

基于EEMD-ARMA模型的中国农药产量预测

舒服华

(武汉理工大学管理学院,湖北 武汉 430070)

摘要:运用EEMD分解和ARMA模型相结合对我国农药产量进行了预测。首先利用EEMD技术对原始信号进行分解,得到频率单一、平稳的本征模函数和光滑的趋势余波,然后运用ARMA模型分别对本征模函数和余波进行预测,最后将二者的预测值合并,实现对目标的精确预测。结果显示:EEMD-ARMA组合模型的平均预测误差为3.231 35%,比单一ARMA模型的平均预测误差6.601 19%减小51.048 98%。由EEMD-ARMA模型预测得到2021年我国农药产量为200.435 2万t。

关键词:农药产量;预测;EEMD;ARMA

中图分类号:F767.2

文献标志码:A

Prediction of Pesticide Production Based on EEMD-ARMA

SHU Fuhua

(College of Management, Wuhan University of Technology, Hubei Wuhan 430070, China)

Abstract: Combining EEMD decomposition and ARMA model, the pesticide production in China is predicted. Firstly, EEMD technology is used to decompose the original signal to obtain the single frequency, stable eigenmode function and smooth trend residual wave. Then ARMA model is used to predict the eigenmode function and residual wave respectively. Finally, the predicted values of the two are combined to realize the accurate prediction of the target. The results show that the average prediction error of EEMD-ARMA model is 3.231 35%, which is 51.048 98% lower than that of single ARMA model 6.601 19%. The EEMD-ARMA model predicts that the pesticide production in China will be 200.435 2 tons in 2021.

Key Words: Pesticide production; Prediction; EEMD; ARMA

农药是现代农产品的产物,是防治农作物遭受病虫害侵扰,提高农产品品质,确保农作物丰收的重要农业生产物质。我国用占世界9%的耕地,养活了占世界22%的人口,粮食生产取得了“十四连丰”,除了种子、肥料、种植技术等手段全面提升外,农药在其中也发挥了重要作用。由于我国人口众多,每年要消耗大量的粮食、蔬菜、水果等农产品,农作物种植规模位居世界第一,为了保证农作物的丰收,因此,对农药的需求自然十分庞大,注定了我国成为全球第一大农药生产国和消费大国的地位,

最高时产量占世界的一半以上。随着绿色、低碳生产生活方式在全球的兴起,以及我国对农产品质量、安全性和农业生态保护的重视,农业生产也开始了转型发展,大力推进减肥减药、节水节能、绿色生态、科技强农等农业发展新模式,促进我国农业提质增效和高质量发展。近些年农药使用实现了零增长,农药产量也呈下降趋势,目前,产量下降到约为世界的1/3。科学预测我国农药的产量,是制订农药行业发展战略的重要依据,对优化调整产业结构,合理配置资源,促进我国农药行业稳步发展,为农业生产保驾护航,确保粮食等农作物持续丰收等具有重要意义。

ARMA模型对于预测光滑分布的时间序列效果较佳,但对于分布非光滑、不平稳的时间序列的

收稿日期:2021-11-21

基金项目:湖北省自然科学基金(2018CFB237)。

作者简介:舒服华(1964-),男,博士,教授,主要从事计量经济学研究工作,E-mail:cannon777@126.com。

预测效果则不尽如人意。经验模式分解(EMD)技术可将杂乱的时间序列分解为若干个比较平稳的本征模函数(IMF)和一个带有趋势的光滑余波(res.)。集成经验模式分解(EEMD)技术是EMD的深化,分解效果比EMD更佳。如果采用EEMD对性能较差的原始信号进行分解,再利用ARMA模型对平稳性较好的本征模函数和光滑的余波分别进行预测,然后将两种成分的预测结果叠加,则可实现对复杂无序的时间序列的精确预测。这一方法广泛应用于各种领域^[1-7]。由于我国农药产量数据组成的时间序列波动起伏,分布极其复杂。因此,运用EEMD-ARMA模型相结合的方法对其进行预测,这样可以有效提高预测的精度和可靠性。

