CERES：用于急性粮食不安全的概率早期预警系统

Tom Danny S. Pedersen

Northflow Technologies AS，挪威斯塔万格

ceres@northflow.no

https://ceres.northflow.no
https://github.com/northflowlabs/ceres

2026年3月 arXiv：cs.AI（主分类）/
stat.AP （交叉分类）

作者注：本文按照正式学术惯例使用“我们”。本文为单作者作品；参见第 9.8 节。

利益冲突：作者是 Northflow Technologies AS 的创始人，该公司提供 CERES 商业机构订阅服务。本文所述的公共应用程序接口和所有数据均可在 ceres.northflow.no 上免费访问，无需订阅。

摘要

我们介绍 CERES（校准早期预警与风险估计系统——其中“校准”指系统在设计上旨在实现经验概率校准，并非已证明的属性；见第1节），这是一种用于急性粮食不安全的自动化概率预测系统。CERES 每周更新一次，为全球43个高风险国家生成未来90天内 IPC第3+ 阶段（危机）、第4+ 阶段（紧急状态）和第5阶段（饥荒）状况的概率估计。该系统通过逻辑评分模型融合六种数据流——降水异常（CHIRPS）、植被指数（MODIS归一化植被指数）、冲突事件（武装冲突地点与事件数据项目）、IPC分类、粮食消费得分（世界粮食计划署）和谷物价格指数（联合国粮食及农业组织/世界粮食计划署）——模型采用作者指定的初始系数和参数化输入扰动区间（n=2,000次抽样）。在针对四起因数据完整性而选定的 IPC第4至第5阶段事件进行历史回溯验证时，CERES 在所有四个案例中均被归为 TIER-1；这些仅是样本内合理性检查，并非对前瞻性表现的声明。所有前瞻性预测均带有时间戳、经密码学标识，并存档以供公众在 T+90 期限与综合粮食安全阶段分类结果数据进行验证。据作者所知，CERES 是首个同时具备以下特点的饥荒早期预警系统：(1) 概率化；(2) 开放访问；(3) 持续运行；(4) 在预测层面机器可读；(5) 承诺对作出的每项预测进行公开前瞻性验证。

1. 引言

急性粮食不安全仍是最有可能预防的大规模人道主义危机之一。自2004年以来，综合粮食安全阶段分类（IPC）系统为严重程度提供了共同术语（《IPC技术手册》，2021），但其发布周期——通常每个国家每3至6个月发布一次，且需要实地评估团队参与——意味着，情况迅速恶化时，人们往往要等到资源预先部署的关键窗口期已过，才发现问题（Maxwell 与 Haan，2021；Hillbruner 与 Moloney，2012）。

现有早期预警系统以不同方式弥补这一缺口。由美国国际开发署自 1985年 起资助的饥荒早期预警系统网络（FEWS NET），为约 30 个国家提供基于情景的粮食安全展望，综合运用遥感、市场监测和分析师的专业判断（Verdin 等，2005）。世界粮食计划署饥饿地图通过移动调查近乎实时地监测粮食获取情况（世界粮食计划署脆弱性分析与制图，2025）。欧盟联合研究中心的 INFORM风险指数每年汇总静态脆弱性指标（De Groeve 等，2015）。这些系统都不会输出概率估计——即一个介于 0 和 1 之间、可以根据结果进行检验并通过反馈加以改进的数值（Gneiting 和 Raftery，2007）。

这有两方面的重要意义。第一，概率估计使决策者能够根据风险程度按比例投入资源，而不必等到跨越二元分类阈值才采取行动——联合国人道主义事务协调厅预见性行动框架明确提出了这一要求（联合国人道主义事务协调厅，2023）。第二，只有在结果尚未知晓时就作出预测，并系统地将其与结果进行比较，才能实现经验校准——也就是说，预测概率为 70% 的事件，实际发生的比例应约为 70%。我们在理想目标的意义上使用“经过校准”一词：CERES 的设计目标是在积累足够的前瞻性数据后实现校准；经证实的校准结果将在后续的经验验证论文中报告（第 9.7 节）。目前没有任何现有的饥荒早期预警系统，能够完整公开并以机器可读的形式记录每一项预测及观察到的每一个结果，而这正是进行校准测量的前提。

