

第六章 数据与方法

6.1 地理分析单元

本通报除中国之外的所有国家和国家内行政单元，均采用全球行政单位层(GAUL)的划分标准，中国本土的行政边界利用国家测绘局地图界线。GAUL 数据可由 FAO 地理数据网下载。

6.1.1 全球农业生态区划

关于全球农业生态区划 (CPSZ) 的具体描述，详见附件 A。CPSZ 的划分主要参考 FAO 森林资源评估的全球生态系统分布图 (39)，在进一步细分或者修改时，还参考了 Köppen 气候带划分产品 (数据来自 Grieser et al. (40)，VASCLimO 1976-2000) 以及全球农业生态区域划分项目 (41) 提供的“最优谷物种植区分布”产品。其他参考数据源还包括：美国农业部 USDA (42)，Ramkuttu 等研究成果中的全球耕地分布图 (43) 以及 Monfreda 等研究成果 (44) 等等。

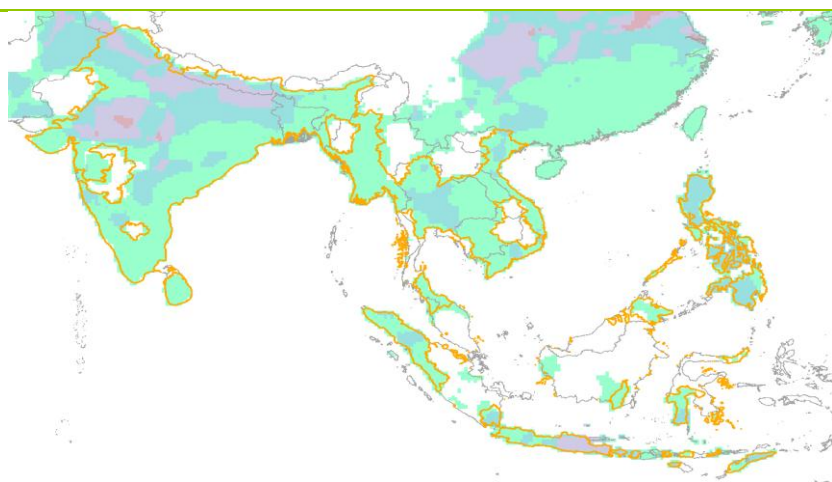
需要特别说明的是，对于中国的农业生态区划，没有采用前述标准进行划分。而是直接使用了中国已有的农业生态区划 (45) 及其英文版 (46)。

6.1.2 主产区分析

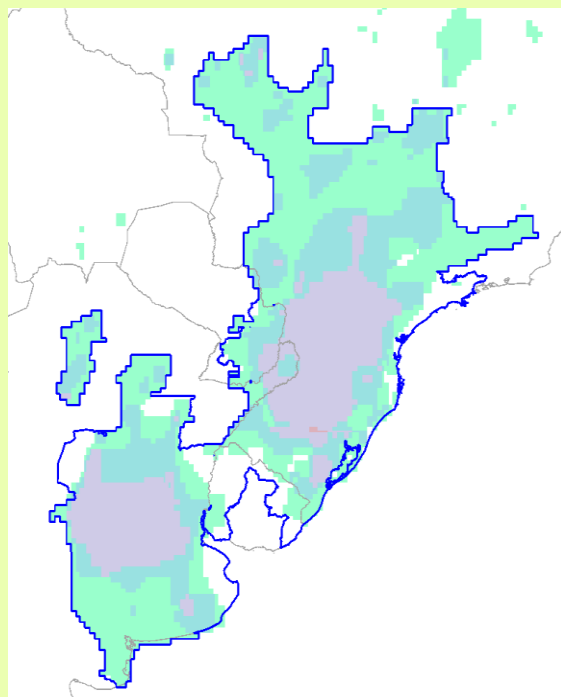
主产区的划分主要基于欧盟联合研究中心 (JRC) 的作物分布图，然后将玉米、水稻、大豆和小麦的分布区域进行合并。每个主产区包含四大主要作物 (玉米、水稻、大豆和小麦) 中的至少一种，主产区边界选择相邻区域没有主要作物种植的国界或者国家级层次以下的行政界线。例如：“中欧-俄罗斯西部”主产区，其北部边界采用 59 号农业生态区 (CPSZ-59) 中的“乌克兰-哈萨克斯坦”边界，东部边界则以乌拉尔山脉为准。