1 EEMD-ARMA 模型

1.1 经验模态分解

EMD分解是将一个频率不规则的复杂信号分解为若干个频率单一的本征模函数和一个趋势性余波。分解出的本征模函数的数量取决于原始信号的特征。原始信号越复杂,分解出来的本征模函数越多。这些本征模函数彼此之间正交,且互不重复,每个IMF表达原始信号的某一些特性,余波的极值点少于2,且十分光滑。IMF只占原始信号极小的一部分,余波占原始信号的大部分^[8-9]。

设有一原始信号 $y(t)$,通过EMD分解,可得到 n 个本征模函数 $\text{imf}_1, \text{imf}_2, \dots, \text{imf}_n$,及其一个余波 res. ,这些本征模函数 $\text{imf}_1, \dots, \text{imf}_n$ 的频率大小依次递减,则 $y(t)$ 可表示为^[8-10]:

$$y(t) = \sum_{i=1}^n \text{imf}_i + \text{res.} \quad (1)$$

1.2 集成经验模态分解

运用EMD分解时,如果原始信号的极值点分布不均匀,很容易出现模态混叠现象,导致分解的IMF分量信息出现差错。EEMD有效地解决了这一问题。EEMD方法与EMD方法差别在于:EMD未对原始信号添加噪声信号,EEMD对原始

信号加入了噪声信号,即首先在待分解的信号中加入白噪声序列,然后再进行经验模态分解。由于加入的白噪声信号的频谱呈均匀分布,这样可使不同时间尺度的信号自动分布到合适的参考尺度上,避免了模态混叠现象的产生,且白噪声均值为零,经过多次平均后,噪声信号会逐步相互抵消,不会对原始信号的特征产生影响。加入白噪声信号后提取到第一个本征模函数 imf_1 ,对剩余的信号再次添加白噪声信号,提取到第二个本征模函数 imf_2 ,重复以上步骤,直到满足停止条件。与EMD相比,由于EEMD分解的本征模函数准确度较高,因此,分解出的本征模函数的数量比EMD要多。

1.3 EEMD-ARMA 模型原理

EEMD-ARMA模型就是将EEMD分解技术与ARMA预测模型相结合,实现对复杂时间序列的精确预测。原始信号经过EEMD分解后可得到若干本征模函数和一个余波。虽然本征模函数并不平滑,但分布比较平稳(这里所说的平稳不是指固定不变,而是在变化过程中均值保持基本稳定),且只占原信号极小一部分;而余波是一条极其光滑的曲线,占原信号的绝大部分。运用ARMA模型分别对本征模函数和余波分别进行预测,然后,将本征模函数和余波的预测值相加,就可还原对原始复杂时间序列的预测,它比单一使用ARMA模型预测的精度更高。

2 我国农药产量预测

图1为2005—2020年我国化学农药原药产量统计数据(数据来源于国家统计局)。从图1可知,从2005年开始,我国农药产量一直保持增长态势,2016年达到最高值477.8万t。随后,在国家调控政策的影响下,产量开始大幅下降,最近3年产量基本保持稳定,年产量在200万t左右。从数据分布特点看,其为波动起伏、非单调、凸凹不一致的复杂时间序列,使用传统ARMA模型预测效果可能不佳,比较适合运用EEMD-ARMA模型预测。