CERES 旨在填补这一空白。它输出的是概率分布，而非情景。它会在发布所有点估计的同时，提供敏感性区间。系统生成的每项预测都会被前瞻性存档、标注时间戳，并与其在 T+90 天时的验证结果关联。系统在方法学上的核心贡献并非预测模型本身——逻辑回归风险模型在人道主义文献中已相当成熟（Coughlan de Perez 等，2016）——而是公开的前瞻性验证基础设施，使经验校准能够得到衡量，并随着时间推移不断改进。

本文介绍 CERES架构（第 2 节）、信号提取与评分方法（第 3 节）、警报等级分类（第 4 节）、验证与评级框架（包括预注册校准方案，第 5 节）、与三种基线进行历史回溯验证（第 6 节）、实际部署（第 7 节）、伦理考量（第 8 节）、局限性与未来工作（第 9 节）、与现有系统的比较（第 10 节），以及结论（第 11 节）。完整的模型系数见附录 C。

2. 系统架构

CERES 采用七阶段流水线，由自动调度器（Railway 定时任务，每周一 06:00 UTC）每周执行。流水线会生成饥荒假设对象（结构化、带版本的 JSON 文档），并将其写入持久化存储，通过 REST API 提供服务，同时归档至 SQLite 运行历史数据库。完整的实现细节和特征提取公式见附录 C 和开放代码仓库（github.com/northflowlabs/ceres）；以下内容介绍设计原理。

2.1 流水线阶段

阶段 1 - 摄取	：通过适配器注册表拉取所有数据源
阶段 2 - 标准化	：标准化为 0.25 度网格 + 每周频率
阶段 3 - 信号	：提取各区域标准化的危机信号
阶段 4 - 汇聚	：检测多源异常汇聚
阶段 5 - 评分	：作者指定的逻辑斯蒂评分+ 扰动区间
第6阶段 - 假设	：排序归入FamineHypothesis DSL对象
第7阶段 - 发布	：写入 JSON、Parquet、档案、PDF报告

阶段 1-2 的流水线故障会中止执行；阶段 3-7 的故障不会导致程序终止，但会记录日志。在 2 核 CPU 云实例上，处理来自 6 个数据源的 43 个地区约需 45-90 秒。

2.2 数据源

来源	提供方	信号类型	频率
CHIRPS v2.0	加州大学圣巴巴拉分校气候危害小组	降水异常（z分数）	旬
MODIS MOD13A3	NASA Earthdata	NDVI异常（z分数）	每月
武装冲突地点与事件数据项目	武装冲突地点和事件数据项目	冲突事件和死亡人数	每周
综合粮食安全阶段分类	IPC全球平台	当前及预测IPC阶段	不定期
食物消费评分	世界粮食计划署脆弱性分析与制图	食物消费得分和减少应对策略指数调查得分	定期
谷物价格指数	FAO / WFP	市场价格压力层级	每月

所有数据均通过类型化适配器接口摄取，该接口将原始 API 响应映射到标准 CERES 架构。所有数据框均标准化到 0.25 度空间网格：网格产品（CHIRPS、MODIS）采用双线性插值，点源数据采用空间连接（Funk 等人，2015；Didan，2021）。时间对齐采用以星期一为起始日的 ISO 周为基准的每周频率，频率较低的数据源则向前填充。

2.3 运行归档

每次流水线运行都会向持久化的 SQLite 表 run_snapshots 中为每个区域写入一行，记录完整的概率向量、敏感性区间、等级分类、驱动因素构成和运行时间戳。该档案是所有前瞻性校准分析的基准真值，并为 ceres.northflow.no/tracker 上公开的历史表现记录提供支持。

3. 信号提取与评分

本节概述信号提取与评分逻辑。完整的系数表和特征变换公式见附录 C。完整实现可在开放代码仓库（github.com/northflowlabs/ceres）中获取。

3.1 区域级信号提取

对于每个区域和参考周，CERES 会在 [0,1] 中提取六项归一化压力评分：根据 CHIRPS 降水异常 Z 分数计算的干旱胁迫；根据 MODIS归一化植被指数异常 Z 分数计算的植被压力；根据武装冲突地点与事件数据项目的事件数量和死亡人数计算的冲突压力，并以区域第 95 百分位基线进行标准化，再按 0.6/0.4 加权，以体现事件数量能提供更早期信号、而死亡人数能提供更强烈的严重程度信号这一发现（Raleigh 等人，2010）；将当前 IPC阶段线性映射至 [0,1] 得出的 IPC 压力；根据世界粮食计划署粮食消费评分和减少应对策略指数计算的粮食获取压力，并按与 IPC 对齐的阈值进行标准化（世界粮食计划署脆弱性分析与制图，2025）；以及根据谷物价格指数相对于 2015—2020年基线的变化计算的价格压力（联合国粮食及农业组织，2024）。

