

基于 MINE 算法的烤烟含氮化合物与 HCN 关系分析

薛宝燕¹, 郭东锋^{2*}, 邵伏文¹, 刘炎红¹, 郭建¹, 裴洲洋³

(1.安徽省烟草专卖局(公司)烟叶管理处, 安徽 合肥 230022; 2.安徽中烟工业有限责任公司技术中心, 安徽 合肥 230088; 3.安徽皖南烟叶有限责任公司, 安徽 宣城 242000)

摘要: 为了考察烤烟烟气氢氰酸释放量与烟叶中主要含氮化合物的关系, 采集全国主要植烟区域的 60 份烤烟样品, 运用机器学习算法新方法 MINE 对氢氰酸与主要含氮化合物之间的关系进行了分析。结果表明: 烟气氢氰酸的释放量与烟叶中的含氮化合物关系最为密切, 与蛋白质、烟碱、游离氨基酸总量 MINE 算法关系强度 *MIC* 分别达 0.396 7、0.403 4 和 0.352 1, 与游离氨基酸中的半胱氨酸、谷氨酸、精氨酸 *MIC* 也分别达 0.388 7、0.326 3 和 0.353 7, 含氮化合物含量显著影响烟气氢氰酸的释放量。

关键词: 烤烟; 氢氰酸; 含氮化合物; MINE

中图分类号: TS41⁺¹ 文献标志码: A 文章编号: 1007-1032(2015)05-0496-07

Study on relationship between HCN yields and nitrogen components of flue-cured tobacco leaf based on MINE algorithm

Xue Baoyan¹, Guo Dongfeng^{2*}, Shao Fuwen¹, Liu Yanhong¹, Guo Jian¹, Pei Zhouyang³

(1. Anhui Provincial Corporation of China National Tobacco Corporation, Hefei 230022, China; 2. Technology Center of Anhui Cigarette Industrial Company Co., Ltd., Hefei 230088, China; 3. Anhui Wannan Tobacco Company Co., Ltd., Xuancheng, Anhui 242000, China)

Abstract: To explore the relationship between HCN yields in smoke stream and main nitrogen compounds in flue-cured tobacco leaf, 60 samples were selected from main tobacco-planting areas in China. The new machine learning algorithm named MINE (maximal information-based nonparametric exploration) was used to study the relationship between the HCN and the main nitrogen compounds. The results showed that HCN yields in smoke stream exhibited close relationship with nitrogen compounds, the relationship strength (*MIC*) with protein, nicotine and total content of amino acid reached 0.396 7, 0.403 4 and 0.352 1, respectively; and the *MIC* with cysteine, glutamic acid and arginine reached 0.388 7, 0.326 3 and 0.353 7, respectively; indicating these nitrogen compounds significantly influence the HCN yields in smoke stream.

Keywords: flue-cured tobacco; HCN; nitrogen components; maximal information based nonparametric exploration (MINE)

大量研究表明, 卷烟主流烟气中的氢氰酸是烟叶中含氮化合物通过燃烧裂解产生的^[1-3], 氢氰酸的释放受到烟叶原料产地、部位、年份等^[4-5]的影响, 卷烟的温度平衡时间^[6]、辅材^[7]、配方打叶^[8]、工艺参数^[9]、化学添加剂^[10-11]、抽吸方式^[12]等对氢

氰酸的释放也有较大影响, 所以, 氢氰酸释放量的大小复杂而多变。围绕主流烟气释放研究, 运用统计学、数据挖掘或机器学习等手段^[13-14]研究了焦油、氢氰酸、一氧化碳等烟气成分。MINE (maximal information-based nonparametric exploration)是 2011

年由 David N. Reshef 等^[15]研究开发的一种探索连续型变量间关系的新算法,目前在烟草科研中鲜有应用。笔者拟以氢氰酸与烟叶中主要含氮化合物的关系疏密为研究对象,运用 MINE 算法探索氢氰酸与含氮化合物之间的关系,以期为卷烟“减害降焦”提供技术参考。