未来发布的 CropWatch 通报将涵盖本次通报 4 个主产区以外的更多作物主产区。

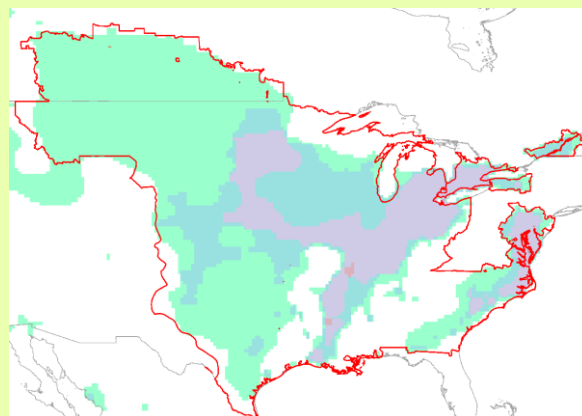
图 6.1 主产区主要作物分布图



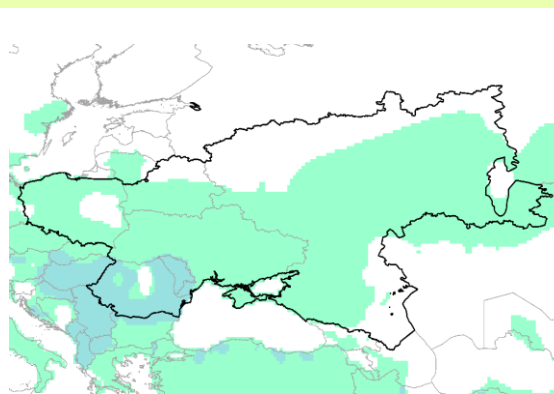
东南亚



南美洲



美国和加拿大



波兰和罗马尼亚到俄罗斯乌拉尔山脉

注：图中只显示玉米、水稻、大豆和小麦的种植分布。白色背景，没有四种主要作物中任何一种分布的区域；浅绿，只有一种主要作物分布的区域；深蓝-深绿，有两种主要作物分布的区域；紫色，三种主要作物分布的区域；深紫-红色，四种主要作物同时分布的区域。四种作物同时分布的区域很少，而且几乎都位于印度。

6.1.3 国家选择

国家的选择主要基于 FAO 产量和粮食交易统计数据，所选国家的主要作物产量及交易量累积占全球该种作物产量和交易量的 80%。然而，本通报也选择了一些在统计中比重较小的国家，这主要是处于这些国家地理位置（如亚洲的乌兹别克斯坦，非洲的埃及）的考虑。关于所选国家的基础信息和农作物概况，详见附件 F。

6.1.4 国家尺度分析

环境指标的计算范围包含：60 个全球农业生态区，4 个作物主产区以及通报所选择的 173 个国家和地区。这些区域基本由全球行政单位层(GAUL)第一级国家单元 (GAUL_0) 构成，同时对小于 25*25 公里的单元进行了剔除。

6.1.5 省-州尺度分析

对于一些面积较大的国家（澳大利亚、巴西、加拿大、中国、印度、哈萨克斯坦和美国），本通报对国家尺度以下的省/州级行政单元进行了分析。

6.1.6 中国省份分析

本通报的监测内容覆盖中国 24 个省级行政单元，这些省级行政单元按照国家标准划分为六大区域，包括华北地区、东北地区、华东地区、中南地区、西南地区以及西北地区（47）。