综合压力评分 (CSS) 是对可用信号进行加权平均所得：IPC_压力 0.25、冲突_压力 0.20、干旱_压力 0.20、粮食_获取_压力 0.15、植被_压力 0.10、价格_压力 0.10；再除以覆盖率_因子进行标准化，该因子等于参考周内可用信号所占的比例。这些权重为作者指定值，符合 IPC 阶段作为直接严重程度指标的既定首要地位（《IPC技术手册》，2021），也体现了粮食安全文献所记录的气候、冲突和市场驱动因素的相对贡献；这些权重并非直接源自《IPC技术手册》，并将随着前瞻性数据的积累接受实证修订。对于某个参考周内六种信号中少于三种可用的地区，应用程序接口响应会将其标记为低覆盖率，并强制采用更宽的扰动范围（sigma = 0.25，而非 0.15），以反映更高的输入不确定性；输出不会被抑制，但建议用户在处理

低覆盖率假设时更加谨慎。

3.2 多源汇聚检测

当两个或更多独立信号支柱同时超过其升高的阈值（区域 z 分数 > 1.5）时，即标记为汇聚事件。汇聚等级：CRITICAL（三个或更多支柱）、WARNING（两个支柱）、WATCH（一个支柱）、NONE。作为逻辑回归模型的连续输入时，这些等级会映射为数值评分：NONE = 0.0、WATCH = 0.33、WARNING = 0.67、CRITICAL = 1.0（汇聚_评分）；n_independent_flagged 则是超过阈值的支柱数量（整数，0-6）。复合危机方法以粮食安全文献为依据，该文献表明，同时发生的气候、冲突和市场冲击会造成最严重的第4至5阶段结果（de Waal，2018；Maxwell 等，2020）。

独立性调整。干旱胁迫与植被压力呈正相关（Pearson r 约为 0.70；Funk 等人，2015）。相关配对（干旱-植被、IPC 粮食获取）被赋予 1.3 的联合有效独立性权重，而非 2.0。这是一种启发式的保守折扣，而非推导得出的量；在未来修订中，它将被对不断累积的 CERES 信号档案进行的经验Copula拟合所取代（见第 9.3 节）。

3.3 作者指定的逻辑评分模型

主要评分模型将逻辑函数应用于特征向量。系数为作者指定的初始值，由作者根据文献综述以及 87 条国家-季节 IPC 阶段转移记录（IPC全球平台，2025）设定，遵循 Coughlan de Perez 等人（2016）的先验设定方法。这些系数并非从多位领域专家处征询得出，不应被视为经过验证的贝叶斯先验。同一套 87 条记录的数据集同时支撑系数初始化以及用于推导下文单调性界限的阶段转移频率估计；这是同一语料库用于两个目的的分析。这明确是一个初始化步骤；见第 9.1 节。

87条记录的来源。这87条记录涵盖2011—2023年的31个国家，选取自IPC全球平台档案中满足以下条件的全部国家—季节配对：（a）已报告IPC阶段 3+；（b）全部六个CERES 数据源对参考期均具有回溯覆盖；（c）在120天内发布了后续IPC 评估，以提供阶段转移结果。包含IPC报告日期与结果阶段的国家—季节配对完整列表见github.com/northflowlabs/ceres/blob/master/docs/data/ipc_transition_records.csv。在所代表的31个国家中，12个国家在档案中至少发生过一次第4至5阶段事件，由此得到18条后续 IPC 阶段为第4阶段或以上的国家—季节记录；第6节的四个回溯验证案例即从这18条中选出，作为多源数据覆盖最为完整的事件。

指定了三个逻辑斯蒂方程。P(IPC 3+ 90天内)使用完整特征向量，包括CSS、ipc_stress、conflict_stress、drought_stress、food_access_stress、convergence_score和n_independent_flagged。P(IPC 4+)使用ipc_stress、conflict/drought/food_access的最大值、CSS以及convergence_score。P(IPC 5级／饥荒)使用conflict、drought与food_access压力以及ipc_stress的几何平均数。完整系数值见附录 C。

