   waywiser::ww_systematic_mpd,
 waywiser::ww_unsystematic_mpd,+   waywiser::ww_systematic_rmpd,
 waywiser::ww_unsystematic_rmpd)R> print(metrics(data = worldclim_testing,
+               truth = response, estimate = predictions), n = 14)
```

```
# A tibble: 14 x 3
   .metric                              .estimator .estimate
   <chr>                                <chr>          <dbl>
 1 willmott_d                           standard   0.919
 2 willmott_d1                          standard   0.729
 3 willmott_dr                          standard   0.759
 4 systematic_mse                       standard   0.0119
 5 unsystematic_mse                     standard   0.0000258
 6 systematic_rmse                      standard   0.109
 7 unsystematic_rmse                    standard   0.00508
 8 agreement_coefficient                standard   0.658
 9 systematic_agreement_coefficient     standard   0.980
10 unsystematic_agreement_coefficient   standard   0.677
11 systematic_mpd                       standard   0.000688
12 unsystematic_mpd                     standard   0.0112
13 systematic_rmpd                      standard   0.0262
14 unsystematic_rmpd                    standard   0.106
```

3.2. 自相关指标

除了提供一组一致性指标外，**waywiser** 还提供了一组用于测量模型残差中空间自相关的函数。这些函数对 **spdep** 软件包（Bivand 和 Wong 2018；Bivand 2022年；Bivand 等2008）提供的函数进行了轻量封装，这意味着（与一致性指标不同）指标计算并未直接在**waywiser**中实现。本节中的公式主要源自Bivand 和 Wong（2018）。

空间自相关指标旨在检测相邻观测值之间的数值是否比随机选取的观测值彼此更相关；也就是说，相似数值是否比随机情况下更集中或更分散。为了评估相邻观测值之间的变量关系，首先必须定义哪些观测值彼此相邻。waywiser 中的实用函数 `ww_build_neighbors()`可自动为 sf 包中的类定义邻居，不过用户通常最好更周全地计算邻居，并将得到的对象传递给函数。处理多边形几何时，



图 1：自动计算得出的法国各行政区邻接关系。行政区质心之间的连线表示邻接关系。

`ww_build_neighbors()`使用 `poly2nb()` 中的 **spdep**函数的默认行为，将共享至少一个边界点的任意多边形定义为邻居。我们可以使用 Guerry (1833年)中的标准“道德统计”数据集来直观展示这一点：

```
R> data("guerry", package = "waywiser")
R> plot(sf::st_geometry(guerry))
R> plot(waywiser::ww_build_neighbors(guerry),
+       sf::st_geometry(guerry), add = TRUE)
```

处理点几何时，waywiser 则使用 spdep 中的 `knearneigh()` 和 `knn2nb()`函数，并结合 `k = 1`，返回每个点的最近邻列表。

```
R> plot(waywiser::ww_build_neighbors(worldclim_testing),
+       sf::st_geometry(worldclim_testing))
R> plot(sf::st_geometry(worldclim_testing), add = TRUE)
```

进行计算时，必须将此邻居列表对象转换为空间权重矩阵，w。waywiser 中的另一个实用函数 `ww_build_weights()`对

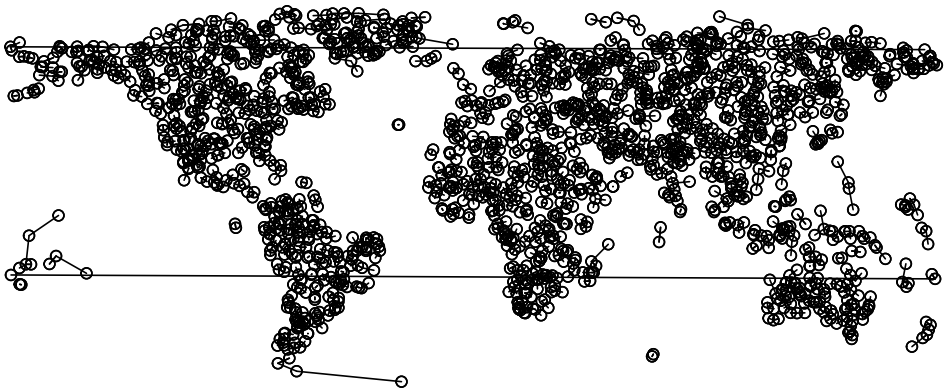

图 2：根据 WorldClim 模拟数据自动计算出的点位空间邻居。点之间的连线表示邻接关系。

spdep 的 `nb2listw()` 函数进行了轻量封装，默认生成行标准化空间权重矩阵。

```
R> waywiser::ww_build_weights(guerry)

Characteristics of weights list object:
Neighbour list object:
Number of regions: 85
Number of nonzero links: 420
Percentage nonzero weights: 5.813149
Average number of links: 4.941176

Weights style: W
Weights constants summary:
   n    nn S0      S1       S2
W 85 7225 85 37.2761 347.6683
```

空间自相关指标在计算中使用此矩阵，以估计相邻位置变量之间的关系。对于**waywiser**，通常假设该变量是模型残差，我们将其记为 x，因此对于给定观测值y_i，$x_i = y_i - \hat{}y_i$。最常用的空间自相关指标当属莫兰指数（莫兰 1950年），其定义如下：

$$I = \frac{n}{W} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(x_i - \bar{x}\right) \left(x_j - \bar{x}\right)}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} \tag{17}$$

其中，n 为观测值的数量，W 为所有w_{ij}之和，且 $i \neq j$。使用行标准化权重矩阵时，I 通常介于 $[-1, 1]$ 之间；显著大于期望值 $E(I)$ 的正值 $= -\dfrac{1}{n-1}$ 表示存在正空间自相关。

Anselin (1995) 后来进一步拓展了 I，提出了一种估计“局部” I 值的方法，可针对每个观测值进行估计，而非依赖单一的自相关统计量来代表

整个研究区域：

$$I_i = \frac{(x_i - \bar{x}) \left(\sum_{j=1}^{N} w_{ij} (x_j - \bar{x}) \right)}{m^2} \tag{18}$$

其中 $m^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}^{\,2})}{n}$(Bivand 和 Wong 2018)。

一个较少使用的替代指标是 I，即 Geary 的 c (Geary1954)，定义如下：

$$c = \frac{n-1}{2W} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - x_j)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{19}$$

与莫兰的 I一样， c 以这种方式表述的指标为整个研究区域的空间自相关提供了一个单一指标。 c 的取值大于或等于 0；相对于期望值 1 而言，较低的取值反映正空间自相关。 Anselin (1995) 以类似方式扩展了 c，提出了一种估计数据集每个观测值局部 I值的方法，进一步阐述见 Anselin (2018)：

$$c_i = \sum_{j=1}^{n} w_{ij} (x_i - x_j)^2 \tag{20}$$

waywiser 提供的另一种局部空间自相关指标是 Getis-Ord 的 G_i (Getis 和 Ord 1992；Ord 和 Getis 1995)。与提供的其他空间自相关指标一样， G_i 在 **waywiser** 中的实现函数是对 **spdep**中某个函数的轻量封装，该函数将 G_i 计算为标准偏差（Bivand 和 Wong 2018；Getis 和 Ord 1996）：

$$Z(G_i) = \frac{\left[\sum_{j=1}^{n} w_{ij} x_j \right] - \left[\sum_{j=1}^{n} w_{ij} \bar{x}_i \right]}{s_i \left[\frac{(n-1)\left(\sum_{j=1}^{n} w_{ij}^2 \right) - \left(\sum_{j=1}^{n} w_{ij} \right)^2}{n-1} \right]^{1/2}}, i \neq j \tag{21}$$

其中 $s_i = \sqrt{\left(\left(\sum_{j=1}^{n} x_j^2 \right) / (n-1) \right) - [\bar{x}_i]^2}$， $i \neq j$和 $\bar{x}_i = \left(\sum_{j=1}^{n} x_j \right) / (n-1)$ $i \neq j$。该指标的一种扩展形式 G_i^*取消了 $i \neq j$ 的要求， i 将其纳入自身的邻居，从而得到公式（ Bivand 和 Wong 2018；Getis 和 Ord 1996）：

$$Z(G_i^*) = \frac{\left[\sum_{j=1}^{n} w_{ij} x_j \right] - \left[\left(\sum_{j=1}^{n} w_{ij} \right) \bar{x}^* \right]}{s^* \left[\frac{(n-1)\left(\sum_{j=1}^{n} w_{ij}^2 \right) - \left(\sum_{j=1}^{n} w_{ij} \right)^2}{n-1} \right]^{1/2}} \tag{22}$$

其中 $s^* = \sqrt{\left(\left(\sum_{j=1}^{n} x_j^2 \right) / n \right) - \bar{x}^{*2}}$和 $\bar{x}^* = \left(\sum_{j=1}^{n} x_j \right) / n$。实际上， G_i和 G_i^*通常提供相似的信息（Getis 和 Ord 1992）。

与模型一致性指标类似，空间自相关指标也在**waywiser** 中实现，采用的是**yardstick** (Kuhn 等人，2023年)提供的基础设施，函数名称以 `ww_` 为前缀，便于自动补全。和之前一样，这些函数接受 `data`、`truth`和

`estimate` 作为参数，并返回一个 tibble，其中列为 `.metric`、 `.estimator`和 `.estimate`：