6.1.7 农作物种植分布及耕地分布图

全球农作物分布数据来源于 JRC。最初的农作物分布数据是为全球水分胁迫目的进行划分的，共包含了 11 种作物类型。本通报只对其中的玉米、水稻、大豆和小麦分布区进行提取，生成后的作物分布数据分辨率为 0.25° 。其他的农作物分布数据还包括：USDA 的主要作物分布图（42），FAO/IRSA（48）（41）的雨养和灌溉农业分布图等等。

本通报所用的耕地分布数据，是将 2010 年和 2011 年 MODIS 土地利用数据产品(49)，第二版国际地圈生物圈计划 (IGBP) 全球土地覆盖数据集(IGBP-DISCover) (50)，全球地表覆盖 2009 (51)(52)这些产品当中的耕地进行合并得到。中国 2000 年，2005 年和 2010 年的耕地分布数据直接采用中国科学院遥感与数字地球研究所的 2000 年，2005 年和 2010 年中国土地覆盖数据（53）。

6.2 数据

6.2.1 NDVI 数据

本通报所用的 NDVI 数据主要是 NASA 提供的 MODIS NDVI 数据，通报所用 NDVI 数据覆盖的时间从 2002 年 1 月至 2013 年 9 月底。通报只选择 MODIS Terra 星陆地 3/4 级产品（54），空间分辨率为 1 公里。在原始数据下载之后，用全球耕地分布数据对 NDVI 进行掩膜处理，剔除非耕地数据，确保 NDVI 数据集适于农作物长势监测及估产等研究。

此外，本通报还使用了 VITO（55）提供的，基于 SPOT-VEGETATION 传感器，分辨率为 0.185° 的长时间序列（1999 年-2012 年）NDVI 平均数据。

6.2.2 气温

本通报生产的气温产品为覆盖全球（ 0.25×0.25 度）的月产品，产品时间范围为 2000 年 1 月至 2013 年 9 月。该产品是基于全球变化总目录资料库中的全球地表日数据集（GSOD）生产的（56）。GSOD 数据集是由美国国家气候中心（NCDC）生产，由地表小时数据集（ISH）DSI-3505 (C00532)插值得到。可下载的在线数据最早可追溯至 1929 年，最新数据在日观测数据获取 1-2 日内更新。该数据集可提供来自全球 9000 多站点的气温、露点温度、海面气压、风速、降雨、雪深等观测参量。

本通报使用了 GSOD 数据集中 2000 年 1 月至 2013 年 9 月的日平均气温计算出月平均气温，考虑高程对温度的影响，结合 STRM_DEM 数据（57）使用克里金插值法得到了全球 0.25×0.25 度的月气温产品。

6.2.3 光合有效辐射 (PAR)

光合有效辐射 (PAR) 是影响作物产量的一个重要参数, PAR 指的是波长在 400-700 纳米之间的太阳辐射, 在不同的光周期内 PAR 和太阳短波辐射有一定的转换关系 (58)。所以光合有效辐射可以由太阳短波辐射计算得到。2001-2012 年全球月平均 PAR 数据来自 NASA(美国航空航天局)的 CERES 下行短波辐射产品 (59)。2013 年全球月平均 PAR 数据由 Geoland-2 项目的全球一小时的下行短波辐射产品累加得到, 在 Geoland-2 产品不能覆盖的区域, 引入了风云卫星 (FY-2D) 的地表辐射产品进行补充 (60-62)。所有上述产品都被转换到了 WGS-84 投影坐标系下, 空间分辨率统一为 0.25 度。