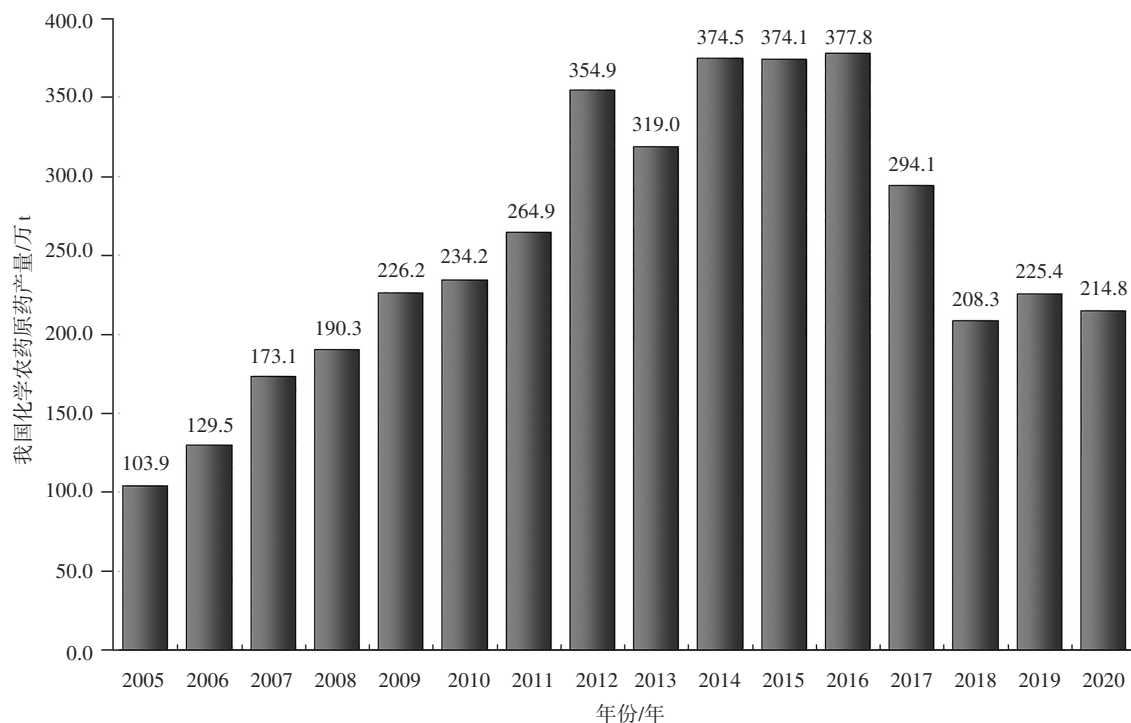


图1 我国农药产量统计数据

以2005—2020年我国农药产量数据为样本, 设其为时间序列 $y(t)$, 则:

$y(t)=[103.9, 129.5, 173.1, 190.3, 226.2, 234.2, 264.87, 354.91, 319, 374.5, 374.1, 377.8, 294.1, 208.3, 225.39, 214.8]$ 。

2.1 EEMD分解

对 $y(t)$ 进行EEMD分解, 结果如图2所示。从图2知, 时间序列 $y(t)$ 通过EEMD技术处理后, 分解出3个本征模函数和1个余波。从图2中还可以看出, 本征模函数虽然不光滑, 但变化比较平稳, 而余波信号十分光滑。各信号具体成分如下:

$\text{imf1}=[7.923\ 321\ 2, -2.969\ 475\ 7, 8.897\ 216\ 7, -2.001\ 321\ 8, 7.760\ 178\ 3, -16.894\ 29, -25.192\ 611, 28.897\ 557, -29.330\ 882, 8.565\ 262\ 6, 0.571\ 903\ 54, 26.227\ 609, -5.683\ 560\ 3, -30.775\ 568, 33.559\ 117, 31.168\ 851]$;

$\text{imf2}=[23.802\ 35, 9.509\ 532\ 8, -3.614\ 142\ 1, -16.568\ 035, -27.277\ 85, -24.972\ 014, -10.875\ 887, 6.090\ 989\ 2, 16.471\ 151, 28.621\ 697, 38.649\ 209,$

$26.591\ 257, -6.992\ 136\ 8, -41.971\ 24, -54.667\ 648, -21.402\ 953]$;

$\text{imf3}=[-4.152\ 053\ 2, -7.137\ 749\ 6, -8.182\ 116, -7.588\ 877\ 6, -5.661\ 76, -2.716\ 17, 0.880\ 084\ 83, 4.724\ 821\ 3, 8.322\ 393\ 2, 10.958\ 099, 12.357\ 853, 12.379\ 673, 10.881\ 576, 7.721\ 577\ 5, 2.757\ 695\ 6, -4.152\ 053\ 2]$;

$\text{res.}=[76.857\ 433, 129.769\ 91, 176.382\ 6, 216.695\ 49, 250.708\ 58, 278.421\ 88, 299.835\ 38, 314.949\ 09, 323.763, 326.277\ 12, 322.491\ 45, 312.405\ 97, 296.020\ 71, 273.335\ 64, 244.350\ 79, 209.066\ 1]$ 。