```
R> waywiser::ww_global_moran_i(worldclim_testing,
+                              truth = response,
+                              estimate = predictions)

# A tibble: 1 x 3
  .metric        .estimator .estimate
  <chr>          <chr>          <dbl>
1 global_moran_i standard       0.809
```

如上所述，waywiser 默认会使用 `ww_build_weights()` 自动创建用于计算的空间权重矩阵。为了让用户能够更改此行为，用于估计空间自相关的函数还接受一个参数 `wt`，其中可以包含用于计算的空间权重矩阵，也可以包含一个从 `data` 创建矩阵的函数：

```
R> waywiser::ww_global_geary_c(worldclim_testing,
+                               truth = response,
+                               estimate = predictions,
+                               wt = waywiser::ww_build_weights)

# A tibble: 1 x 3
  .metric        .estimator .estimate
  <chr>          <chr>          <dbl>
1 global_geary_c standard       0.159
```

由于这些函数的 `_vec` 变体不接受 `data` 参数，waywiser 无法自动创建空间权重矩阵，因此必须将其提供给 `wt` 参数：

```
R> waywiser::ww_global_geary_c_vec(
+   truth = worldclim_testing$response,
+   estimate = worldclim_testing$predictions,
+   wt = waywiser::ww_build_weights(worldclim_testing))

[1] 0.1593391
```

如前所述，waywiser 中的函数在类型和大小上都保持稳定，因此函数的输出始终具有已知的数据类型和维度。对于模型一致性指标和全局自相关统计量，这意味着 waywiser 的输出始终是仅含一行的 tibble（对于分组数据框，则每组一行）。不过，局部自相关指标的行为有所不同：局部自相关函数不会返回单行，而是返回一个 tibble，其行数与 `data` 中的观测值（或对于 `_vec` 变体，与 `truth` 和 `estimate`中的值）相同。这些估计值的排列顺序与输入数据框相同，因此可以将这些函数的输出与 `cbind()` 搭配使用，把自相关估计值关联到相应的观测值。

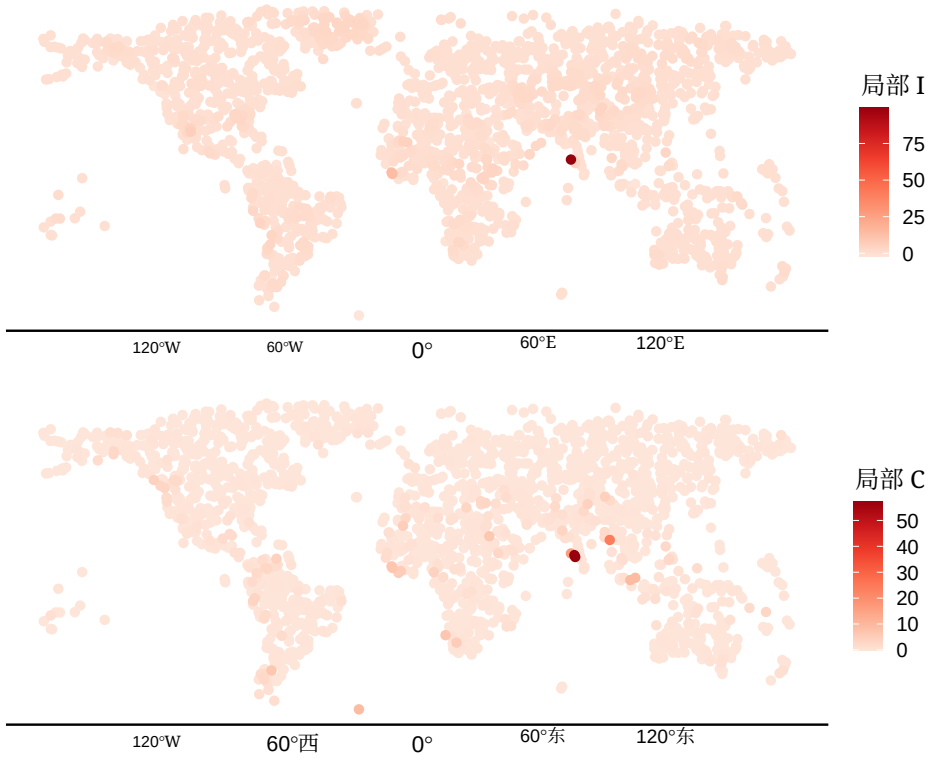

图 3：线性模型拟合 WorldClim 模拟数据所得模型残差的局部 I 和 c 值。

```
R> waywiser::ww_local_moran_i(worldclim_testing,
+                             truth = response,
+                             estimate = predictions) |> head(2)
```

