www.sciengine.com/CSD
ISSN 2096-2223
CN 11-6035/N

OPEN ACCESS
CC BY

文献 CSTR：
32001.14.11-6035.nasdc.2023.0022.zh

文献 DOI：
10.11922/11-6035.nasdc.2023.0022.zh
数据 DOI：
10.57760/sciencedb.agriculture.00278

文献分类：信息科学

收稿日期：2023-11-28
录用日期：2024-04-19
发表日期：2026-02-04

2012–2018 年基于 HP 滤波法的西红柿批发价格波动统计数据集

孔汇鑫[1,2]，韩昀[3]，胡林[1,2]，曹姗姗[1,2,5*]，孔繁涛[4,5*]

1. 中国农业科学院农业信息研究所，北京 100081
2. 国家农业科学数据中心，北京 100081
3. 中国农业科学院农业资源与农业区划研究所，北京 100081
4. 中国农业科学院农业经济与发展研究所，北京 100081
5. 三亚中国农业科学院国家南繁研究院，海南三亚 572024

摘要：我国西红柿产量位居全球第一，西红柿的历史价格及其价格波动数据对影响因素分析和政府宏观调控具有重要意义。本数据集数据主要来源于全国农产品商务信息公共服务平台，由其他权威平台和项目组实际市场调研数据作为补充。按照权威性、连续性、可用性等原则，对收集到的基础价格数据进行数据补充、饱满度稳定性检验、剔除不规范数据、填补缺失值等操作后，形成中国西红柿价格波动数据集的基础数据。通过计算趋势数据和价格波动率数据，对数据集进行进一步挖掘。本数据集时间上包括周度、月度 2 个尺度，空间上将 25 个省（自治区、直辖市）划分为六大地理区划，方便观测各省周度、月度的价格波动特征。本数据集可为我国西红柿价格波动规律和监测预警研究提供基础。

关键词：西红柿；价格波动；科学数据；蔬菜交易

数据库（集）基本信息简介

数据库（集）名称	2012–2018年基于HP滤波法的西红柿批发价格波动统计数据集
数据作者	孙伟
数据通信作者	孙伟（sunwei02@caas.cn）
数据时间范围	2012–2018年
地理区域	东北地区（辽宁省、黑龙江省），华北地区（北京市、天津市、河北省、山西省、内蒙古自治区），西北地区（陕西省、甘肃省、宁夏回族自治区、新疆维吾尔自治区），华东地区（上海市、江苏省、浙江省、安徽省、福建省、江西省、山东省），西南地区（四川省、云南省），中南地区（河南省、湖南省、湖北省、广东省、广西壮族自治区）
数据量	4.86 MB
数据格式	*.xlsx, *.jpg
数据服务系统网址	https://doi.org/10.57760/sciencedb.agriculture.00278
基金项目	中国农业科学院创新工程任务（CAAS-ASTIP-2016-AII）

* 论文通信作者
曹姗姗: caoshanshan@caas.cn
孔繁涛: kongfantao@caas.cn

数据库（集）组成	本数据集包含：（1）6个EXCEL文件，分别为2012–2018年全国6大地理区划25个省份的4200个月度价格时间序列和18200个周度价格时间序列数据，包括日期、周均价、月均价、周均波动率、月均波动率。（2）25张月度价格趋势循环序列分解图，75张价格波动变化图，包括周均价历史轨迹、月均价历史轨迹、月度价格波动率。

引　言

西红柿，又名番茄、洋柿子、番柿，原产于南美洲，于 19 世纪引入中国，因其生长对土壤要求较低，且加工食用方式多样，可以生食、煮食、加工制成番茄酱等，是我国主要的蔬菜作物之一。我国是全球最大的西红柿生产国，据 FAO 统计，2021 年我国西红柿种植面积约为 111.3 万公顷，产量约为 6609 万吨，占全球产量的 1/3 以上。