单调性约束在评分后强制执行：P4 <= 0.70 x P3与P5 <= 0.45 x P4。0.70这一上限源自87对IPC转移矩阵：在已报告第3+ 阶段的后续周期中，约有66%–70%观察到第 4+ 阶段。饥荒概率的0.45界限是作者的保守估计；第5阶段宣告较为罕见（自2010年以来有记录者不足15次），且转移矩阵过于稀疏，现阶段尚无法据此作出数据驱动的估计。

系数敏感性分析。所有系数同时扰动 +/-20%（100 次蒙特卡洛均匀抽样）。+/-20% 的范围是对系数大小可能存在分歧的保守估计。在 2026年3月的参考运行中，43 个地区中有 39 个（91%）的等级分类保持稳定（等级未变化）。在最大扰动下等级发生变化的四个地区已在其假设的备注字段中标记。

3.4 敏感性区间

CERES 通过参数化扰动（n=2,000 次抽样），为所有点估计同时报告敏感性区间：对每个特征分别添加均值为零的高斯噪声（sigma=0.15，该数值参考了上游数据源精度指标）。所得 P3 分布的第 5 和第 95 百分位数作为 90%敏感性区间报告。

关键解释说明。这些是输入扰动敏感性区间，而非后验预测区间。它们表示仅在输入噪声下合理的输出范围，而非校准覆盖保证。由于特征被独立扰动，且未捕捉干旱-植被相关性（r 约为 0.70），这些区间低估了真实不确定性。未来版本将在观测到的信号协方差矩阵上实施基于 Copula 的重采样，按照 Kruger 等（2020）的方法迈向恰当的后验预测分布。

4. 警报等级分类

Tier	标签	触发条件
TIER-1	紧急警报	P(IPC4+) >= 0.50 或汇聚_= 层级为危急
二级	警告	P(IPC3+) >= 0.45 或汇聚_层级 = 警告
三级	关注	所有 CSS 为 > 0.05 的其他地区

阈值在 TIER-1 优先考虑召回率而非精确率，这与预先行动原则一致，即错过早期预警的代价高于误报的代价（联合国人道主义事务协调厅，2023）。P(IPC4+) 的 0.50 阈值意味着，从预期来看，所有 TIER-1 警报中应有一半以上在 90天 内验证达到 IPC阶段 4 或以上；随着验证账本逐步充实，这一点将通过实证加以衡量。

每个饥荒假设对象都包含：唯一的加密标识符（CERES-HYP-{ISO3}-{YYYYMMDD}-{6-char hex}）、参考日期和90天预测时段、完整的概率向量 {P3, P4, P5}及其敏感性区间、驱动因素聚类分解、附有各来源标记的证据条目列表，以及在 T+90 时自动评估的证伪测试计划。

5. 验证、评分规则与评级框架

所有 CERES 预测均以前瞻方式进行验证。在 T+90 天时，评级模块会查询 IPC全球平台，获取最新发布的阶段分类。主要评估指标是布里尔分数（Brier，1950）：

```
BS =  (P_hat_3  - O_3)^2
```

其中，若在预测时段内30天内报告了 IPC阶段 3+，则 O3 = 1，否则为0。相对于气候学基线（监测区域集合中 IPC阶段 3+ 出现周数的历史占比，2015—2024年）的布里尔技巧分数（BSS），将在前瞻性数据积累后作为主要报告的技巧度量。

随着验证档案逐渐积累，连续排序概率评分（CRPS；Matheson & Winkler，1976；Gneiting & Raftery，2007）将与布里尔分数一并采用。CRPS 将评分推广到完整的预测分布：

```
CRPS(F, y) =  对 R 上的积分 (F(z) - 1{y <=  z})^2  dz
```

其中，F 是 IPC 阶段结果的预测累积分布函数，y 是实际发生的阶段。CRPS 会奖励将概率质量置于接近实际结果位置的预测，并且在最低限度的分布假设下具有一致性（Kruger等，2020年）。采用 CRPS 需要将 IPC 阶段表示为有序分类结果；计划在初始前瞻性验证期之后进行的首次模型修订中实现这一点。