```
# A tibble: 2 x 3
  .metric       .estimator .estimate
  <chr>         <chr>          <dbl>
1 local_moran_i standard       0.290
2 local_moran_i standard       4.00
```

自相关指标有助于对模型性能进行定性评估，尤其是局部统计量，与推断相比，更适用于数据探索任务（Anselin 2018）。例如，在评估拟合模拟数据的线性模型时，通过目视检查局部 I 统计量（使用 ggplot2 （Wickham 2016）和 patchwork（Pedersen 2022）绘制于图 3）可发现印度南部的残差呈现明显聚集。与此同时，局部 c 统计量也指出同一地点值得关注，同时还凸显了东南亚的其他几个地区，以及西非和南美洲的一些区域。

模型残差中存在空间自相关，表明底层模型存在设定错误。存在残差自相关的区域之间的相似性可能有助于

识别缺失的预测变量，或解释模型的不足。

4. 多尺度评估

空间模型通常用于预测感兴趣的最精细尺度，之后再根据需要将预测值聚合到更大尺度。例如，森林地上生物量模型通常用于预测各个测量样地的生物量，随后将其聚合，以估计区域或景观尺度的生物量（Johnson 等 2022；Blackard 等 2008）。同样，线虫生物量模型用于估计精细分辨率下的生物量，之后将预测值聚合以得出全球估计值（van den Hoogen 等 2019）。遗憾的是，不能假定模型性能估计在不同空间尺度上保持一致（Nelson 等 2009）。因此，必须针对每个相关的感兴趣尺度分别计算性能指标。

为此，Riemann 等 （2010年）提出了一种评估方案，用于评估模型在多个聚合尺度上的性能。该方法要求将模型预测值和测量值聚合到由规则多边形构成的多个网格中，并据此计算多个性能估计值，以评估模型性能如何随空间位置及多个尺度而变化。此方法在**waywiser** 中通过函数 `ww_multi_scale()`实现。与前文介绍的模型评估函数一样，该函数接受参数 `data`、`truth`和 `estimate`，还接受一个新参数 `metrics` ，该参数接收指标函数列表（或 `yardstick::metric_set()`的输出），以便在每个聚合尺度上进行计算。还可以通过 `...` 向 `sf::st_make_grid()` 传递其他参数，从而评估聚合到间距均匀的规则多边形网格上的预测值。例如，以下代码将 WorldClim 模型聚合到 2×2、5×5 和 10×10 网格后，计算其 RMSE 和 Willmott 的 d_1 值：

```
R> (multi_scale_output <- waywiser::ww_multi_scale(
+   worldclim_testing,+   truth = response,
+   estimate = predictions,
+   metrics = list(yardstick::rmse, waywiser::ww_willmott_d1),
+   n = c(2, 5, 10)))

# A tibble: 6 x 6
  .metric     .estimator .estimate .grid_args       .grid          .notes
  <chr>       <chr>          <dbl> <list>           <list>         <list>
1 rmse        standard      0.0299 <tibble [1 x 1]> <sf [4 x 5]>   <tibble>
2 willmott_d1 standard      0.586  <tibble [1 x 1]> <sf [4 x 5]>   <tibble>
3 rmse        standard      0.0650 <tibble [1 x 1]> <sf [25 x 5]>  <tibble>
4 willmott_d1 standard      0.759  <tibble [1 x 1]> <sf [25 x 5]>  <tibble>
5 rmse        standard      0.0761 <tibble [1 x 1]> <sf [100 x 5]> <tibble>
6 willmott_d1 standard      0.810  <tibble [1 x 1]> <sf [100 x 5]> <tibble>
```

请注意，此函数要求 `data` 是一个来自 sf 包、包含点几何的 `sf` 对象，以确保将观测值分配到各个网格单元时，能够正确处理数据的坐标参考系和单位。

`ww_multi_scale()`的输出通常会为每种指标与聚合级别的组合各生成一行。不过，与模型一致性指标函数类似，如果`data`是由`group_by()`生成的分组数据框，此函数则会分别计算每个组的指标，输出中每种唯一分组、指标与聚合级别的组合各占一行：

```
R> waywiser::ww_multi_scale(
+   dplyr::group_by(worldclim_testing, group),
+   truth = response,+   estimate = predictions,
+   metrics = list(waywiser::ww_willmott_d1),
+   n = 2)
```

```
# A tibble: 2 x 7
  group .metric     .estimator .estimate .grid_args       .grid        .notes
  <int> <chr>       <chr>          <dbl> <list>           <list>       <list>
1     1 willmott_d1 standard       0.776 <tibble [1 x 1]> <sf [8 x 6]> <tibble>
2     2 willmott_d1 standard       0.408 <tibble [1 x 1]> <sf [8 x 6]> <tibble>
```

除了熟悉的`.metric`、`.estimator`和`.estimate`列之外，`ww_multi_scale()`还会返回三个新列。`.grid_args`列是一个由tibble数据框组成的列表，来自**tibble**软件包（Müller and Wickham 2022年），其中包含通过`sf::st_make_grid()`构建网格时所使用的参数。此列可以展开，例如使用`unnest()`函数（来自**tidyr**（Wickham 等人，2023b）将这些参数作为列添加到输出表中：

```
R> tidyr::unnest(multi_scale_output, .grid_args)
```

```
# A tibble: 6 x 6
  .metric     .estimator .estimate     n .grid          .notes
  <chr>       <chr>          <dbl> <dbl> <list>         <list>
1 rmse        standard      0.0299     2 <sf [4 x 5]>   <tibble [0 x 2]>
2 willmott_d1 standard      0.586      2 <sf [4 x 5]>   <tibble [0 x 2]>
3 rmse        standard      0.0650     5 <sf [25 x 5]>  <tibble [0 x 2]>
4 willmott_d1 standard      0.759      5 <sf [25 x 5]>  <tibble [0 x 2]>
5 rmse        standard      0.0761    10 <sf [100 x 5]> <tibble [0 x 2]>
6 willmott_d1 standard      0.810     10 <sf [100 x 5]> <tibble [0 x 2]>
```

在可视化此过程的输出时，这一点尤其方便：

```
R> tidyr::unnest(multi_scale_output, .grid_args) |>
+   ggplot2::ggplot(ggplot2::aes(n^2, .estimate, color = .metric)) +
+   ggplot2::geom_line() ++   ggplot2::scale_x_continuous(
+     name = "Number of grid cells",
+     breaks = (c(2, 5, 10)^2)) ++   ggplot2::facet_wrap(~ .metric)
```

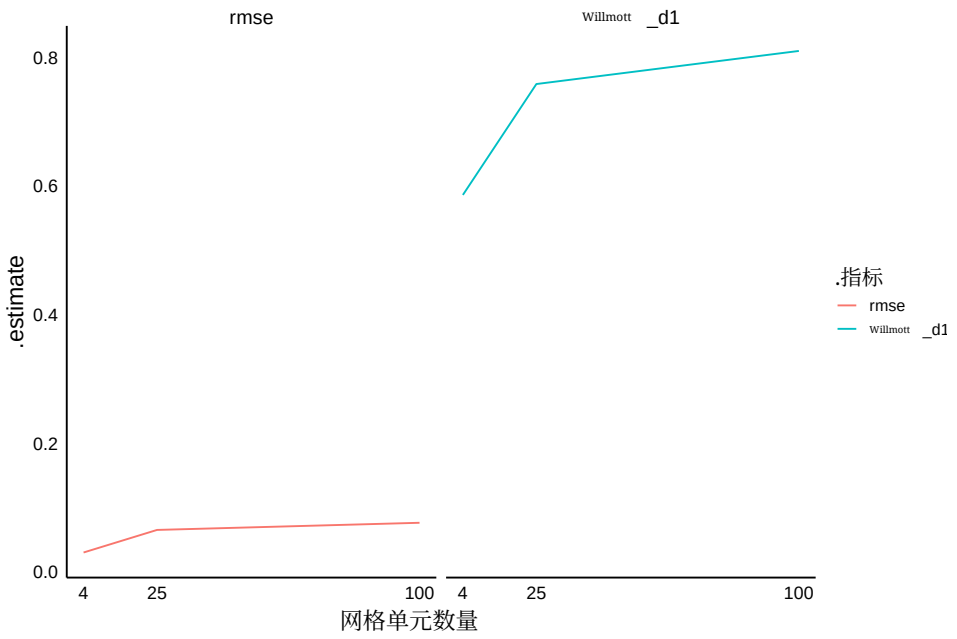

图 4：对聚合到多个空间尺度的 WorldClim 模拟数据拟合线性模型所得的 RMSE 和 Willmott d_1 估计值。“网格单元数量”指数据被聚合到的分区数量；网格单元越多，聚合粒度越细。

而 `.grid` 列是一个包含 `sf` 对象的列表，每个对象都包含用于在计算指标之前汇总预测值和观测值的实际多边形。这些对象还包含 `.truth` 和 `.estimate`的汇总值，以及多边形内 `.truth` 和 `.estimate` 各自的非缺失观测值数量。