我国西红柿主要优势产区有黄淮海及环渤海设施生产区域、北部高纬度夏秋生产区域、长江流域早春生产区域、西南冬季生产区域和华南番茄生产区域[1]。诸如季节更替、气象条件、政策导向以及社会舆论等多元要素，极易引发西红柿市场定价出现频繁且剧烈的震荡，这种不稳定性严重制约了该产业的持续性稳健运行[2]。对价格波动变化率和波动情况进行研究可以有效探究价格的波动规律和各个因素对于价格的影响，从而针对性地提出更好的种植建议，减少不必要的损失。历史价格数据对于价格研究来说是不可或缺的，使用已处理好的价格波动率数据可以减少后续研究中的数据处理时间，提高效率。当前科学数据论文所涉及的专业范围较为广泛，且内容呈现多元化特征。其中，尽管蔬菜相关主题的研究文献数量众多，但聚焦于西红柿的专项探讨仍相对有限。本数据集以西红柿的历史定价记录为依据，对其价格波动规律进行分析，从而直观展示西红柿价格的变动趋势与周期性特征，为后续深入探讨西红柿价格影响因素及开展预测研究提供数据基础。本研究构建的数据集，以西红柿过往时期的价格数据作为核心基础，在此前提下对西红柿价格的波动变化态势展开计算分析。

本数据集以西红柿历史价格数据为基础，计算西红柿价格的波动趋势和波动率，并使用 HP 滤波法去除了季节的影响，对预测西红柿价格变化趋势、监测预警西红柿价格来说具有重要意义。为了支撑价格预测研究，本文基础数据源自国内 29 个省级行政区域内的主要农产品批发市场。样本具体涵盖了从 2012 年起至 2018 年止的西红柿每日平均成交价，并以此构建了研究用的核心数据库。为保证数据的完整性，经过饱满度检验，删除了 4 个省份的数据，保留 25 个省份的数据，形成本数据集。本数据集包含数据饱满度较高的 25 个省（自治区、直辖市）的西红柿周度、月度平均价格数据及价格波动变化率数据，统计图共 50 张，包括历史价格变化趋势图和价格波动变化率折线图，HP 滤波分解趋势图 25 张。趋势分解图可直观呈现西红柿价格的波动状态及波动周期，为后续开展西红柿价格影响因素的分析工作与价格预测研究提供坚实的数据支撑。

1　数据采集和处理方法

1.1　数据来源与数据概况

本数据集的基础数据为 2012 年至 2018 年这 7 年间，国内各省份不同类型农产品批发市场的西红柿每日平均成交价。数据来源以商务部提供的全国农产品商务信息公共服务平台数据为主，该平台提供的数

据占比超过 90%，对于缺失的调研数据，采用算术平均法进行计算补全。在数据处理环节，首先运用人工查验的方式对数据饱满度进行检测，若某组数据的缺失时长超过 1 周，则判定其数据饱满度未达到要求，对该组数据执行删除处理。为保障最终数据的合理性与准确性，将所有采集到的数据转化为散点图形式，通过可视化观察排查其中存在的异常不合理数据，并将此类数据予以剔除。完成上述全部数据整理操作后，最终获取到各农产品批发市场西红柿的日均交易价格数据集。

1.2 数据处理

本研究借助 Excel 软件中的 AVERAGEIFS 函数，通过系统计算得出各省份西红柿的周平均价格、月平均价格，以及周度与月度平均价格波动率，累计生成相关信息 22400 条。同时，运用 Eviews12 软件对月度价格数据实施 HP 滤波法处理，最终得到 2012－2018 年各省份西红柿月度价格趋势循环序列分解图，共计 25 张。本数据集依据地理区划标准，将涉及的 25 个省份划分为 6 大地理区域；在时间维度上，将数据划分为周度和月度两个部分；在表现形式上，则分为文本和图片两种类型。数据处理的具体流程详见图 1。