从分解结果可知, 本征模函数依次递减, 虽然数量较多, 但只占原始信号少量部分, 余波占原始信号的比重比较大。

为了验证EEMD-ARMA模型的性能, 分别采用传统ARMA方法和EEMD-ARMA方法对我国农药产量进行预测。

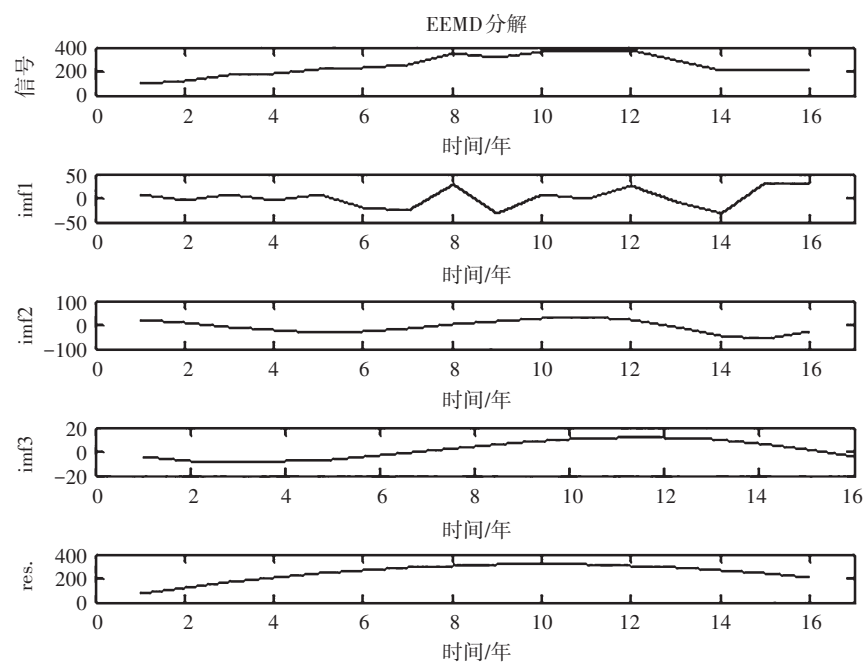


图2 EEMD分解结果

2.2 ARMA模型预测

以 $y(t)$ 样本建立ARMA模型,经检验 $y(t)$ 为非平稳系列,进行一次差分后, $d(y(t))$ 变为平稳系列,可以用于建模。通过比较,模型的最佳阶数为ARMA(2,1),估计模型的参数,结果表1。

根据估计得到的参数,得到预测方程:

$$d(y_i)=-49.082\ 22+0.818\ 222d(y_{i-1})\ 0.031\ 679d(y_{i-2})+\varepsilon_i-2.244\ 940\varepsilon_{i-1}\quad(2)$$

由预测方程(2)得到 y_i 的预测结果 y_j ,结果如表2所示,由于原始序列 $y(t)$ 进行过一次差分分析,且

表1 ARMA模型的参数估计结果

变量	系数	标准差	t-统计量	显著性概率
c	-49.082 22	152.575 6	-0.321 691	0.855 0
AR(1)	0.818 222	0.321 836	2.542 361	0.031 6
AR(2)	0.031 679	0.356 126	0.088 955	0.931 1
MA(2)	-2.244 940	0.921 153	-2.697 642	0.024 5
R^2	0.826 726	因变量均值	3.207 692	
R^2_{Adj}	0.768 968	因变量标准差	49.764 37	
回归标准差	23.919 65	赤池信息准则	9.434 935	
残差平方和	5 149.332	施瓦茨信息准则	9.608 765	
对数似然函数值	-57.327 07	汉南-奎因准则	9.399 204	
F-统计量	14.313 565	德宾-沃森统计	2.482 536	
模型显著性概率	0.000 899			