```
R> multi_scale_output$.grid[[2]]
```

```
Simple feature collection with 4 features and 4 fieldsGeometry type: POLYGON
Dimension:     XY
Bounding box:  xmin: -178.6972 ymin: -59.39802 xmax: 179.6074 ymax: 83.85131
Geodetic CRS:  +proj=longat +datum=WGS84 +no_defs
.truth .truth count .estimate .estimate count_                          _
10.3221379 790 0.3613896 790 20.4078587 1191 0.3860225 1191 30.3148534 10
0.3539468 10 40.4041676 9 0.4096796 9geometry1 POLYGON ((-178.6972 -59.398...
2 POLYGON ((0.4550702 -59.398...3 POLYGON ((-178.6972 12.2266...
4 POLYGON ((0.4550702 12.2266...
```

这些数据通常有助于可视化误差的空间分布：

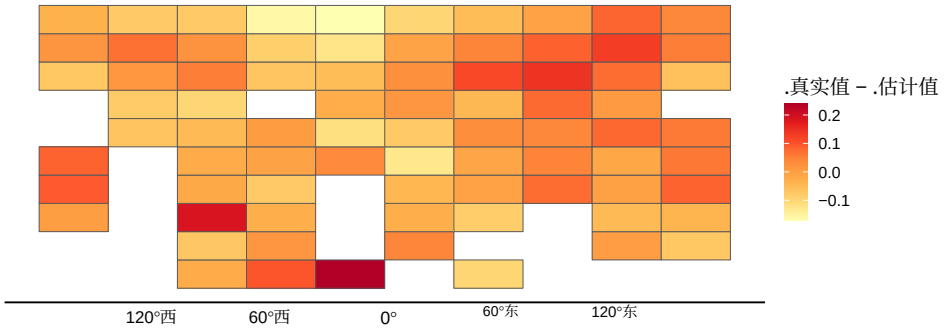

图 5：将预测值和观测值汇总到 10×10 网格后得到的模型误差。

```
R> multi_scale_output$.grid[[6]] |>
+   tidyr::drop_na() |>
+   ggplot2::ggplot(ggplot2::aes(fill = .truth - .estimate)) +
+   ggplot2::geom_sf() +
+   ggplot2::scale_fill_distiller(palette = "YlOrRd", direction = 1)
```

最后一列 `.notes` 包含诊断信息，其中包括哪些观测值（如果有）落在网格边界之外。

当模型的所有使用场景尚不明确时，将预测值和观测值系统地汇总到规则多边形网格中的方法，有助于为未来将预测值汇总到其关注边界的用户提供有用的性能估计值。不过，如果未来的使用场景已十分明确，直接汇总到这些边界后评估模型性能通常更有用；例如，除了其他评估之外，通常也适合在将预测值汇总到行政边界（如城镇或区域多边形）后评估模型性能。

因此，`ww_multi_scale()` 允许用户通过 `grids` 参数提供自己预先计算好的多边形。随后，预测值和观测值将汇总到所提供的多边形中。除此之外，函数接口和返回值均相同：

```
R> waywiser::ww_multi_scale(
+   worldclim_testing,+   truth = response,
+   estimate = predictions,
+   metrics = list(yardstick::rmse, waywiser::ww_willmott_d1),
+   grids = list(sf::st_make_grid(worldclim_testing)))
```

```
# A tibble: 2 x 6
  .metric     .estimator .estimate .grid_args       .grid          .notes
  <chr>       <chr>          <dbl> <list>           <list>         <list>
1 rmse        standard      0.0761 <tibble [0 x 0]> <sf [100 x 5]> <tibble>
2 willmott_d1 standard      0.810  <tibble [0 x 0]> <sf [100 x 5]> <tibble>
```

请注意，使用预计算多边形时，`.grid_args` 是一个有 0 行的 tibble，因为未传入任何参数来创建网格。

要求 `sf` 对象作为输入数据面临的一个挑战是，必须先将所有观测值加载到内存中，才能计算任何聚合；对于观测单元数量庞大且相对较小的建模工作（例如高分辨率景观模型或全球模型），这会带来问题。此类工作通常依赖栅格格式来提高效率，并尽可能避免将栅格转换为矢量表示。因此，`ww_multi_scale()` 也可用于使用栅格输入计算性能指标。如果 `data` 为 `NULL` 或缺失，用户可以提供 `SpatRaster` 来自**terra** 包（Hijmans 2023年）的对象，同时用于 `truth` 和 `estimate`，以进行多尺度评估。此方法通过 **exactextractr** 包（Baston 2022年）执行聚合，以提高内存和计算效率。除接口略有不同且使用 exactextractr 进行聚合外，此 `ww_multi_scale()` 方法的运行方式完全相同，返回的输出也与 `sf` 方法完全一致。

```
R> r1 <- matrix(nrow = 10, ncol = 10)
R> r1[] <- 1
R> r1 <- terra::rast(r1)
R> r2 <- matrix(nrow = 10, ncol = 10)
R> r2[] <- 2
R> r2 <- terra::rast(r2)
R> waywiser::ww_multi_scale(truth = r1, estimate = r2, n = 1)
```