1 材料与方法

1.1 材料

选取 2012 年度贵州(兴义、贵阳),辽宁凤城,云南(大理、石林、玉溪),江西赣州,陕西(商洛、安康),湖南(凤凰、桂阳、宁乡),河南(宜阳、三门峡),广西百色,湖北宜昌,安徽宣城,重庆黔江濯水,福建南平,四川凉山烟区的 B2F、C3F、X2F 共计 60 个样品,采用统一辅材进行单料烟的卷制。

1.2 检测方法

1.2.1 烟气氢氰酸检测

按照 YC/T 403—2011 测定烟样氢氰酸。

1.2.2 烟叶化学成分的测定

氨基酸含量的测定按照 YC/T 282—2009 进行;总氮及烟碱的测定按照 YC/T 160—2002 和 YC/T 161—2002 进行;烟叶硝酸根离子和蛋白质含量测定分别参照文献[16]和[17]方法进行。

1.3 数据处理方法

数据整理在 Microsoft Excel 中进行,统计分析

及作图在 R3.1 开源软件平台进行。采用拉依达(PauTa)准则(3σ 准则)对数据进行清洗整理,对整理后出现的缺失值采取 KNN(最近邻法)进行补缺。HCN 与含氮化合物间相关分析采用皮尔逊(Pearson)相关分析法。MINE 算法在 R3.1 开源软件平台实现。

2 结果与分析

对烟叶中氢氰酸以及主要含氮化合物和氨基酸含量的统计描述列于表 1。氢氰酸含量变幅为 154~316 mg/支,数据形态呈现出右偏态尖顶峰,变异系数达 16.43%,数据分布较为离散;烟叶中硝酸根离子变异系数为 52.10%,变异较大,数据形态呈现出右偏态平顶峰,变幅在 0.004 8%~0.068 2%,数据分布离散较大;烟叶中总氮和蛋白质变异系数均在 15%以下,总氮变幅在 1.14%~2.09%,蛋白质变幅在 3.74%~6.49%,数据分布相对集中;烟碱含量变异系数 33.70%,变幅在 1.13%~3.94%,数据呈现出右偏态平顶峰,数据分布较为离散;氨基酸含量变异广泛,除亮氨酸变异系数为 19.17%外,其余氨基酸变异系数均超过了 20%,其中苏氨酸、谷氨酸、丙氨酸、缬氨酸、蛋氨酸、色氨酸、赖氨酸、脯氨酸和氨基酸总量变异在 20%~50%,而天冬氨酸、甘氨酸、半胱氨酸、异亮氨酸、络氨酸、苯丙氨酸、组氨酸和精氨酸变异系数均超过了 50%,氨基酸含量在烟叶中存在较大变异,数据离散程度较高。

表 1 烤烟烟叶氢氰酸与含氮化合物含量统计描述

化学成分	极小值	极大值	均值	标准差	偏度	峰度	变异系数/%
氢氰酸	154.000 0	316.000 0	214.094 3	35.170 1	0.739 4	0.565 7	16.43
硝酸根离子	0.004 8	0.068 2	0.030 9	0.016 1	0.434 3	-0.264 3	52.10
蛋白质	3.740 0	6.490 0	5.102 5	0.618 4	0.243 0	0.317 4	12.12
总氮	1.140 0	2.090 0	1.560 2	0.229 3	0.443 5	-0.390 7	14.70
烟碱	1.130 0	3.940 0	2.260 9	0.7620	0.402 1	-0.891 1	33.70
天冬氨酸	0.066 8	0.625 2	0.205 3	0.1149	1.386 9	2.3841	55.99
苏氨酸	0.133 5	0.382 9	0.233 1	0.071 3	0.559 3	-0.942 4	30.60
谷氨酸	0.045 2	0.936 3	0.396 3	0.196 3	0.433 9	-0.147 7	49.54
甘氨酸	0.003 1	0.109 7	0.025 2	0.024 1	2.465 2	5.734 8	95.83
丙氨酸	0.012 1	0.394 1	0.216 8	0.092 8	-0.399 9	0.291 8	42.82
半胱氨酸	0.000 0	0.016 5	0.005 8	0.003 8	1.071 3	0.934 7	65.14
缬氨酸	0.016 9	0.059 8	0.031 6	0.010 2	0.794 6	0.213 1	32.35

表 1(续)