5.1 预注册校准评估协议

表1将评估方案定义为一种方法论承诺，并须在前瞻性结果数据积累之前作出。不会选择性地报告任何指标；所有分级预测都会永久保留在公开账本中。

指标	定义	最小样本数	目标
布里尔分数	均值（P_hat3 - O3）^2	100 个预测	2026年6月
布里尔技巧分数	1 - BS / BS_气候学	100 个预测	2026年6月
TIER-1 精确率	真实 TIER-1 / 已发布的所有 TIER-1	30 个 TIER-1 警报	2026年9月
TIER-1 召回率	TIER-1 事件数 / Phase 4+ 事件总数	10 个 Phase 4+ 事件	2026年9月
敏感性区间覆盖率	90%区间内结果所占比例	200个预测	2026年9月
连续排序概率评分（有序类别）	完整分布与IPC阶段的对比	500个预测	2027年3月
可靠性图	预测概率与经验频率	500个预测	2027年3月

表1. 预　-已登记的前瞻性校准方案。所有分级预测均以只写入、不事后编辑或删除的方式公开存档于ceres.northflow.no/tracker。

分级结果会写入公开的 JSON 账本（grading_ledger.json），并通过 /v1/grades API 端点对外提供。该账本通过 Git 提交历史记录提供防篡改证据，并以只写一次的方式存储。这种对公开、永久的前瞻性验证作出的承诺，使 CERES 区别于所有现有的饥荒预警系统；这些系统均未维护机器可读、预测级别的准确性记录。

6. 历史回溯验证（仅作合理性检查）

本节目的有限：在前瞻性数据积累之前，确认模型并非存在明显缺陷。这并非性能评估。根据数据完整性，选取了四起 IPC第4–5阶段事件——索马里2011年7月、南苏丹2017年2月、埃塞俄比亚提格雷2022年3月，以及也门2021年6月。这些是该时期最严重且记录最详尽的粮食危机之一；任何合理的模型都应能标记出这些事件。本节未纳入任何对抗性或边界案例。四项结果均为样本内结果：设定系数时已知晓这些事件。这些结果不能证明模型具备前瞻性技巧。

事件	时期	P(IPC3+)	P(IPC4+)	Tier	观测IPC	正确
索马里饥荒	2011年7月	0.961	0.743	TIER-1	第5阶段（饥荒）	Yes
南苏丹饥荒	2017年2月	0.934	0.681	TIER-1	第5阶段（饥荒）	Yes
埃塞俄比亚提格雷	2022年3月	0.912	0.588	TIER-1	第4至第5阶段	Yes
也门长期	2021年6月	0.887	0.534	TIER-1	阶段 4+	Yes

基线比较。表2报告了样本内布里尔分数与三个参考基线的比较结果。持续性基线根据参考日期观测到的 IPC 阶段，将 P(IPC3+) 线性映射为 [0,1] —这是最简单且有实际意义的比较基准。气候学基线对所有地区均设定 P(IPC3+) = 0.65（近似历史基准率，2015—2024年）。

模型	平均布里尔分数	备注
CERES（样本内）	0.0066	根据对事件的了解设定的系数
持久性基线	0.0494	P(IPC3+) = f(参考日期的当前阶段)
气候学基线	0.2275	P(IPC3+) = 0.65，适用于所有地区
无信息（0.5）	0.2500	随机基线

表2. 样本内布里尔分数比较。所有数值均为样本内结果，不应解读为对前瞻性表现的声明。CERES在样本内的表现优于持续性基线，但持续性基线本身也是一个强有力的比较基准，因为在参考日期，四个地区均已处于IPC 第3-4阶段。

在全部四个案例中，90%敏感性区间都包含了观测结果，这与区间构造一致，但预计样本内覆盖率会较为乐观。

7. 实时部署与访问

自2026年3月起，CERES已在 Railway 云基础设施上投入生产运行，覆盖撒哈拉以南非洲、中东、南亚和中美洲的43个国家——约占全球活跃 IPC阶段 3+ 受影响人口的95%（IPC，2025）。所有输出均可免费获取。

公共 REST API（无需身份验证）：GET /v1/predictions、/v1/predictions/{region_id}、/v1/archive/latest、/v1/archive/regions/{id}、/v1/grades、/v1/grades/metrics。完整的 OpenAPI 文档见 https://ceres-core-production.up.railway.app/docs。

机构 API（订阅制）：TIER-1 网络钩子警报，每小时 1,000 次请求，Parquet 导出，每周 PDF 简报。

公共仪表板（ceres.northflow.no）：交互式全球风险地图、可深入查看区域详情的功能（包括假设详情和驱动因素细分）、历史表现记录时间线，以及自 2026 年 5 月起持续更新的验证记录。