```
# A tibble: 2 x 6
  .metric .estimator .estimate .grid_args       .grid        .notes
  <chr>   <chr>          <dbl> <list>           <list>       <list>
1 rmse    standard           1 <tibble [1 x 1]> <sf [1 x 5]> <tibble [0 x 2]>
2 mae     standard           1 <tibble [1 x 1]> <sf [1 x 5]> <tibble [0 x 2]>
```

5. 适用区域

waywiser 中的最后一组函数**waywiser** 旨在帮助用户评估模型能否可靠地泛化到新观测值。通过计算新观测值在预测变量空间中与用于训练模型的数据的相似程度，可以判断新观测值是否处于模型的“适用域”内，以及模型的预测值是否可能很好地代表这些观测值（Netzeva 等，2005）。虽然这些方法并非专门针对空间问题，但模型预测变量中的空间自相关会增加这样一种可能性：使用模型进行地理外推时，也需要在预测变量空间中进行外推；而且随着外推距离增加，预测效果会变差。因此，在数据稀少、难以用于模型训练和评估的区域进行预测时，这些技术尤其有用。

Meyer 和 Pebesma (2021) 提出了一种基于“不相似性指数”的新适用域方法，DI，该指数表示预测变量空间中某个点与其最近邻之间的欧几里得距离，并以用于训练模型的数据中的此类距离的平均值进行标准化。使用经过缩放和中心化、再按变量加权的变量

重要性分数，观测值 k 的 DI 计算如下：

$$\begin{aligned} d(a,b) &= \sqrt{\sum_{i=1}^{p}(a_i - b_i)^2} \\ d_k &= \arg\min_z d(k,z) \\ DI_k &= \frac{d_k}{\bar{d}} \end{aligned} \tag{23}$$

其中，p 是拟合模型时使用的预测变量数量，a_i 和 b_i 分别是观测值 a 和 b的缩放后加权预测变量值，z 是用于训练模型的数据中的观测值数量，¯和 d 是用于训练模型的数据中所有观测值对的平均 d 。

这些 DI 值本身有助于表征新观测值与用于训练模型的观测值之间的相似性，而 DI 的增加通常与预测误差的增加相关（Meyer andPebesma 2021）。Meyer 和 Pebesma (2021) 还提出了一种阈值化方法，用于计算布尔值“适用区域”（AOA）：将训练数据中大于第75百分位数 DI 值加上1.5倍四分位距的 DI 值的点定义为超出模型的 AOA，因此其预测误差可能高于测试集观测值所报告的误差。

用于计算 DI 和 AOA 的函数最初在**CAST** (Meyer 等 2023) 中实现，重点是支持使用**caret** 建模框架拟合的模型（Kuhn 2022a）。在**waywiser**中，`ww_area_of_applicability()` 函数提供了一个与框架无关的接口，用于计算 DI 和 AOA，并额外支持使用 tidymodels 框架的 workflows（第6节）（Kuhnand Silge 2022）。`ww_area_of_applicability()`的接口设计借鉴了**applicable** 程序包（GottiandKuhn 2022），并模仿了常见的模型拟合函数，例如 `lm()`。标准调用需要提供所用的模型公式、训练集和测试集，以及变量重要性分数：

```
R> (aoa <- waywiser::ww_area_of_applicability(
+   formula(worldclim_model),+   worldclim_training,
+   testing = worldclim_testing,
+   importance = vip::vi_model(worldclim_model)))
```

```
# Predictors:4
Area-of-applicability threshold:
```

0.103787

等效的调用会移除 formula 参数，改为传入训练数据和测试数据，并将其子集限定为仅包含用于拟合模型的变量：

```
R> waywiser::ww_area_of_applicability(
+   as.data.frame(worldclim_training)[1:4],
+   testing = as.data.frame(worldclim_testing)[1:4],
+   importance = vip::vi_model(worldclim_model))
```

```
# Predictors:4
Area-of-applicability threshold:
```

0.103787

如示例所示， `ww_area_of_applicability()`原生支持由**vip** 包（Greenwell 和 Boehmke 2020）返回的变量重要性分数。此 `importance` 参数也接受包含以下列的任意数据框：`term` 和 `estimate`，分别存放变量名和重要性估计值，因此可以使用任何计算变量重要性分数的方法。

与 waywiser 中的其他函数不同， `ww_area_of_applicability()` 的输出是一个自定义类“`ww_area_of_applicability`”，它继承自 hardhat 包中的“`hardhat_model`”和“`hardhat_scalar`”类。可以将此类视为模型对象，类似于 `lm()`返回的对象，因此它实现了一个 `predict()` 方法，用于计算新观测值的 DI 和 AOA。此函数返回一个包含两列的 tibble： `di`，其中包含每个新观测值的 DI 值；以及 `aoa`，这是一个布尔值，用于指示该观测值位于 AOA 范围内（`TRUE`）还是范围外（`FALSE`）：

```
R> predict(aoa, worldclim_testing)
# A tibble: 2,000 x 2
di aoa<dbl> <lgl>
1 0.0676 TRUE
2 0.109  FALSE
3 0.0264 TRUE4 0.0382 TRUE
5 0.0398 TRUE6 0.0243 TRUE
7 0      TRUE8 0.0462 TRUE
9 0.0285 TRUE
10 0.0178 TRUE# ... with 1,
990 more rows
```

包含缺失值的观测值，其 `NA`的 `di`和 `aoa`均会有值，从而保证返回的行数始终与 `newdata` 对象中的行数一致。因此，这些预测值可以通过 `cbind()` 及类似函数与预测变量值安全地合并，从而便于可视化和分析。

```
R> library("patchwork")
R> (cbind(worldclim_testing, predict(aoa, worldclim_testing)) |>
+     ggplot2::ggplot(ggplot2::aes(color = di)) +
+     ggplot2::geom_sf(alpha = 0.7) +
+     ggplot2::scale_color_distiller(palette = "Reds", direction = 1)) /
+   (cbind(worldclim_testing, predict(aoa, worldclim_testing)) |>
+      ggplot2::ggplot(ggplot2::aes(color = aoa)) +
+      ggplot2::geom_sf(alpha = 0.7))
```

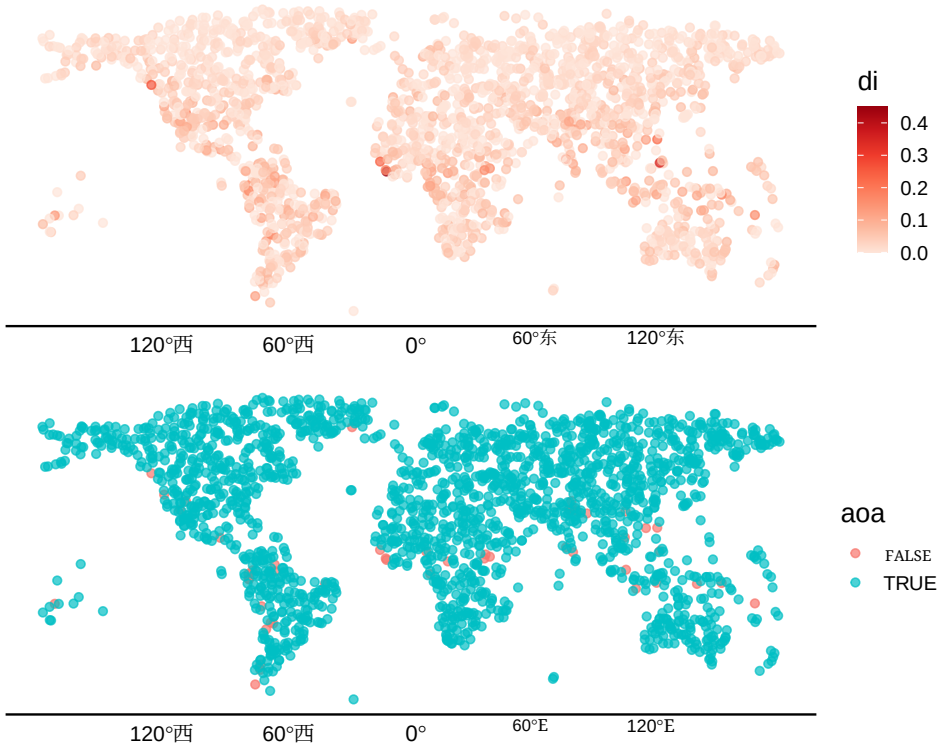

图 6： DI 以及对 WorldClim 模拟数据拟合的线性模型的 AOA。 DI 较高的区域在用于训练模型的数据中代表性较差，并且可能落在模型的 AOA 之外（下图中的“FALSE”）。

最后， `ww_area_of_applicability()` 还支持使用交叉验证的数据划分来计算 AOA，而非使用独立的训练集和测试集。任何 `rset` 对象，例如由**rsample** 或**spatialsample** 程序包（Frick 等人 2022年；Mahoney 和 Silge， 2023）生成的对象，都可用于计算 AOA。该方法将 d_k 计算为评估集中的一个点与其在分析集中的最近邻之间的距离。第6节提供了完整的演示。

使用交叉验证划分代替测试集数据时，waywiser 与 CAST 中的实现略有不同。CAST 使用所有观测值进行缩放和中心化，并将所有折中观测值之间的平均距离计算为 $\bar{d}$；而 waywiser 则针对每个交叉验证折，分别重新缩放分析集和评估集，并将仅分析集内观测值之间的平均距离计算为每折的 $\bar{d}$。这些调整旨在限制分析集与评估集之间的数据泄漏，确保评估数据不会用于确定分析数据中心化和缩放所需的参数，也不会用于计算 $\bar{d}$。在实践中，这意味着 waywiser 计算出的 d_k 略高于 CAST，因此得到的 DI_k 和阈值也略高。对于预测，waywiser 会对整个数据集进行统一缩放和中心化，并将 $\bar{d}$ 设为各折的平均值 $\bar{d}$，前提是最终使用的模型已经过重新训练

并使用整个数据集。

6. 互操作性

如前几节所示，waywiser 旨在适用于大多数空间建模工作流，无论使用何种模型或框架。waywiser 中的函数具有大小稳定性和类型稳定性，无论使用什么输入参数，都会返回数据类型相同且维度可预测的输出。函数依赖成熟的数据类型作为输入和输出，适当使用标准数据框和向量，并依赖广受欢迎的 sf 包处理空间数据类。通过依赖标准类并返回可预测的输出，waywiser 力求轻松集成到尽可能多的建模工作流中。

不过，**waywiser** 尤其适用于采用 tidymodels 建模包生态系统的工作流（Kuhn and Silge 2022）。这些包为常见的建模任务提供了一致且易用的界面，旨在让用户只需以最直接的方式使用包中的函数，就能轻松遵循统计最佳实践。 **waywiser** 中的函数经过设计，可自动兼容 tidymodels 包。为展示这一点，我们将首先使用 10 折空间聚类交叉验证方法将数据划分为多个折，该方法由 **spatialsample**（Mahoney 和 Silge，2023）实现：