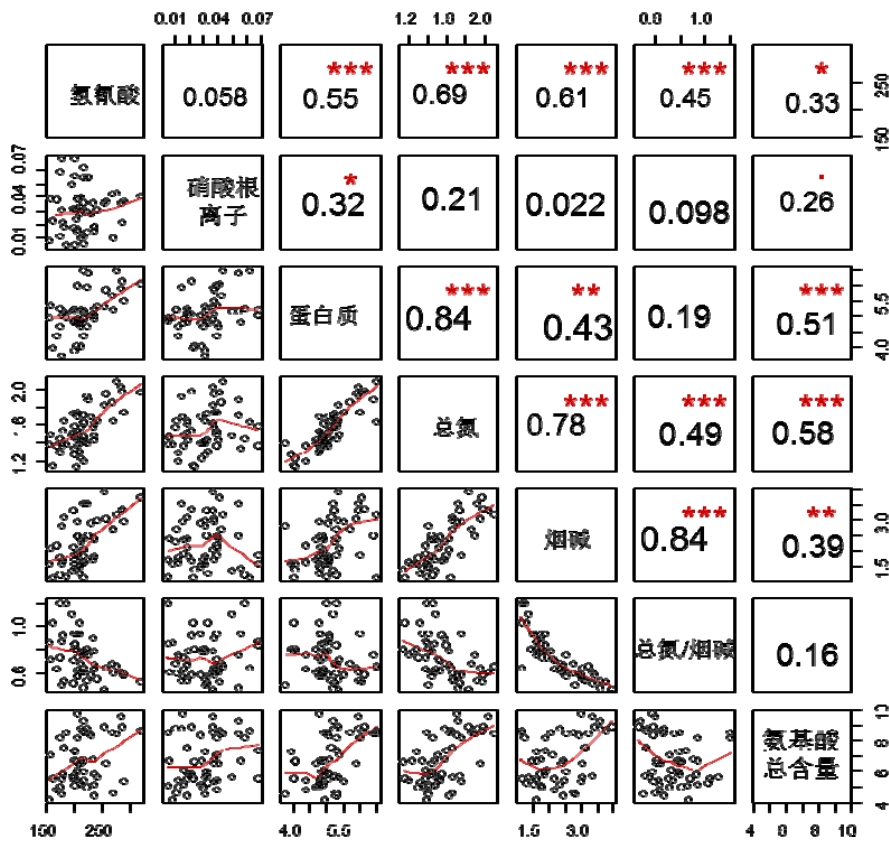
化学成分	极小值	极大值	均值	标准差	偏度	峰度	变异系数/%
蛋氨酸	0.011 7	0.060 3	0.033 1	0.013 1	0.529 9	-0.525 4	39.59
异亮氨酸	0.000 0	0.061 2	0.025 4	0.020 5	-0.145 3	-1.559 2	80.51
亮氨酸	0.384 6	0.876 8	0.586 6	0.112 5	0.294 9	-0.335 9	19.17
苯丙氨酸	0.000 0	0.651 6	0.244 9	0.168 8	0.293 6	-0.200 9	68.92
组氨酸	0.008 4	0.216 2	0.091 4	0.048 1	0.688 5	-0.033 9	52.66
色氨酸	0.027 6	0.230 4	0.113 7	0.052 6	0.370 2	-0.722 1	46.30
赖氨酸	0.019 7	0.294 1	0.144 5	0.059 8	0.124 6	-0.206 7	41.42
精氨酸	0.000 8	0.052 2	0.017 1	0.009 5	1.375 8	2.584 6	55.65
脯氨酸	1.474 2	6.629 0	4.317 0	1.199 3	0.140 3	-0.661 9	27.78
氨基酸总量	4.168 1	9.776 7	6.751 3	1.583 7	0.275 4	-1.265 8	23.46

氢氰酸含量单位为 mg/支；氨基酸含量单位为 mg/g；烟叶化学成分含量为质量分数(%)。

对氢氰酸与主要含氮化合物及氨基酸总量进行简单相关分析，结果见图 1。氢氰酸与烟叶中总氮、烟碱、蛋白质、氨基酸总含量呈显著正相关，与氮碱比呈显著负相关。图 2 表明，氢氰酸释放量与丙氨酸、半胱氨酸、络氨酸、组氨酸、色氨酸和精氨酸呈显著正相关。说明氢氰酸释放不仅与含氮化合物总量相关，而且与含氮化合物的种类也存在