6.2.4 降水产品

CropWatch 通过集成不同来源的降水产品, 获取了 2000 年 10 月至 2013 年 9 月, 覆盖整个陆地表面的, 空间分辨率为 $0.25^{\circ} \times 0.25^{\circ}$, 时间分辨率为月的降水时间序列。该降水数据集使用的两大数据源包括: (i) TRMM 降水数据集: 由 NASA 提供的第 7 版空间分辨率为 $0.25^{\circ} \times 0.25^{\circ}$, 覆盖南北纬 50° 之间地区的 TRMM 降水产品(63); (ii) GPCC 降水数据集(64): 通过重采样技术将其由 0.5° 与 1° 空间分辨率采样成 0.25° 。2000 年 10 月-2010 年 12 月之间的 CropWatch 降水时间序列中南北纬 50° 范围的降水量采用 TRMM 月降水产品数据 (TRMM 3B43), 其他区域的采用的是 GPCC 数据。2011 年 1 月-2013 年 6 月之间, 南北纬 50° 之外的地区采用的是 GPCC 监测分析数据(65)。2013 年 7 月-2013 年 9 月, 南北纬 50° 之间的地区, 通过合成时间分辨率为 3 小时, 空间分辨率为 $0.25^{\circ} \times 0.25^{\circ}$ 的 TRMM 3B42 近实时降水产品数据而来, 其他区域采用的是 GPCC First_guess 产品(66)。将 CropWatch 降水产品与实测站点的月降水量进行比较, 结果表明二者具有较高的一致性。由于采用不同的数据源合成降水时间序列数据, 在一些月份二者的拼接处有明显的(切割线)现象。

6.2.5 植被健康指数 (VHI)

植被健康指数 (VHI) 可以有效地指示作物生长状况。本通报采用温度状态指数 (TCI) 和植被状态指数 (VCI) 加权的方法计算植被健康指数(67) (68) (69)。温度状态指数和植被状态指数数据均可以通过 NOAA 数据中心的卫星数据应用和研究数据库下载 (70)。该方法的计算方程如下:

$$VHI = a * TCI + (1 - a) * VCI$$

但是, 本文使用的经验系数 (a) 在全球不同区域是变化的, 而不是前面研究者使用的常数 0.5。

6.2.6 生物量

通报根据 Lieth 的“迈阿密”模型(71) (72)计算了净初级生产力 (NPP), 以此作为全球潜在生物量的度量。在迈阿密模型中, NPP 可表达为温度和降雨两个环境因子的函数, 如下式所示:

$$NPP = \min(NPP_T, NPP_P)$$

$$NPP_T = \frac{3000}{1 + e^{1.315 - 0.119T_c}}$$

$$NPP_P = 3000(1 - e^{-0.000664Prec})$$

其中, T_c 为年平均气温, 单位为摄氏度 ($^{\circ}C$), Prec 表示年平均降雨, 单位为毫米 (mm)。

6.3 方法

6.3.1 环境指标 (EI)

本通报基于作物生育期的环境指标，比较了不同年份、不同地区国家的环境因子对农作物生长的影响。通过对农业区赋予较高权重，关注区域每年一个数值，从而可以在相同的空间和时间尺度上与其他变量或指标（如农业统计数据和社会经济指标）进行比较。

对于降雨，前人研究了降雨环境指标 (73)，并实际应用于综合分析 (74)，(75)。本次通报将这种指标计算方法推广到其他环境因子（气温和光合有效辐射 (PAR)），在不同区域（如 CPSZ, 国家或其他空间单元）分析了三种环境指标（降水、气温和 PAR）对农业生产的影响。

环境指标的计算基于 25km 空间分辨率的环境因子数据，利用潜在净初级生产力(NPPP) (72) 作为权重（象元的潜在生产力越高，权重值越大），结合耕地掩膜计算了各区域在用户定义时段内的累积值。其中，PAR 和降水两个因子采用了简单累加，而对温度选取大于 5°C 的值进行累加。对于时段的选取，大部分分析采用全年的累加(从 2012 年 10 月到 2013 年 9 月)，当使用特定时段时，文中都会标注(如北半球冬季作物使用了 1 月到 5 月生长期的累加)。

6.3.2 总产，面积和单产

总产

在 30 个主产国的总产估算中，通报充分考虑了各个国家的作物物候特征，采用各国当前生长季内的平均 NDVI 结合作物掩膜以及前两年的作物产量统计数据，估算了今年四种粮油作物（玉米、小麦、大豆和水稻）的总产量，计算公式如下所示：

$$\text{总产}_i = a + b * \text{NDVI}_i + c * \text{总产}_{i-1} + d * \text{总产}_{i-2}$$