```
R> library("tidymodels") |>
+   suppressPackageStartupMessages()
R> (worldclim_resamples <-
+     spatialsample::spatial_clustering_cv(worldclim_training))

#  10-fold spatial cross-validation
# A tibble: 10 x 2
   splits              id
   <list>              <chr>
 1 <split [7180/820]>  Fold01
 2 <split [7338/662]>  Fold02
 3 <split [6757/1243]> Fold03
 4 <split [7254/746]>  Fold04
 5 <split [6900/1100]> Fold05
 6 <split [7226/774]>  Fold06
 7 <split [7288/712]>  Fold07
 8 <split [7452/548]>  Fold08
 9 <split [7261/739]>  Fold09
10 <split [7344/656]>  Fold10
```

每一行 `worldclim_resamples`都包含一次交叉验证迭代，数据根据空间位置划分为分析集和评估集；每个观测值恰好被分配到一个评估集中。随后，我们可以使用 tidymodels 生态系统中的函数，分别来自 workflows、**parsnip**、**tune** 和 **yardstick** ，为每个分析集分别拟合线性模型，并使用相应的评估集对其进行评估（Vaughan 和 Couch

2022年；Kuhn 和 Vaughan 2022年；Kuhn 2022年b；Kuhn 等人 2023年）。请注意，我们可以轻松地将 **waywiser** 中的模型评估函数集成到此工作流程中，这是因为这些函数扩展了 **yardstick**的基础设施：

```
R> workflow(response ~ bio2 + bio10 + bio13 + bio19) |>
+   add_model(linear_reg()) |>
+   fit_resamples(worldclim_resamples,
+                 metrics = metric_set(rmse, mae, waywiser::ww_willmott_d1)) |>
+   collect_metrics()

# A tibble: 3 x 6
  .metric     .estimator   mean     n std_err .config
  <chr>       <chr>       <dbl> <int>   <dbl> <chr>
1 mae         standard   0.0968    10 0.00875 Preprocessor1_Model1
2 rmse        standard   0.120     10 0.0106  Preprocessor1_Model1
3 willmott_d1 standard   0.642     10 0.0465  Preprocessor1_Model1
```

我们也可以将**waywiser** 中的模型评估函数用于其他用途，例如使用**dials**包（Kuhn 和 Frick 2022年）自动评估随机森林模型的超参数：

```
R> rf_workflow <- workflow(response ~ bio2 + bio10 + bio13 + bio19) |>
+   add_model(rand_forest("regression", mtry = tune(),
+                         trees = tune(), min_n = tune()))
R> rf_parameters <- extract_parameter_set_dials(rf_workflow) |>
+   finalize(worldclim_resamples)R> rf_res <- rf_workflow |>
+   tune_grid(grid = grid_latin_hypercube(rf_parameters, size = 9),
+             resamples = worldclim_resamples,
+             metrics = metric_set(waywiser::ww_willmott_d1))
R> collect_metrics(rf_res)

# A tibble: 9 x 9
   mtry trees min_n .metric     .estimator  mean     n std_err .config
  <int> <int> <int> <chr>       <chr>      <dbl> <int>   <dbl> <chr>
1     2   971    32 willmott_d1 standard   0.935    10 0.0102  Preprocessor1_Mo~
2     1   325    17 willmott_d1 standard   0.916    10 0.0129  Preprocessor1_Mo~
3     2   482     7 willmott_d1 standard   0.945    10 0.00945 Preprocessor1_Mo~
4     2    95    26 willmott_d1 standard   0.936    10 0.0101  Preprocessor1_Mo~
5     1  1657    29 willmott_d1 standard   0.910    10 0.0144  Preprocessor1_Mo~
6     1  1522     5 willmott_d1 standard   0.926    10 0.0123  Preprocessor1_Mo~
7     2   728    23 willmott_d1 standard   0.938    10 0.00985 Preprocessor1_Mo~
8     1  1779    36 willmott_d1 standard   0.907    10 0.0143  Preprocessor1_Mo~
9     2  1134    14 willmott_d1 standard   0.942    10 0.00954 Preprocessor1_Mo~

R> select_best(rf_res)
```

```
# A tibble: 1 x 4
   mtry trees min_n .config
  <int> <int> <int> <chr>
1     2   482     7 Preprocessor1_Model3
```

找到随机森林的最优超参数后，我们便可以使用完整的训练数据集拟合调优后的模型，并使用 yardstick 和 waywiser 中的函数，将预测值与测试集进行比较评估：

```
R> tuned_rf_workflow <- rf_workflow |>
+   finalize_workflow(select_best(rf_res)) |>
+   fit(worldclim_training)
R> worldclim_testing$predictions <- predict(tuned_rf_workflow,
+                                           worldclim_testing)$.pred
R> metrics <- metric_set(waywiser::ww_willmott_d1,
+                        waywiser::ww_agreement_coefficient)
R> metrics(worldclim_testing, response, predictions)
```

```
# A tibble: 2 x 3
  .metric               .estimator .estimate
  <chr>                 <chr>          <dbl>
1 willmott_d1           standard       0.983
2 agreement_coefficient standard       0.998
```

最后，正如第 5 节所述，我们还可以使用 `ww_area_of_applicability()` 和交叉验证对象来估计此随机森林的 AOA。为此，我们首先通过 `vip::vi_permute()` 估计变量重要性分数：