密切关系。

由于相关分析只能简单描绘 2 个变量间关系的疏密程度，并能严格定义变量间的因果逻辑关系，因此，为了进一步探讨氢氰酸与烟叶中含氮化合物的关系，采取 MINE 算法对 2 组变量间的关系进一步探索。



*表示达到 0.05 显著水平；**表示达到 0.01 极显著水平；***表示达到 0.001 极显著水平。

图 1 氢氰酸与主要含氮化合物相关分析矩阵

Fig.1 Scatterplot matrix between HCN and main nitrogen compositions

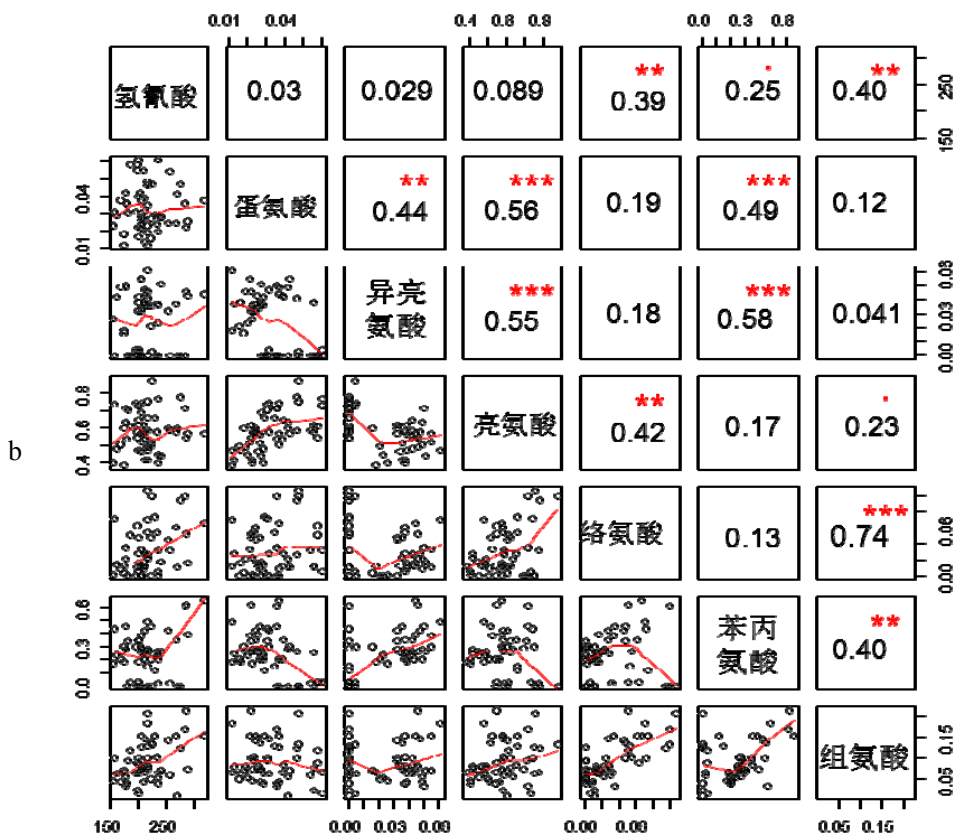
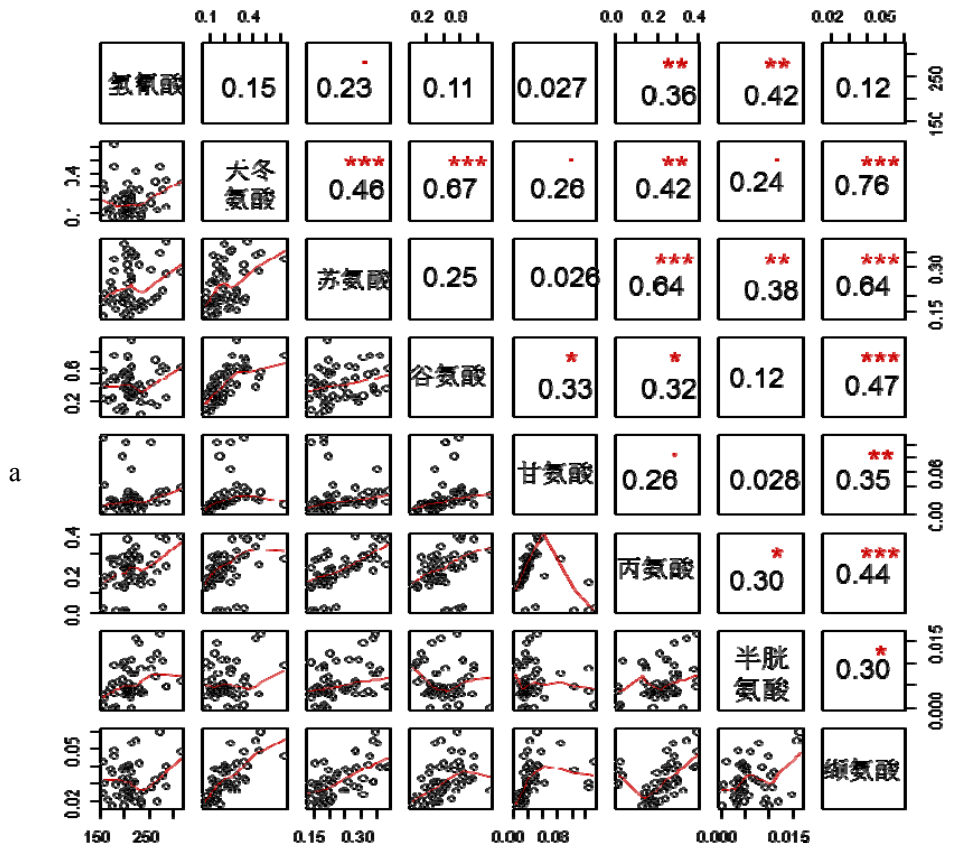