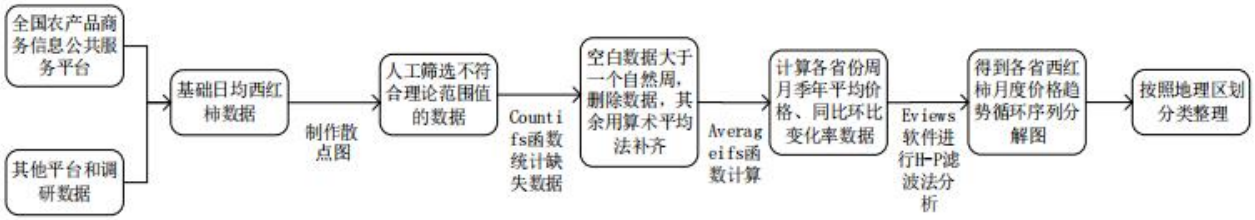


图 1 总体流程图

Figure 1 Overall flowchart

数据处理的第一步的核心是保障数据连续性，在此基础上统一数据的名称、长度、格式及单位，核心在于甄别无效信息去除以及处理缺失值。在样本筛选阶段，将连续 7 天设定为保留数据的阈值。鉴于长时段的断档不仅会导致插值失真，还会严重干扰后续图表的可视化呈现，故凡是中断跨度超出一周的样本，均予以整体舍弃。反之，针对那些零散分布且未破坏整体趋势的缺失值，则采取临近值填补策略，即利用该空缺日相邻的前后两期数据求取均值来完善序列。第二步为将数据导入 Eviews 软件，使用 HP 滤波法对数据进行价格趋势循环序列分解，并绘制成图像，以此得到 25 个省（自治区、直辖市）西红柿价格的趋势循环序列分解图。第三步按照设定的时间周期，通过公式计算周度和月度的价格波动率，格式设置为百分数形式，保留 2 位小数。

1.3 数据转换与模型处理

1.3.1 价格波动率计算

价格波动率可以反映价格的波动变化程度，体现价格的波动特征和周期，从而便于后期的价格预测，提前做出对策。价格波动率是在周/月平均价格的基础数据之上，通过公式计算得出：

$$\text{价格波动率} = \ln\left(S_t / S_{t-1}\right) \tag{1}$$

其中，ln 为自然对数，S_t为单位时间价格，S_{t-1}为上一单位时间价格。

1.3.2 HP 滤波法模型处理

HP 滤波是一种常用的时间序列分析工具，把宏观经济运行看作是潜在增长和短期波动的某种组合，

运用计量方式将价格序列分解为趋势曲线与周期曲线[3]。该方法在保留原序列完整性的基础上，抽取出一条平滑的曲线，弱化了趋势性、波动性等因素之间的相互影响，方便进一步对数据进行价格冲击或价格预警等研究分析。

根据 HP 滤波法，将价格数据进行分解，得到趋势数据g_t和周期波动数据c_t；g_t是从时间序列中得到的平滑的序列，表示的是价格的总体趋势：

$$y_t = g_t + c_t \quad t = 1, 2, \dots, n \tag{2}$$

其中，y_t为第 t 期西红柿实际价格，n 为价格条数，g_t为第 t 期西红柿价格趋势数据，在图上体现为一条平滑的曲线，c_t为周期波动数据，在图中表现为波动幅度较大的周期性曲线。HP 滤波分解原理是把原价格序列中的趋势元素分解出来，原价格序列中可观测部分趋势元素常被定义为最小化问题的解，即 HP 滤波的问题就是使损失函数最小[4]。即式（2）最小化问题：

$$min\left\{\sum_{t=1}^{n}(y_t - g_t)^2 + \lambda \sum_{t=1}^{n}[(g_{t+1} - g_t) - (g_t - g_{t-1})]^2\right\} \tag{3}$$

对式（2）进行一阶求导，并令导数为 0，用矩阵形式表示为：$c = \lambda F g$，整理后得：

$$g = (\lambda F + I)^{-1} y \tag{4}$$

其中，λ为平滑参数，用于平衡趋势成分与周期成分的权重，I 为单位矩阵，维度与价格时间序列长度一致，F 为辅助矩阵，由 HP 滤波法的二阶差分结构推导得出，用于分离趋势成分与周期成分，y是原始的西红柿实际价格，g是西红柿价格趋势数据。HP 滤波分解方法依赖于λ的值，所以在进行滤波分解操作前，要先确定λ的值[5]。当λ=0 时，如趋势线满足最小化问题，则该趋势线为实际产出线，随着λ值不断增加，趋势线更加平滑，当λ趋近于无穷大时，趋势线将接近直线。