当期预测值与前期的预测值和误差有关,故前3期的值不能预测。

2.3 EEMD-ARMA 模型预测

2.3.1 本征模函数预测

以 imf1 (简写为 i_1) 为样本建立 ARMA 模型,经检验 imf1 为平稳序列,通过比较,模型的最佳阶数为 ARMA(1,2),估计模型的参数,得到预测方程为:

$$i_{1t}=0.841\ 439-0.191\ 883i_{1t-1}+\varepsilon_t-2.046\ 481\varepsilon_{t-1}-1.251\ 819\varepsilon_{t-2}\tag{3}$$

由预测方程(3)得到 i_1 预测结果 i_{1f} ,结果如表2。

以 imf2 (简写为 i_2) 为样本建立 ARMA 模型,经检验 imf2 为平稳序列,通过比较,模型的最佳阶数为 ARMA(2,2),估计模型的参数,得到预测方程为:

$$i_{2t}=2.405\ 598+1.603\ 460i_{2t-1}-1.320\ 021i_{2t-2}+\varepsilon_t-0.077\ 203\ 3\varepsilon_{t-1}-0.921\ 968\varepsilon_{t-2}\tag{4}$$

由预测方程(4)得到 i_2 预测结果 i_{2f} ,结果如表2。

以 imf3 (简写为 i_3) 为样本建立 ARMA 模型,经检验 imf3 为非平稳序列,进行一次差分后, $d(i_{3t})$ 变

为平稳系列,通过比较,模型的最佳阶数为 ARMA(1,2),估计模型的参数,得到预测方程为:

$$d\left(i_{3t}\right)=-4.488\ 957+0.931\ 470d\left(i_{3t-1}\right)+\varepsilon_t+0.659\ 283\varepsilon_{t-1}+0.165\ 805\varepsilon_{t-2}\tag{5}$$

由预测方程(5)得到 i_3 预测结果 i_{3f} ,结果如表2。

2.3.2 余波预测

以 res. (简写为 r) 为样本建立 ARMA 模型,经检验, r_t 为非平稳序列,进行一次差分后, $d(r_t)$ 平稳系列,通过比较,模型的最佳阶数为 ARIMA(2,2),估计模型的参数,得到预测方程为:

$$d\left(r_t\right)=8.273\ 800+1.921\ 509d\left(r_{t-1}\right)-0.957\ 036d\left(r_{t-2}\right)+\varepsilon_t-0.073\ 704\varepsilon_{t-1}-0.925\ 682\varepsilon_{t-2}\tag{6}$$

由预测方程(6)得到 r 预测结果 r_f ,结果如表2。

2.3.3 合并各分量预测值

将 imf1, imf2, imf2, res. 的预测值合并,最终获得 EEMD-ARMA 模型的预测值为:

$$y_{1f}=i_{1f}+i_{2f}+i_{3f}+r_f\tag{7}$$

结果如表2所示。

表2 预测结果及比较

年份/年	实际值/万 t	ARMA 模型预测		EEMD-ARMA 模型预测					预测值 y_{1f}	误差/%
		预测值 y_f	误差/%	i_{1f}	i_{2f}	i_{3f}	r_f			
2005	103.9	-	-		-	-	-	-	-	-
2006	129.5	-	-	-0.517 47	-	-	-	-	-	-
2007	173.1	-	-	6.590 70	-	-	176.392 8	-	-	-
2008	190.3	202.218 3	6.262 91	-2.354 98	-14.295 80	-7.686 52	216.703 8	192.366 5	1.085 93	
2009	226.2	228.003 8	0.797 43	-2.224 26	-31.961 30	-5.704 23	250.709 2	210.819 4	-6.799 56	
2010	234.2	253.234 2	8.127 34	-21.361 80	-23.894 30	-2.755 65	278.421 9	230.410 1	-1.618 24	
2011	264.9	281.814 8	6.397 41	-17.397 20	-9.648 55	0.903 50	299.830 4	273.688 1	3.329 23	
2012	354.9	324.958 0	-8.439 32	16.197 34	11.495 25	4.7658 58	314.943 9	347.402 4	-2.115 36	
2013	319.0	347.758 3	9.015 14	-20.774 10	17.962 35	8.462 41	323.755 9	329.406 6	3.262 26	
2014	374.5	356.565 5	-4.788 91	8.243 49	17.540 28	11.282 96	326.270 2	363.336 9	-2.980 79	
2015	374.1	366.840 5	-1.940 54	9.412 79	36.645 84	12.152 80	322.485 5	380.696 9	1.763 41	
2016	377.8	350.124 2	-7.325 52	18.583 05	30.041 47	12.380 02	312.401 0	373.405 6	-1.163 17	
2017	294.1	304.674 9	3.595 67	-8.606 64	-5.836 93	10.844 10	296.019 0	292.419 5	-0.571 39	
2018	208.3	244.642 8	17.447 32	-13.458 50	-39.931 80	7.684 71	273.336 3	227.630 8	9.280 25	
2019	225.4	218.387 4	-3.106 88	38.688 06	-50.782 50	2.736 43	244.356 5	234.998 5	4.263 04	
2020	214.8	211.887 2	-1.356 07	26.738 32	-24.949 00	-4.173 98	209.076 1	206.691 4	-3.774 94	
2021	预测值	192.071 0	6.064 19	-7.624 15	53.727 79	-13.164 00	167.495 5	200.435 2	3.231 35	