表 2 基于 MINE 算法的氢氰酸与含氮化合物关系信息提取

Table 2 Extracted information for the relationship between HCN and nitrogen compositions based on MINE algorithm

应变变量	自变量	MIC	P *
氢氰酸	总氮/烟碱	0.479 7***	< 0.001
	总氮	0.456 8**	< 0.01
	烟碱	0.403 4**	< 0.01
	蛋白质	0.396 7**	< 0.01
	半胱氨酸	0.388 7*	< 0.05
	精氨酸	0.353 7*	< 0.05
	氨基酸总量	0.352 1*	< 0.05
	谷氨酸	0.326 3*	< 0.05
	络氨酸	0.308 0	> 0.05
	组氨酸	0.306 7	> 0.05
	脯氨酸	0.305 6	> 0.05
	苏氨酸	0.301 3	> 0.05
	异亮氨酸	0.291 7	> 0.05
	亮氨酸	0.290 7	> 0.05
	丙氨酸	0.283 0	> 0.05
	色氨酸	0.276 7	> 0.05
	硝酸根离子	0.274 7	> 0.05
	赖氨酸	0.251 1	> 0.05
	甘氨酸	0.249 6	> 0.05
	缬氨酸	0.249 0	> 0.05
苯丙氨酸	0.248 9	> 0.05	
蛋氨酸	0.241 5	> 0.05	
天冬氨酸	0.216 5	> 0.05	

*是基于 $n=60$ 样本容量计算的 MIC 指标 P 值。

3 小结与讨论

氢氰酸是卷烟烟气中 7 种有害成分之一，对卷烟的危害性指数有着重要的影响。对氢氰酸和烟叶中主要含氮化合物的相关分析表明，主流烟气中的氢氰酸与烟叶中的总氮、烟碱、蛋白质、氨基酸总含量以及游离氨基酸中的丙氨酸、半胱氨酸、络氨酸、组氨酸、色氨酸和精氨酸呈显著正相关关系，与氮碱比呈显著负相关关系。运用 MINE 算法考察氢氰酸与主要含氮化合物的关系表明，氢氰酸与总氮、烟碱、蛋白质以及氨基酸关系最为密切。