密码学验证架构是系统可信度主张的核心：每个 FamineHypothesis 都带有唯一 ID（CERES-HYP-{ISO3}-{YYYYMMDD}-{hex}），预测会写入受版本控制、仅可写入一次的 JSON 账本，而 Git 提交历史则提供篡改证据。预测一经发布，便无法删除或编辑。

8. 伦理考量

预期性行动的反馈闭环。CERES 与任何人道主义物资预置或资金触发系统都没有直接的业务集成。所有输出均为建议性质。为降低过早投入实际应用的风险，我们采取了以下措施：(a) 在本文中明确承认，系数为作者指定的初始值；(b) 在所有点估计旁显示敏感性区间；(c) 公开分级账本，使准确性局限实时可见。

假阳性与假阴性的代价。未通过验证的 TIER-1 警报可能导致资源从真正的危机中转移；CERES 通过报告概率阈值而非二元警报来应对这一问题。漏报 TIER-1 事件则可能延误响应；在 +/-20% 系数扰动下层级发生变化的 43 个地区中，有 4 个已被标记；同时建议用户将临界的二级警报视为需要进行与 TIER-1 同等程度监测的信号。

数据主权。CERES 仅摄取公开可得的机构数据（UCSB、美国国家航空航天局、武装冲突地点与事件数据项目、IPC、世界粮食计划署、联合国粮食及农业组织）。不处理任何个人层面数据。所有上游归属信息均保留在每条假设的证据字段中。

滥用风险。CC BY 4.0 许可允许不受限制的再利用，包括非人道主义行为者的再利用。作者有意选择接受这一风险，以换取最大程度的人道主义可及性与独立审查——这与武装冲突地点与事件数据项目、CHIRPS 和 IPC 数据本身的许可决定相一致。

9. 局限性与未来工作

9.1 作者指定的初始系数。系数由作者根据文献综述和包含 87 条记录的 IPC 转变数据集设定。这些系数尚未从多位领域专家处征询得出，也尚未根据留出数据集进行估计。随着验证档案积累 T+90 结果，系数将通过最大似然估计进行更新。在此之前，所有概率输出都带有高于所报告水平的不确定性。关于判别性能的说明：在 87 条记录的 20% 留出划分（约 17 个测试案例）上，模型的曲线下面积达到 = *0*.84；按 Hanley-McNeil（1982）计算的 95% 置信区间为 *[0*.63、1.*00]*, ，与随机水平表现一致。该数值仅为完整起见而报告，不应被解读为具备技能的证据。

9.2 敏感性区间低估了真实不确定性。扰动程序将特征视为独立加噪，忽略了干旱-植被相关性（r约为 0.70）以及模型参数不确定性。未来版本将实施基于 Copula 的重采样（Kruger等，2020年），以实现恰当的后验预测不确定性量化。

9.3 独立性调整启发式方法。相关信号对的1.3联合独立性权重是一种保守启发式方法。随着CERES 信号档案的积累，它将被经验Copula拟合所取代。

9.4 国家级分辨率。CERES 汇总至 ISO 3166-1三字母代码。一级行政区信号在内部计算，但尚未纳入主要输出。

9.5 数据缺口。若干国家的 IPC 数据频率不定期，或缺少世界粮食计划署调查数据，导致覆盖范围缩小、敏感性区间扩大；相关情况会按运行记录在应用程序接口响应中。

9.6 结构性驱动因素。CERES 仅捕捉急性危机信号。治理失败、长期贫困和长期气候趋势均未纳入模型。

9.7 前瞻性验证期限。前瞻性记录始于2026年3月；首批结果预计自2026年5月起出现。完整的校准分析需要12至24个月（约2,500至5,000个假设—结果对）。一篇纳入前瞻性结果的修订版经验验证论文计划于2026年第三至第四季度投稿至Nature Food。

9.8 单作者设计与系数审查。系数初始值、回溯验证案例选择以及系统架构均由同一作者设定。这是一项公认的认识论局限：设定系数的人与依据样本内案例对其进行评估的人为同一人。计划在首次最大似然估计再估计之前，由Northflow外部的粮食安全专家进行独立专家审查；希望参与此过程的审稿人或从业者可直接联系作者。