从前文可知,ARMA模型的平均预测误差为6.601 19%,EEMD-ARMA模型的平均预测误差为3.231 35%,EEMD-ARMA模型的平均预测误差比ARMA模型的平均预测误差减小了51.048 98%。再考察两个模型对2021年我国化学农药原药产量的预测。ARMA模型的预测值为192.071 0万t,比2020年的产量减少了10.581 5%,EEMD-ARMA模型的预测值为200.435 2万t,比2020年产量减少了6.687 752%。虽然我国农药产量呈递减之势,2020年我国新冠肺炎疫情肆虐,产量也只减小了4.702 8%,2021年我国基本控制住了疫情,各项生产基本

恢复正常,农药产量不太可能出现这样大幅度的下降。可见,ARMA模型的预测值显然过低,可信度不大。而EEMD-ARMA模型预测值比2020年降幅稍大,从国家要实现农药化肥零增长的政策措施来看,农药产量会逐步减少,这一预测结果与实际情况基本吻合,有较高的可信度。可见,不论是从预测精度还是预测结果来看,EEMD-ARMA模型优势都十分明显。两种模型的预测曲线如图3所示。从图3可见,EEMD-ARMA模型的预测曲线比实际值曲线更为贴近。

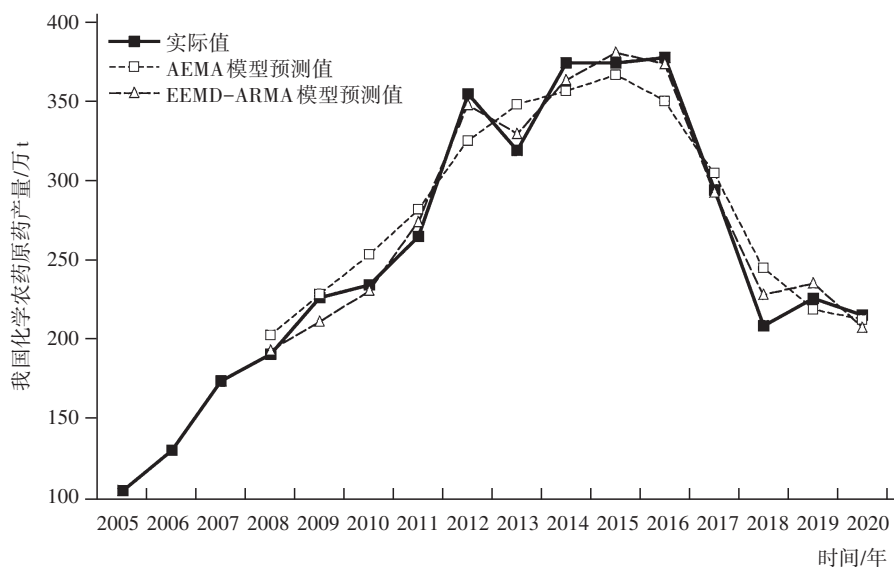


图3 模型的预测曲线及比较

3 结语

农药是预防、消灭、控制危害农作物病虫害和其他有害生物的有效武器,也是调节、控制、促进农作物生长的有效手段,对确保农作物正常生长、农产品质量、农作物增产丰收具有不可替代的作用,是农业活动的重要生产资料。科学、规范使用农药,不仅不会对农产品安全性带来影响,而且还可以改善和提高农产品的品质。因此,人们对农药要正确认识,理性客观看待。没有农药,就没有丰衣足食,没有美味佳肴,没有粮食安全,没有国家安定。农药是现代农业的有机组成部分,对保证农业丰收,确保居民的菜篮子、米袋子,维护粮食安全,改善老百姓饮食生活水平,维护社会稳定具有十分重要的作用。虽然农药的益处尽人皆

知,但农药的毒副作用也是显而易见的。过量、泛滥使用农药,不仅会对水、土、大气等环境造成污染,而且影响农产品的质量和安全性,危害消费者的身体健康。

因此,必须科学、正确、依归、依法、适量、适度使用农药,充分发挥农药的作用,把农药的副作用降到最低。为了维护人民群众的身体健康,推进农业的可持续发展,国家开展了化肥农药使用量零增长行动,经过5年的不懈努力,2020年,我国顺利实现了化肥农药零增长的预期目标,完成了农业生产方式的绿色转型和高质量发展。国家也加强了对农药行业调整整顿,压缩产能,优化升级产业结构,农药的产量出现了大幅下降。经过整顿升级之后,近些年我国农药产量基本稳定,使供需关系基本保持平衡,产品结构进一步优化,一大批高效、低毒、