根据一般研究经验，λ的选择如图 2 所示。

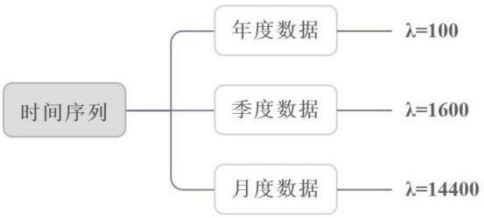


图 2　λ数值选择示意图

Figure 2　Schematic diagram of λ value selection

最后将数据进行可视化，数据集图中 trend 代表趋势线，cycle 代表周期波动线，通过观察 trend 线的趋势可以看出价格的总体波动趋势。

2　数据样本描述

2.1　文本数据

在 2012–2018 年基于 HP 滤波法的西红柿价格波动统计数据集中，价格单位为元/公斤，波动变化率以百分比表示，所有数值均保留两位小数。数据集涵盖的 25 个省份划分为华北、东北、华东、中南、西南及西北六大地理区域，时间上按周与月进行划分，共包含文本数据 22400 条。表 1 汇总了本数据集文本数据按时间尺度分类的各项数据指标。

表 1　文本数据涉及数据项

Table 1　Data items included in the textual dataset

数据项	地理区划	数据量分布
周/月度价格数据	东北地区	辽宁省、黑龙江省，周度价格数据 1456 条，月度价格数据 336 条。
	华北地区	北京市、天津市、河北省、山西省、内蒙古自治区，周度平均价格数据 3640 条，月度价格变化数据 840 条。
	西北地区	陕西省、甘肃省、宁夏回族自治区、新疆维吾尔自治区，周/月度平均价格数据 2912 条，周/月度价格变化数据 672 条。
	华东地区	上海市、江苏省、浙江省、安徽省、福建省、江西省、山东省，周/月度平均价格数据 5096 条，周/月度价格变化数据 1176 条。
	西南地区	四川省、云南省，周/月度平均价格数据 1456 条，周/月度价格变化数据 336 条。
	中南地区	河南省、湖南省、湖北省、广东省、广西壮族自治区，周/月度平均价格数据 3640 条，周/月度价格变化数据 840 条。