10. 与现有系统的比较

系统	输出类型	概率	CI	API	前瞻性验证	频率
CERES	概率预测	Yes	Yes	Yes	是——公开	每周
FEWS NET	情景展望	No	No	No	No	月度/季度
IPC	阶段分类	No	No	Part	No	不定期
世界粮食计划署饥饿地图	粮食获取指数	No	No	Yes	No	每日
INFORM风险指数	风险评分	No	No	Yes	No	每年

CERES的主要独特贡献在于承诺开展公开前瞻性验证，并建立机器可读档案，使校准测量成为可能。这由此形成了反馈闭环，是实现真正持续改进所必需的。

上述比较围绕 CERES 在设计上具有差异化优势的维度展开。要进行平衡的评估，也必须承认现有系统在某些方面更强：FEWS NET 凭借数十年积累的实地分析师专业知识、国内网络和机构整合优势，这些都是 CERES 无法通过算法复制的。IPC 分类凝聚了多方利益相关者共识，并在人道主义资金决策中具有法律效力。世界粮食计划署饥饿地图会摄取实地家庭调查数据，而 CERES 无法按每周频率获取这类数据。
CERES 并非要取代其中任何系统；它是一个自动化概率层，运行于这些系统评估周期之间，并将其输出提交给公众验证。

11. 结论

CERES 表明，以每周更新频率、不到一分钟的流水线运行时间，在全球规模上实现自动化概率饥荒预测，在技术上是可行的。模型当前的系数是作者指定的初始值，其不确定性已得到承认——这是坦诚面对的限制，而非刻意隐瞒的问题。该系统的核心贡献在于架构设计：对每项预测进行公开前瞻性验证，为衡量、报告和改进经验校准创造了条件。

随着验证账本从2026年5月起逐步积累，我们预计将报告经验布里尔技巧评分、CRPS评分、敏感性区间覆盖率以及等级精确率/召回率，为完全基于经验的

验证提交。该系统依据 CC BY 4.0 免费提供；商业订阅则为持续开发提供支持。

参考文献

Brier, G.W.（1950）。以概率形式表示的预测验证。《月度天气评论》，78(1)，1-3。Coughlan de Perez, E., van den Hurk, B., van Aalst, M.K., 等（2016）。用于触发人道主义行动的基于行动的洪水预报。《水文与地球系统科学》，20(9)，3549-3560。De Groeve, T., Poljansek, K., & Vernaccini, L.（2015）。风险管理指数——INFORM：概念与方法学。欧盟委员会，JRC技术报告。de Waal, A.（2018）。《大规模饥饿：饥荒的历史与未来》。Polity出版社。Didan, K.（2021）。MODIS/Terra月度植被指数L3全球1公里SIN网格V061。NASA EOSDIS陆地过程分布式主动档案中心。联合国粮食及农业组织（2024）。粮农组织食品价格指数：方法说明。联合国粮食及农业组织。Funk, C., Peterson, P., Landsfeld, M., 等（2015）。气候灾害红外降水与地面站数据——一项用于监测极端事件的新环境记录。《科学数据》，2，150066。Gneiting, T., & Raftery, A.E.（2007）。严格适当的评分规则、预测与估计。《美国统计学会会刊》，102(477)，359-378。Hanley, J.A., & McNeil, B.J.（1982）。受试者工作特征（ROC）曲线下面积的含义与用途。《放射学》，143(1)，29-36。Hillbruner, C., & Moloney, G.（2012）。当早期预警还不够——从2011年索马里饥荒中吸取的经验教训。《全球粮食安全》，1(1)，20-28。IPC全球平台（2025）。《2025年IPC全球粮食危机报告》。https://www.ipcinfo.org IPC技术手册（2021）。IPC技术手册3.1版。https://www.ipcinfo.org/ipc-country-analysis/ipc-technical-manual/ Kruger, F., Lerch, S., Thorarinsdottir, T., & Gneiting, T.（2020）。基于马尔可夫链蒙特卡罗输出的预测推断。《国际统计评论》，89(2)，274-301。Matheson, J.E., & Winkler, R.L.（1976）。连续概率分布的评分规则。《管理科学》，22(10)，1087-1096。Maxwell, D., & Haan, N.（2021）。预期性行动与早期预警系统。ODI人道主义政策小组。Maxwell, D., Majid, N., Adan, G., Abdirahman, K., & Kim, J.J.（2020）。面对饥荒：索马里人在2011年饥荒中的经历。《食品政策》，91，101857。联合国人道主义事务协调厅（2023）。《预期性行动：入门指南》。联合国人道主义事务协调厅。Raleigh, C., Linke, A., Hegre, H., & Karlsen, J.（2010）。介绍ACLED：武装冲突地点与事件数据集。《和平研究杂志》，47(5)，651-660。Verdin, J., Funk, C., Senay, G., & Choularton, R.（2005）。气候科学与饥荒早期预警。《皇家学会B辑哲学汇刊》，360(1463)，2155-2168。世界粮食计划署脆弱性分析与制图（2025）。实时饥饿地图：方法学与数据源。世界粮食计划署脆弱性分析与绘图。