其中，i 代表关注年份；a, b, c, d 为利用 2002-2012 年平均 NDVI 和 2000-2011 年 FAOSTAT 产量统计数据由逐步线性回归得到的四个系数。

对于中国，各种作物的总产通过单产与面积的乘积进行估算，公式如下所示：

$$\text{总产} = \text{单产} * \text{面积}$$

面积

通报基于 CropWatch 系统利用作物种植比例（播种面积/耕地面积）和作物种植结构（某种作物播种面积/总播种面积）对中国四种粮油作物（玉米、小麦、大豆和水稻）的播种面积进行了估算。其中，种植比例基于高分辨率的环境星 (HJ-1 CCD) 数据由非监督分类获取，中国的作物种植结构通过 GVG 系统由田间采样获取（可参考 6.3.6 节）。通过农田面积乘以作物种植比例和作物种植结构得到不同作物的播种面积。

对 30 个主产国的作物面积估算，通报采用了两种方法。第一种方法是基于时间序列 MODIS NDVI 数据提取不同作物的特征曲线进行面积的估算(78) (79) (80) (81)，另一中方法是基于前两年面积统计数据的回归，公式如下：

$$\text{面积}_i = a + b * \text{面积}_{i-1} + c * \text{面积}_{i-2}$$

其中 a, b, c 是基于 FAOSTAT 统计数据或各国面积统计数据通过逐步线性回归得到的三个系数。

单产

30 个主产国的单产由总产和面积计算，如下式：

$$\text{单产} = \text{总产} / \text{面积}$$

中国单产的估算采用了农业气象模型以及遥感指数两个模型(82) (83)。通过将两种模型的估算结果进行综合平均，对中国 2013 年四种作物的单产进行了估算。

6.3.3 复种指数 (CI)

复种是指在同一田地上一年内接连种植两季或两季以上作物的种植方式，复种指数是用来描述耕地在生长季中利用程度的指标，通常以全年总收获面积与耕地面积比值计算 (84)，也可以用来描述某一区域的粮食生产能力 (85)。本通报采用范锦龙等人 (86) 提出的方法，基于经过 S-G 方法 (87) 平滑后的 MODIS 时间序列 NDVI 曲线，提取曲线峰值个数、峰值宽度和峰值等指标，计算耕地复种指数。

复种指数的提取最终相当于时间序列 NDVI 曲线峰值频数的提取。本通报利用差分方法提取经 S-G 平滑后的 NDVI 散点序列的极大值个数。

假设一个像素由包含 N 个元素的离散点系列 S 构成，由 S 的前后两个元素值的差构成包含 N-1 个元素的点系列 S1，即

$$S1 = \text{DIFF} (S)$$

其中，DIFF 为取前后两个元素值差的函数。接着判断 S1 各个元素值的正负，如果 S1 的某个元素值小于 0，就把这个元素的值赋为 -1，如果大于等于 0，就把这个元素的值赋为 1，存为点系列 S2，即

$$S2 = \text{SIGN} (S1)$$

其中，SIGN 为判断元素值为正负的函数。然后，再求 S2 的前后两个元素值的差，形成点系列 S3，即

$$S3 = \text{DIFF} (S2),$$

最后，统计点系列 S3 中元素值为 -2 的元素个数，元素值为 -2 的总个数即为求得的极大值个数，即 NDVI 曲线中峰值的个数。基于上述公式，再利用 NDVI 阈值等条件加以限制，最终计算得到取值为 1, 2 和 3，代表一季，两季和三季种植的像元级复种指数分布图。

6.3.4 未种植耕地比率 (UAL)