低残留的环保农药新产品问世,取得了良好效果,为建设和谐环保的社会贡献了力量。准确预测我国农药产量,对制订农药产业发展战略,控制总量,调优结构,促进农药行业转型发展,使其更好地服务农业生产等具有重要的意义。

本研究运用EEMD-ARMA模型对我国农药产量进行了预测。首先利用EEMD技术对原始信号进行分解,得到频率单一的3个本征模函数和1个趋势型的光滑余波,然后利用ARMA模型对它们进行逐一预测,最后将它们的预测值进行合并,实现对原始数据序列的高精度预测。结果显示:EEMD-ARMA模型的平均预测误差为3.231 35%,比ARMA模型的平均预测误差6.601 19%减小51.048 98%。由EEMD-ARMA模型预测得到2021年我国农药产量为200.435 2万t。

参考文献:

- [1] 李智强,邹红霞,齐斌,等.基于EEMD-ARIMA的年降水预测拟合模型研究[J].计算机应用与软件,2020,37(11):46-50,78.
- [2] 李宁,刘瑜,王大刚.基于ARIMA和EEMD的东江流域季节降水预报研究[J].人民珠江,2019,40(3):52-58,70.
- [3] 张莹,谭艳春,彭发定,等.基于EEMD和ARIMA的海温预测模型研究[J].海洋学研究,2019,37(1):9-14.
- [4] 郭建鹏,余颖铃,温步瀛.基于EEMD-ARIMA模型的地区月负荷量预测[J].电力工程技术,2018,37(6):28-32,74.
- [5] 张金良,王明雪.基于EEMD,SVM和ARMA组合模型的电价预测[J].电力需求侧管理,2020,22(3):63-68.
- [6] 田波,朴在林,郭丹,等.基于改进EEMD-SE-ARMA的超短期风功率组合预测模型[J].电力系统保护与控制,2017,45(4):72-79.
- [7] 冉茂亮,陈彦如,杨新彪.基于EEMD-LMD-LSTM-LEC深度学习模型的短时物流需求预测[J].控制与决策,2021,40(8):1134-1143.
- [8] 谢丽蓉,王斌,包洪印,等.基于EEMD-WOA-LSSVM的超短期风电功率预测[J].太阳能学报,2021,42(7):290-296.
- [9] 刘洋,谢庐乐,汪寿阳,等.基于EEMD-AWNN集成学习的中国经常账户预测研究[J].系统工程理论与实践,2021,41(5):1240-1251.
- [10] 舒服华.基于EMD-ARMA的福建省GDP增长速度预测[J].三明学院学报,2018,35(5):12-16.