```
R> d1_wrapper <- function(actual, predicted) {
+   waywiser::ww_willmott_d1_vec(actual, predicted)+ }
R> pred_wrapper <- function(object, newdata) {
+   object |>+     predict(newdata) |>
+     ranger::predictions()+ }
R> importance <- vip::vi_permute(
+   extract_fit_engine(tuned_rf_workflow),
+   train = as.data.frame(worldclim_training)[c(1:4, 6)],
+   target = "response",+   metric = d1_wrapper,
+   smaller_is_better = FALSE,
+   pred_wrapper = pred_wrapper)
```

然后，我们可以将交叉验证对象、模型公式和重要性分数传递给 `ww_area_of_applicability()`，从而计算适用区域：

```
R> waywiser::ww_area_of_applicability(
+   worldclim_resamples,
+   response ~ bio2 + bio10 + bio13 + bio19,
+   importance = importance)

# Predictors:4
Area-of-applicability threshold:
0.1791143
```

因此，waywiser 的函数旨在与 tidymodels 生态系统中的软件包自然集成，将这些软件包一致且易用的接口扩展至空间模型评估任务。不过，正如本节开头所强调的，waywiser 的函数与框架无关，也接受标准数据框和向量作为输入，并尽可能返回标准数据框和向量。这样一来，waywiser 力求与尽可能多的建模工作流实现最大程度的互操作性。

7. 结论

该**waywiser** 软件包提供了一套符合人体工学的工具包，用于评估空间模型，重点是为多种成熟的评估方法提供一致的接口。waywiser 提供的函数包括多种模型评估指标、一种在跨多个空间尺度汇总预测值时评估模型性能的方法，以及一种计算模型适用域的方法。这些函数接受标准数据框和向量作为输入，并在可能的情况下以标准数据框和向量的形式返回结果，因此能够兼容广泛的建模工作流。其他功能使其特别易于与 tidymodels 生态系统中的软件包结合使用。该软件包未来的发展方向包括新增评估指标和提升计算速度。waywiser 的发布版本可从 ComprehensiveR Archive Network（CRAN）获取，网址为 https://CRAN.R-project.org/package=waywiser；开发版本则可从 GitHub 下载，网址为 https://github.com/ropensci/waywiser。

致谢

waywiser 的初期开发得到了 Posit, PBC 的支持。waywiser 的 0.3.0 版本由 Virgilio Gómez-Rubio 博士和 Jakub Nowosad 博士为 rOpenSci 进行评审，他们的反馈极大地改进了该软件包。Lucas Johnson 就适用区域的实现和多尺度评估工作流提供了宝贵的反馈。

参考文献

Anselin L（1995）。“空间关联局部指标（LISA）。”*Geographical Analysis*，**27**（2），93–115。 doi:10.1111/j.1538-4632.1995.tb00338.x。

Anselin L（2018）。“多变量空间关联局部指标：扩展 Geary’s c。”Geographical Analysis，51（2），133–150。 doi:10.1111/gean.12164。Baston D（2022）。exactextractr：使用多边形从栅格数据集中快速提取数据。R 包版本 0.9.1，网址 https://CRAN.R-project.org/package=exactextractr。Bivand R（2022）。“用于分析空间数据的 R 包：区域数据的比较案例研究。”Geographical Analysis，54（3），488–518。 doi:10.1111/gean.12319。Bivand RS, Pebesma EJ, Gómez-Rubio V（2008）。《使用 R 进行应用空间数据分析》。Springer New York，纽约。doi:10.1007/978-0-387-78171-6。Bivand RS, Wong DWS（2018）。“全局与局部空间关联指标的实现方式比较。”TEST，27（3），716–748。doi:10.1007/s11749-018-0599-x。Blackard J, Finco M, Helmer E, Holden G, Hoppus M, Jacobs D, Lister A, Moisen G, Nelson M, Riemann R（2008）。“利用全国森林清查数据和中等分辨率信息绘制美国森林生物量地图。”《环境遥感》，112（4），1658–1677。 doi:10.1016/j.rse.2007.08.021。Correndo AA, Moro Rosso LH, Schwalbert R, Hernandez C, Bastos LM, Nieto L, Holzworth D, Ciampitti IA（2022）。metrica：预测性能指标。R 包版本 2.0.1，网址 https://CRAN.R-project.org/package=metrica。Draper NR, Yang Y（1997）。“几何平均函数关系的推广。”计算统计与数据分析，23（3），355–372。doi:10.1016/s0167-9473(96)00037-0。Fick SE, Hijmans RJ（2017）。“WorldClim 2：全球陆地区域的全新 1 公里空间分辨率气候表面。”《国际气候学杂志》，37（12），4302–4315。 doi:10.1002/joc.5086。Frick H, Chow F, Kuhn M, Mahoney M, Silge J, Wickham H（2022）。rsample：通用重抽样基础设施。R 包版本 1.1.1，网址 https://CRAN.R-project.org/package=rsample。Geary RC（1954）。“邻接比率与统计制图。”The Incorporated Statistician，5（3），115。 doi:10.2307/2986645。Getis A, Ord JK（1992）。“利用距离统计量分析空间关联。”Geographical Analysis，24（3），189–206。 doi:10.1111/j.1538-4632.1992.tb00261.x。Getis A, Ord JK（1996）。《GIS环境中的空间分析：建模》，章节“局部空间统计量概述”，第261–277页。GeoInformation International，剑桥，英国。Gotti M, Kuhn M（2022）。applicable：适用域方法汇编。R 包版本 0.1.0，网址 https://CRAN.R-project.org/package=applicable。Greenwell BM, Boehmke BC（2020）。“变量重要性图——vip 包简介。”The R Journal，12（1），343–366。doi:10.32614/RJ-2020-013。

Guerry AM（1833年）。《法国道德统计论》。Crochard，巴黎。Hijmans RJ（2023年）。terra：空间数据分析。R 软件包版本 1.7-3，网址 https://CRAN.R-project.org/package=terra。Ji L、Gallo K（2006年）。“用于图像比较的一致性系数。”《摄影测量工程与遥感》，72(7)，823–833。 doi:10.14358/pers.72.7.823。Johnson LK、Mahoney MJ、Bevilacqua E、Stehman SV、Domke GM、Beier CM（2022年）。“利用时空拼接的 Li-DAR 覆盖数据进行精细分辨率的景观尺度生物量制图。”《国际应用地球观测与地理信息学杂志》，114，103059。doi:10.1016/j.jag.2022.103059。Kuhn M（2022a）。caret：分类与回归训练。R 软件包版本 6.0-93，网址 https://CRAN.R-project.org/package=caret。Kuhn M（2022b）。tune：整洁调参工具。R 软件包版本 1.0.1，网址 https://CRAN.R-project.org/package=tune。Kuhn M、Frick H（2022年）。dials：创建调参参数值的工具。R 软件包版本 1.1.0，网址 https://CRAN.R-project.org/package=dials。Kuhn M、Silge J（2022年）。《使用 R 进行整洁建模》。O’Reilly，塞巴斯托波尔，加利福尼亚州。Kuhn M、Vaughan D（2022年）。parsnip：用于建模与分析函数的通用 API。R 软件包版本 1.0.3，网址 https://CRAN.R-project.org/package=parsnip。Kuhn M、Vaughan D、Hvitfeldt E（2023年）。yardstick：对模型性能进行整洁表征。