本通报中，引入未种植耕地比率目的是为了表征某一行政区划内在过去一年内未被耕种的耕地面积占耕地总面积的比例。众多研究成果表明，NDVI 与鲜活植被具有较高的相关性(88) (89)，因此本通报中使用时间序列 MODIS NDVI 数据用于耕种与未耕种耕地的识别。对于每一像元，利用 S-G 平滑算法(87)对每年的作物生长曲线进行平滑，并逐像元的提取 2002 年至 2013 年每年的 NDVI 峰值，同时计算 2002 年-2012 年间各像元年度 NDVI 峰值的均值 (NDVIm) 以及多年 NDVI 峰值的标准差 (NDVlstd)。基于这些参数，利用 NDVI 阈值法(90)(91)和决策树算法对 2012 年 10 月-2013 年 9 月期间的耕种与未耕种耕地进行区分。NDVIm 与 NDVlstd 的差值被引入到决策树分类过程中，可以在一定程度上削弱年度间作物长势差异造成的分类误差。分类树的规则如下：

$$\text{耕地状态} = \begin{cases} \text{未耕种, 如果 Peak}_{2013} \leq 0.15 \\ \text{未耕种, 如果 } 0.15 < \text{Peak}_{2013} < 0.45 \text{ 且 } \text{Peak}_{2013} < (\text{NDVI}_m - \text{NDVI}_{\text{std}}) \\ \text{耕种, 如果 } 0.15 < \text{Peak}_{2013} < 0.45 \text{ 且 } \text{Peak}_{2013} \geq (\text{NDVI}_m - \text{NDVI}_{\text{std}}) \\ \text{耕种, 如果 Peak}_{2013} \geq 0.45 \end{cases}$$

6.3.5 潜在生物量比例 (PBR)

潜在生物量比例 (PBR) 是本通报提出的一个用于衡量当前作物生物量潜力的新指数。由于 NDVI 与植被生产力与生物量具有很好的相关性(92)(93)(94)(95)，本通报使用 NDVI 作为作物生物量的近似，利用长时间序列的 NDVI 指数计算了 PBR，公式如下所示：

$$\text{PBR} = \frac{\text{NDVI}_{\text{max}_c} - \text{NDVI}_{\text{min}_h}}{\text{NDVI}_{\text{max}_h} - \text{NDVI}_{\text{min}_h}}$$

其中， $\text{NDVI}_{\text{max}_c}$ 代表当前年份最大 NDVI 值， $\text{NDVI}_{\text{max}_h}$ 代表历史最大 NDVI 值， $\text{NDVI}_{\text{min}_h}$ 代表历史最小 NDVI 值。考虑到 NDVI 峰值可以反映纯植被信息，较少受土壤等背景因素的影响， $\text{NDVI}_{\text{max}_c}$ 和 $\text{NDVI}_{\text{max}_h}$ 可通过遍历 NDVI 时间序列获取。但是 $\text{NDVI}_{\text{min}_h}$ 会受土壤、云雨等因素的影响，不能简单使用遍历方法，因此本通报引入了最小 NDVI 经验值 (0.15) 进行 $\text{NDVI}_{\text{min}_h}$ 的计算，具体如下所示：

$$\text{NDVI}_{\text{min}_h} = \max(0.15, \text{NDVI}_{\text{min}_h0})$$

其中， $\text{NDVI}_{\text{min}_h0}$ 是由遍历 NDVI 时间序列得到的时间序列 NDVI 最小值。

作为当前生物量与历史最大生物量的比值，PBR 的取值范围在 0-1 间。PBR 值越高，代表该年度的潜在生物量越大。

6.3.6 作物种植结构

作物种植结构是指在某一行政单元或区域内，每种作物的播种面积占总播种面积的比例。作物种植结构仅用在国内的作物种植面积估算中(76)。作物种植结构数据通过利用 GVG 采样系统在特定区域内开展地面观测，通过系统采集的大量的作物照片来估算每一区域各种作物的种植比例(96)(97)(98)(85)。

6.3.7 时间序列聚类

时间序列聚类方法就是自动或半自动地比较时间序列栅格影像中各像元的时间序列曲线，并且把具有相似特征曲线的像元归为同一类别，最终输出不同分类结果的过程。这种方法的优势在于能够综合分析时间序列数据，捕捉其典型空间分布特征。在全球农情遥感速报中，我们应用了 VITO 为 JRC/MARS 开发的 SPIRITS 软件，对 NDVI 和 VHI 时间序列影像（当前作物生长季与近 5 年平均的差值）进行了时序聚类分析。