图 3 以月平均价格（单位：元/千克）及其月度变化率为例，展示了数据集中南地区的部分数据。

日期	河南省（	湖北省（	湖南省（	广东省（	广西省（	河南月度变	湖北月度变	湖南月度变	广东月度变	广西月度变化率
2012年1月	3.31	3.63	3.17	2.59	2.27					
2012年2月	3.64	3.90	3.34	2.86	2.33	9.62%	6.99%	5.39%	9.91%	2.64%
2012年3月	4.92	4.48	4.04	3.41	2.91	30.25%	13.96%	18.99%	17.61%	22.21%
2012年4月	5.20	4.91	4.63	3.79	2.86	5.48%	9.24%	13.49%	10.56%	-1.59%
2012年5月	2.70	4.64	5.08	4.85	4.23	-65.73%	-5.79%	9.29%	24.80%	39.01%
2012年6月	1.09	2.50	3.31	3.23	3.44	-90.42%	-61.61%	-42.71%	-40.84%	-20.54%
2012年7月	1.23	2.22	2.44	3.03	2.46	11.66%	-12.29%	-30.65%	-6.43%	-33.50%
2012年8月	2.24	3.14	2.95	3.64	2.70	60.11%	34.75%	18.91%	18.60%	8.98%
2012年9月	2.95	3.89	3.50	3.77	2.85	27.75%	21.56%	17.29%	3.35%	5.65%
2012年10月	1.92	3.21	2.71	2.77	2.43	-42.91%	-19.11%	-25.71%	-30.80%	-16.11%
2012年11月	2.19	3.13	2.42	2.40	2.38	13.06%	-2.59%	-11.39%	-14.31%	-2.04%
2012年12月	2.88	3.44	3.08	2.79	2.36	27.23%	9.35%	24.40%	15.01%	-0.96%
2013年1月	3.66	4.02	3.57	2.73	2.67	24.02%	15.67%	14.59%	-2.16%	12.51%
2013年2月	3.54	3.70	3.56	2.13	1.57	-3.33%	-8.37%	-0.13%	-24.69%	-52.87%
2013年3月	3.31	3.36	2.96	1.98	1.55	-6.72%	-9.73%	-18.49%	-7.56%	-1.44%
2013年4月	4.25	4.15	3.84	2.95	2.32	25.09%	21.24%	25.87%	40.19%	40.17%
2013年5月	2.62	3.49	3.44	3.17	2.38	-48.20%	-17.25%	-10.84%	7.07%	2.76%
2013年6月	1.68	2.38	2.43	2.81	2.35	-44.59%	-38.21%	-34.91%	-12.09%	-1.44%
2013年7月	1.55	2.36	2.25	2.48	2.23	-7.95%	-0.99%	-7.75%	-12.34%	-5.38%
2013年8月	1.75	2.58	2.60	2.69	2.06	12.24%	9.01%	14.74%	7.98%	-7.59%
2013年9月	2.70	3.67	3.69	3.57	2.84	43.14%	35.16%	34.84%	28.16%	31.95%
2013年10月	3.47	4.51	4.43	4.31	3.51	24.95%	20.57%	18.24%	18.99%	21.16%
2013年11月	3.60	4.91	4.72	4.99	3.68	3.87%	8.47%	6.30%	14.57%	4.71%
2013年12月	3.69	4.21	4.02	3.89	3.36	2.43%	-15.19%	-15.95%	-24.81%	-9.00%
2014年1月	4.84	5.28	4.57	3.88	3.50	27.20%	22.46%	12.73%	-0.28%	4.12%
2014年2月	4.68	5.03	4.35	3.30	2.64	-3.34%	-4.75%	-4.81%	-16.21%	-28.31%
2014年3月	4.82	5.00	4.38	3.57	3.02	2.82%	-0.69%	0.60%	7.73%	13.57%
2014年4月	3.92	4.23	3.63	3.38	2.36	-20.66%	-16.54%	-18.83%	-5.40%	-24.71%
2014年5月	3.01	4.01	3.88	3.62	2.28	-26.25%	-5.55%	6.84%	6.88%	-3.64%