R 软件包版本 1.2.0，网址 https://CRAN.R-project.org/package=yardstick。Legates DR、McCabe GJ（1999年）。“评估水文与水文气候模型验证中‘拟合优度’度量的使用。”《水资源研究》，35(1)，233–241。 doi:10.1029/1998wr900018。Legendre P、Fortin M（1989年）。“空间格局与生态分析。”《Vegetatio》，80(2)，107–138。doi:10.1007/bf00048036。Leroy B、Meynard CN、Bellard C、Courchamp F（2015年）。“virtualspecies：用于生成虚拟物种分布的 R 软件包。”《Ecography》。doi:10.1111/ecog.01388。Li X、Anselin L（2023年）。rgeoda：用于空间数据分析的 R 库。R 软件包版本 0.0.10-2，网址 https://CRAN.R-project.org/package=rgeoda。Mahoney M、Silge J（2023年）。spatialsample：空间重采样基础设施。R 软件包版本 0.3.0，网址 https://CRAN.R-project.org/package=spatialsample。Meyer H、Milà C、Ludwig M（2023年）。CAST：用于时空模型的“caret”应用。R 软件包版本 0.7.1，网址 https://CRAN.R-project.org/package=CAST。Meyer H、Pebesma E（2021年）。“对未知空间进行预测？估计空间预测模型的适用区域。”《生态学与进化方法》，12(9)，1620–1633。 doi:10.1111/2041-210x.13650。

Moran PAP（1950年）。“连续随机现象札记。”《生物统计学》，37(1/2)，17。doi:10.2307/2332142。Müller K、Wickham H（2022年）。tibble：简单数据框。R 软件包版本 3.1.8，网址https://CRAN.R-project.org/package=tibble。Nelson MD、McRoberts RE、Holden GR、Bauer ME（2009年）。“卫星影像空间聚合与分辨率对森林土地面积估计值的影响。”《国际遥感杂志》，30(8)，1913–1940。doi:10.1080/01431160802545631。Netzeva TI、Worth AP、Aldenberg T、Benigni R、Cronin MT、Gramatica P、Jaworska JS、Kahn S、Klopman G、Marchant CA、Myatt G、Nikolova-Jeliazkova N、Patlewicz GY、Perkins R、Roberts DW、Schultz TW、Stanton DT、van de Sandt JJ、Tong W、Veith G、Yang C（2005年）。“（定量）构效关系适用域定义方法的现状。”《实验动物替代方案》，33(2)，155–173。doi:10.1177/026119290503300209。Ord JK、Getis A（1995）。“局部空间自相关统计量：分布问题及其应用。”Geographical Analysis，27(4)，286–306。doi:10.1111/j.1538-4632.1995.tb00912.x。Pebesma E（2018）。“R 的简单要素：为空间矢量数据提供标准化支持。”The R Journal，10(1)，439–446。doi:10.32614/RJ-2018-009。Pedersen TL（2022年）。patchwork：图形组合器。R 软件包版本 1.1.2，网址https://CRAN.R-project.org/package=patchwork。Reich Y、Barai S（1999年）。“评估用于工程问题的机器学习模型。”《工程中的人工智能》，13(3)，257–272。doi:10.1016/s0954-1810(98)00021-1。Riemann R、Wilson BT、Lister A、Parks S（2010年）。“利用美国林务局森林清查与分析（FIA）数据，对森林特征的连续地理空间数据集进行有效评估的方案。”《环境遥感》，114(10)，2337–2352。 doi:10.1016/j.rse.2010.05.010。Roehm T、Tiarks R、Koschke R、Maalej W（2012年）。“专业开发人员如何理解软件？”2012年第34届国际软件工程会议（ICSE）。doi:10.1109/icse.2012.6227188。网址 http://dx.doi.org/10.1109/ICSE.2012.6227188。van den Hoogen J、Geisen S、Routh D、Ferris H、Traunspurger W、Wardle DA、de Goede RGM、Adams BJ、Ahmad W、Andriuzzi WS、Bardgett RD、Bonkowski M、Campos-Herrera R、Cares JE、Caruso T、de Brito Caixeta L、Chen X、Costa SR、Creamer R、Mauro da Cunha Castro J、Dam M、Djigal D、Escuer M、Griffiths BS、Gutiérrez C、Hohberg K、Kalinkina D、Kardol P、Kergunteuil A、Korthals G、Krashevska V、Kudrin AA、Li Q、Liang W、Magilton M、Marais M、Martín J、Matveeva E、Mayad EH、Mulder C、Mullin P、Neilson R、Nguyen TAD、Nielsen UN、Okada H、Rius JEP、Pan K、Peneva V、Pellissier L、Carlos Pereira da Silva J、Pitteloud C、Powers TO、Powers K、Quist CW、Rasmann S、Moreno S、Scheu S、Setälä H、Sushchuk A、Tiunov AV、Trap J、van der Putten W、Vestergård M,

Villenave C, Waeyenberge L, Wall DH, Wilschut R, Wright DG, Yang Ji, Cro wther TW（2019）。“全球范围内的土壤线虫丰度与功能群组成”。《自然》，572(7768)，194–198。 doi:10.1038/s41586-019-1418-6。Vaughan D, Couch S（2022）。workflows：建模工作流。R 软件包版本 1.1.2，URL https://CRAN.R-project.org/package=workflows。Wickham H（2016）。ggplot2：数据分析的优雅图形。施普林格纽约出版社。ISBN 978-3-319-24277-4。URL https://ggplot2.tidyverse.org。Wickham H, François R, Henry L, Müller K, Vaughan D（2023a）。dplyr：数据处理语法。R 软件包版本 1.1.0，URL https://CRAN.R-project.org/package=dplyr。Wickham H, Vaughan D, Girlich M（2023b）。tidyr：整理杂乱数据。R 软件包版本 1.3.0，URL https://CRAN.R-project.org/package=tidyr。Willmott CJ（1981）。“论模型验证”。《自然地理学》，2(2)，184–194。doi:10.1080/02723646.1981.10642213。Willmott CJ（1982）。“关于模型性能评估的一些评论”。《美国气象学会通报》，63(11)，1309–1313。 doi:10.1175/1520-0477(1982)063<1309:scoteo>2.0.co;2。Willmott CJ, Ackleson SG, Davis RE, Feddema JJ, Klink KM, Legates DR, O’Donnell J, Rowe CM（1985）。“模型评估与比较统计”。《地球物理研究杂志》，90(C5)，8995。doi:10.1029/jc090ic05p08995。Willmott CJ, Robeson SM, Matsuura K（2011）。“改进的模型性能指数”。《国际气候学杂志》，32(13)，2088–2094。 doi:10.1002/joc.2419。Willmott CJ, Wicks DE（1980）。“一种在加利福尼亚州境内对月降水量进行空间插值的经验方法”。《自然地理学》，1(1)，59–73。 doi:10.1080/02723646.1980.10642189。Zambrano-Bigiarini M（2020）。hydroGOF：用于比较模拟与观测水文时间序列的拟合优度函数。 doi:10.5281/zenodo.839854。R 软件包版本 0.4-0，URL https://github.com/hzambran/hydroGOF。

所属机构：
Michael J Mahoney 纽约州立大学环境科学与林业学院环境科学研究生项目
林业大道 1 号 美国纽约州锡拉丘兹 电子邮箱： mjmahone@esf.edu 网址：
https://mm218.dev/