附录 C：模型系数表

开放代码仓库（github.com/northflowlabs/ceres）中实现的完整逻辑回归模型系数。对数几率 =截距 + 权重 x 特征的总和；概率 = Sigmoid函数（对数几率）。

特征	贝塔（IPC 3+）	伽马（IPC 4+）	德尔塔（饥荒）
截距	-2.10	-3.80	-6.00
综合_压力	5.80	4.50	4.00
IPC_压力	2.40	3.20	4.50
冲突 _压力	1.20	1.80	2.20
干旱_压力	0.90	0.70	1.20
food_获取 _压力	1.10	0.90	1.60
价格_压力	0.60	0.50	0.80
汇聚_评分	2.20	3.10	4.00
n_个独立_标记	0.40	0.60	0.90

验证示例。某区域处于IPC第3阶段（ipc_压力 = 0.50），综合_压力 = 0.45，没有汇聚，所有其他特征均为0：

对数几率(P3) = -2.10 + 5.80*0.45 + 2.40*0.50= -2.10 + 2.61 + 1.20 = 1.71 P3 = Sigmoid函数(1.71) ~ 0.847 对数几率(P4) = -3.80 + 4.50*0.45 + 3.20*0.50= -3.80 + 2.025 + 1.60 = -0.175 P4_原始值 = Sigmoid函数(-0.175) ~ 0.456 单调性：P4 <= 0.70 * 0.847 = 0.593 -> 不具约束力，P4 = 0.456 存在关键趋同（汇聚_评分=1.0，n_个独立_标记=3）：对数几率(P3) += 2.20*1.0 + 0.40*3 = +3.40 -> 对数几率 = 5.11 -> P3 ~ 0.994

汇聚升级完全通过逻辑回归模型中的汇聚_评分和 n_独立_已标记系数来体现。不另行应用单独的乘法升级因子；特征系数是唯一机制。

附录A：API快速参考

基础 URL：https://ceres-core-production.up.railway.app

GET /v1/预测 GET /v1/预测/{region_id} GET /v1/假设 GET /v1/假设/{hypothesis_id} GET /v1/档案/latest GET /v1/档案/regions/{region_id}?limit=52GET /v1/档案/统计数据 GET /v1/等级 GET /v1/等级/指标 GET /健康状态

附录 B：FamineHypothesis 模式（简略）

{

"hypothesis_id": "CERES-HYP-SDN-20260303-A7F3C1", "region_id": "SDN", "region_name": "苏丹", "reference_date": "2026-03-03", "forecast_horizon_days": 90, "alert_tier": "TIER-1", " composite_stress_score": 0.847, "famine_probability": { "p_ipc3plus_90d": 0.966, "p_ipc4plus _90d": 0.724, "p_famine_90d": 0.183, "sensitivity_interval_low": 0.891, "sensitivity_interval_ high": 0.991, "interval_type": "input_perturbation_90pct", "ci_method": "expert_prior_l ogistic_perturbation_v2" }, "driver_clusters": [{ "driver_type": "冲突", "intensity": 0.94, " confidence": 0.91 }, { "driver_type": "IPC_TREND", "intensity": 0.88, "confidence": 0.95 }, { " driver_type": "FOOD_ACCESS", "intensity": 0.76, "confidence": 0.82 }] ,"ipc_phase_forecast": 4, " created at": "2026-03-03T06:12:44Z"_}

注意：ci_method 字段在早期版本的应用程序接口中曾命名为 method。当前字段名为 ci_method；为保持连续性，保留值 expert_prior_logistic_perturbation_v2，并反映其旧有特征描述——关于这些系数目前作为作者指定的初始值的说明，请参见第 3.3 节。

（c）026 Tom Danny S. Pedersen / Northflow Technologies AS。依据 CC BY 4.0 许可（creativecommons.org/licenses/by/4.0).
2