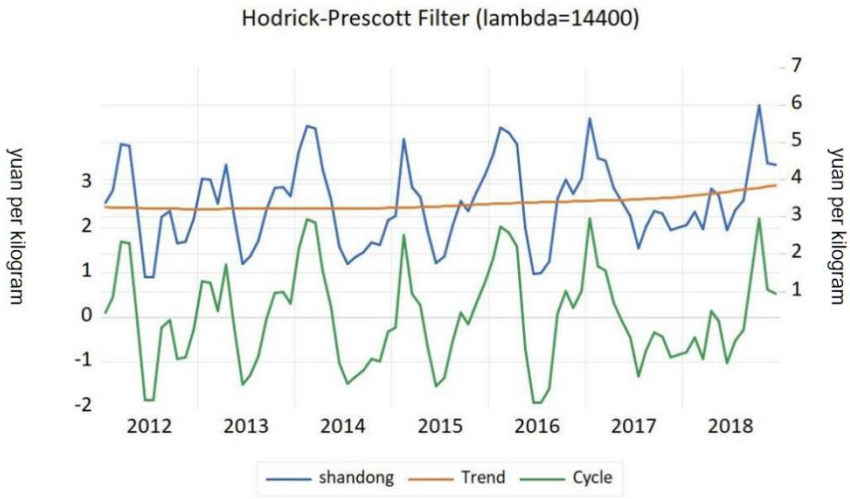
图 3　华南地区数据示意图

Figure 3　Schematic diagram of data South China

2.2　图像数据

根据计算出的各省份（自治区、直辖市）的平均价格和价格波动变化率绘制周度/月度平均价格趋势图、

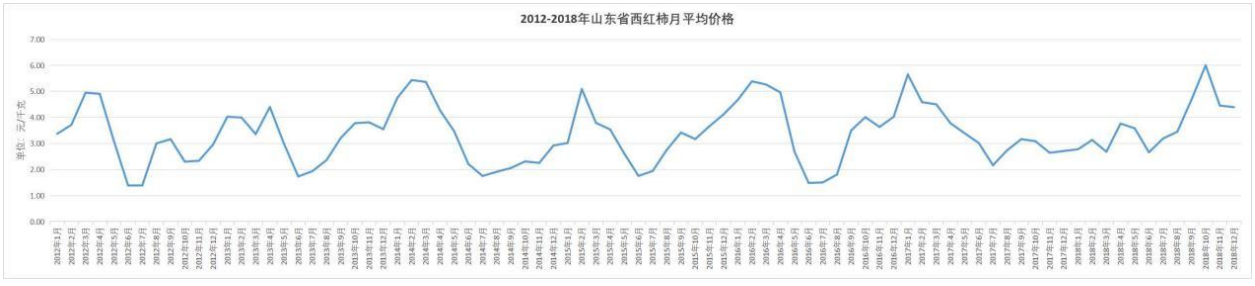
周度/月度价格波动变化率趋势图及西红柿月度价格趋势循环序列分解图。通过绘制折线图，可直观呈现西红柿价格在特定时段内的变化情况。图 4 展示了以月为单位时，山东省月平均价格趋势图、月度价格变化率图和月度价格趋势循环序列分解图，其中价格趋势循环序列分解图横坐标单位为年份，纵坐标单位为价格（元/公斤）。本数据集包含反映 25 个省（自治区、直辖市）价格波动的折线图共 75 张。



a 月度价格趋势循环序列分解图



b 2012–2018 年山东省西红柿月度价格变化率图



c 2012–2018 年山东省西红柿月平均价格趋势图

图 4　山东省月度价格趋势分析与周期分解

Figure 4　Monthly Price Trend Analysis and Cycle Decomposition for Shandong Province

3 数据质量控制与评估

本数据集采用的西红柿每日价格信息，主要来源为商务部主办的全国农产品商务信息公共服务平台。针对其中存在的部分数据缺失情况，通过整合其他权威渠道进行补充，包括全国农产品批发市场价格信息系统，并结合了针对实际市场的调研信息。数据处理过程主要分为两个阶段。第一阶段，对各省内所有批发市场报告的西红柿每日价格进行均值计算，所得结果即视为该省当日的西红柿代表价格。第二阶段，对上述价格序列实施质量审核与饱满度校验，其核心在于识别并标记显著偏离正常范围的异常值。所有被标记的疑似异常数据均需经过人工复核，以判定其是否应予剔除，从而确保后续分析所用数据的可靠性。在 Excel 中通过使用 COUNTIFS 函数统计缺失值数量来检验饱满度，若数据中存在零星缺失，则通过前后相邻交易日的价格均值进行插补，以维持数据连贯性；如数据完整性过低且难以有效填补，则予以剔除，减少大量空白数据在图表中造成的价格误差。

对最终成型的数据集，将进一步开展人工质量核验工作，重点核查其中是否存在公式运算失误、数据格式不统一等各类问题，以此确保数据集的连贯性与准确性。所有数据处理环节均严格遵循既定流程规范执行，最终形成能够反映各不同地理区域在各类时间尺度下平均价格水平及其变动情况的完整数据集。

4 数据价值

本数据集核心来源为国家层面官方平台所发布的数据，并融合了市场调研所得信息及其他权威统计平台的记录，最终形成一份覆盖 2012 至 2018 年西红柿批发市场交易的原始价格数据集合。本数据集在时间上具有连续性，数据质量可靠且来源专业，足以支撑各类经济学实证研究。

本数据集的主要应用价值包括：（1）提供了连续、准确的西红柿周度与月度均价序列，为后续的学术探索奠定了基础，有助于基于数据驱动的方法，对西红柿产业的产销全流程实现更具科学性、客观性与实用性的评估与施策[6]。（2）提供剔除季节波动后的价格趋势和价格波动变化率数据，为后续探究影响价格的因素和价格预测提供便利，达到提高研究效率的目的。（3）根据价格波动规律，有利于政府决策人员更好地了解西红柿价格波动现状，从而根据规律提前采取预防措施。（4）本数据集能够作为构建西红柿价格监测与预警体系所必需的数据基础，可为相关研究提供高质量的量化依据，有效提高价格趋势研判的准确性；该数据集的开放应用，有助于农业生产者及时掌握未来价格波动的预测资讯，并以此为基础，对生产与销售计划做出更为及时的调整，最终达到缓解因气候变化等外部冲击所导致的经济风险的目的。

作者分工职责

孔汇鑫（2000—），女，山东淄博人，硕士研究生，研究方向为农产品市场信息分析。主要承担工作：数据整理分析及论文撰写。

韩昀（1967—），女，山东省临沂人，硕士，副研究员，研究方向为农村区域发展。主要承担工作：文章框架构建指导。

胡林（1967—），男，内蒙古商都县人，博士，研究员，研究方向为科学数据管理。主要承担工作：数据分析方法指导。

曹姗姗（1984—），女，黑龙江哈尔滨人，博士，副研究员，研究方向为农林时空信息智能分析。主要承

担工作：论文撰写指导。

孔繁涛（1968—），男，山东滕州人，博士，研究员，研究方向为农业信息技术。主要承担工作：方案设计与组织实施。

参考文献

[1] 李君明, 项朝阳, 王孝宣, 等. “十三五”我国番茄产业现状及展望[J]. 中国蔬菜, 2021(2): 13－20. DOI: 10.19928/j.cnki.1000-6346.2021.2009. [LI J M, XIANG C Y, WANG X X, et al. Current situation of tomato industry in China during ‘the thirteenth Five-Year Plan’ period and future prospect[J]. China Vegetables, 2021(2): 13－20. DOI: 10.19928/j.cnki.1000-6346.2021.2009.]

[2] 谭雅蓉, 王一罡, 于金莹. 新冠肺炎疫情对北京市蔬菜价格影响实证研究[J]. 北方园艺, 2020(24): 153－161. DOI: 10.11937/bfyy.20201717. [TAN Y R, WANG Y G, YU J Y. An empirical study on the impact of COVID-19 on vegetable prices in Beijing city[J]. Northern Horticulture, 2020(24): 153－161. DOI: 10.11937/bfyy.20201717.]

[3] 张连城, 韩蓓. 中国潜在经济增长率分析: HP 滤波平滑参数的选择及应用[J]. 经济与管理研究, 2009, 30(3): 22－28, 86. DOI: 10.3969/j.issn.1000-7636.2009.03.004. [ZHANG L C, HAN B. Analysis of China’s potential economic growth rate: selection and application of HP filter smoothing parameters[J]. Research on Economics and Management, 2009, 30(3): 22－28, 86. DOI: 10.3969/j.issn.1000-7636.2009.03.004.]

[4] 韩蓓. HP 滤波法及其在地区潜在经济增长率测算中的应用[J]. 经济师, 2009(1): 28－29, 36. DOI: 10.3969/j.issn.1004-4914.2009.01.012. [HAN B. HP filtering method and its application in the calculation of regional potential economic growth rate[J]. China Economist, 2009(1): 28－29, 36. DOI: 10.3969/j.issn.1004-4914.2009.01.012.]

[5] 王桂芝, 陆金帅, 陈克垚, 等. 基于 HP 滤波的气候产量分离方法探讨[J]. 中国农业气象, 2014, 35(2): 195－199. DOI: 10.3969/j.issn.1000-6362.2014.02.012. [WANG G Z, LU J S, CHEN K Y, et al. Exploration of method in separating climatic output based on HP filter[J]. Chinese Journal of Agrometeorology, 2014, 35(2): 195－199. DOI: 10.3969/j.issn.1000-6362.2014.02.012.]

[6] 戴小文, 漆雁斌, 陈文宽. 农业现代化背景下大数据分析在农业经济中的应用研究[J]. 四川师范大学学报(社会科学版), 2015, 42(2): 70－77. DOI: 10.13734/j.cnki.1000-5315.2015.02.008. [DAI X W, QI Y B, CHEN W K. On the applications of big data analysis to the agricultural economy under the background of agricultural modernization[J]. Journal of Sichuan Normal University (Social Sciences Edition), 2015, 42(2): 70－77. DOI: 10.13734/j.cnki.1000-5315.2015.02.008.]

论文引用格式

孔汇鑫, 韩昀, 胡林, 等. 2012－2018 年基于 HP 滤波法的西红柿批发价格波动统计数据集[J/OL]. 中国科学数据, 2026, 11(1). (2026-02-04). DOI: 10.11922/11-6035.nasdc.2023.0022.zh.

数据引用格式

孙伟. 2012－2018 年基于 HP 滤波法的西红柿批发价格波动统计数据集[DS/OL]. V1. Science Data Bank, 2024. (2026-01-28). DOI: 10.57760/sciencedb.agriculture.00278.

A dataset of tomato wholesale price fluctuations based on the HP filtering method (2012-2018)

KONG Huixin[1,2], HAN Yun[3], HU Lin[1,2], CAO Shanshan[1,2,5*], KONG Fantao[4,5*]

1. Agricultural Information Institute of CAAS, Beijing 100081, P.R. China
2. National Agriculture Science Data Center, Beijing 100081, P.R. China
3. Institute of Agricultural Resources and Regional Planning, CAAS, Beijing 100081, P.R. China
4. Institute of Agricultural Economics and Development, CAAS, Beijing 100081, P.R. China
5. National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya 572024, P. R. China

*Email: caoshanshan@caas.cn (CAO Shanshan); kongfantao@caas.cn (KONG Fantao)

Abstract: Tomato is one of the most important vegetable crops in China, and the yield of tomato in China ranks first in the world. The historical price data and price fluctuation data of tomato are of great significance for analyzing the factors affecting price changes and for supporting macro-regulation by government authorities. The main source of this dataset is the national agricultural product business information public service platform, supplemented by other authoritative platforms and the actual market research data of the project team. According to the principles of authority, continuity and availability, the collected basic price data were supplemented, checked for completeness and stability, the irregular data were removed, and the missing values were filled, resulting in a dataset of tomato wholesale price fluctuations. Subsequently, trend components and price volatility indicators were calculated to further refine the dataset. In terms of time, it includes two temporal scales: weekly and monthly. In terms of space, 25 provinces (autonomous regions and municipalities) are divided into six geographical divisions, enabling the observation of weekly and monthly price fluctuation characteristics in each province. The dataset provides a data basis for scientific research on tomato price fluctuations and for monitoring and early warning in China.

Keywords: tomatoes; price fluctuations; scientific data; vegetable trade

Dataset Profile

Title	A dataset of tomato wholesale price fluctuations based on the HP filtering method (2012-2018)
Data authors	SUN Wei
Data corresponding author	SUN Wei (sunwei02@caas.cn)
Time range	2012–2018
Geographical scope	Northeast China (Liaoning Province, Heilongjiang Province), North China (Beijing, Tianjin, Hebei Province, Shanxi Province, Inner Mongolia Autonomous Region), Northwest China

	(Shaanxi Province, Gansu Province, Ningxia Hui Autonomous Region, Xinjiang Uygur Autonomous Region), East China (Shanghai, Jiangsu Province, Zhejiang Province, Anhui Province, Fujian Province, Jiangxi Province, Shandong Province), Southwest China (Sichuan Province, Yunnan Province), and South China (Henan Province, Hunan Province, Hubei Province, Guangdong Province, Guangxi Province).
Data volume	4.86 MB
Data format	*.xlsx, *.jpg
Data service system	https://doi.org/10.57760/sciencedb.agriculture.00278
Sources of funding	The Innovation Project Task of Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2016-AII)
Dataset composition	This data set contains: (1) 6 EXCEL files, comprising 4,200 monthly price time series and 18,200 weekly price time series of 25 provinces in the six major geographical divisions of China from 2012 to 2018, including the date, weekly average price, monthly average price, weekly average volatility and monthly average volatility. (2) 25 monthly price trend cycle sequence decomposition charts and 75 price fluctuation maps, including historical trajectories of weekly and monthly average prices, and monthly price volatility.