综合分析认为，烟气中氢氰酸的释放量与烟叶中的总氮关系最为密切，同时氢氰酸的释放还与蛋白质、游离氨基酸中的特定氨基酸(总量、半胱氨酸、

谷氨酸、精氨酸)高度相关，与郑宏伟等^[20]研究天门冬氨酸的裂解行为结论有相似之处。

由于氢氰酸是卷烟在 700~1 000 °C 燃烧时产生的，而卷烟燃烧又受辅材、配方、燃烧状态等的影响，烟叶中的含氮化合物还包括叶绿素、TSNAs、氨、酰胺等重要化合物，对这些化合物还需要进一步跟踪检测。MINE 算法虽可以得出氢氰酸与含氮化合物的关系亲疏，但是其具体函数或非函数关系需要进一步探索研究，由于本研究所用方法为机器学习 MINE 算法，与传统线性分析略有不同，因此，该方法的进一步拓展应用尚需要更深入的实践和研究。

参考文献:

- [1] 谢剑平, 刘惠民, 朱茂祥. 卷烟烟气危害性指数研究[J]. 烟草科技, 2009(2): 5-15.
- [2] 谢剑平. CORESTA 热点研究问题[J]. 中国烟草学报, 2006, 12(1): 13-16.
- [3] 王彦亭, 谢剑平, 张虹, 等. 降低卷烟烟气中有害成分的技术研究及应用[J]. 中国烟草学报, 2003, 9(3): 3-9.
- [4] 王涛, 鲍峰伟, 王刘胜, 等. 卷烟主流烟气七种有害成分释放量与烟叶产地、年份之间的关系[J]. 湖北农业科学, 2014(6): 1330-1333.
- [5] 陈敏, 郭吉兆, 郑赛晶, 等. 烟叶部位、产地与卷烟主流烟气 7 种有害成分释放量关系研究[J]. 中国烟草学报, 2012, 18(5): 16-22.
- [6] 许永, 张霞, 刘巍, 等. 温湿度平衡时间对卷烟主流烟气氢氰酸释放量影响的研究[J]. 应用化工, 2012(7): 1140-1142.
- [7] 曾万怡, 向能军, 龚为民, 等. 纸质滤嘴棒对卷烟主流烟气中有害物质的影响[J]. 中国造纸, 2014(6): 35-39.
- [8] 赵乐, 彭斌, 于川芳, 等. 辅助材料设计参数对卷烟 7 种烟气有害成分释放量的影响[J]. 烟草科技, 2012(10): 46-50.
- [9] 陈昆燕, 冯广林, 李东亮, 等. 打叶复烤各工序对卷烟主流烟气氢氰酸与苯并[a]芘和巴豆醛释放量的影响[J]. 湖南农业大学学报: 自然科学版, 2014, 40(2): 144-147.
- [10] 者为, 廖头根, 王明锋, 等. 均苯三甲酸合铜选择性降低卷烟主流烟气中的 HCN[J]. 烟草科技, 2014(10): 45-48.
- [11] 周宛虹, 孙文梁, 王律, 等. 胺基修饰的介孔二氧化硅选择性降低卷烟烟气中的氢氰酸[J]. 烟草科技, 2013(4): 42-45.

- [12] 许永, 张霞, 刘巍, 等. 抽吸方式对卷烟主流烟气中氢氰酸释放量的影响[J]. 中国烟草学报, 2011, 17(6): 4-7.
- [13] 王浩军, 郭东锋, 杜丛中, 等. 基于关联规则的皖南焦甜香烟叶外观特征分析[J]. 中国烟草科学, 2014, 35(5): 93-97.
- [14] 侯小东, 杜咏梅, 刘新民, 等. 分类树模型在烟草农业研究中的应用[J]. 中国烟草科学, 2012, 33(5): 93-96.
- [15] Reshef D, Reshef A, Finucane K, et al. Detecting novel associations in large data sets[J]. Science, 2011, 334: 1518-1524.
- [16] 许自成, 张莉, 肖汉乾, 等. 烤烟硝酸盐、亚硝酸盐含量与若干品质指标的典型相关分析[J]. 郑州轻工业学院学报, 2005(1): 43-46.
- [17] 王晶, 胡立中, 朱栋梁, 等. 烟叶中游离态氨基酸与卷烟主流烟气中氢氰酸的相关关系[J]. 光谱实验室, 2012(6): 3793-3797.
- [18] 拉罗斯. 数据挖掘方法与模型[M]. 刘燕权, 等, 译. 北京: 高等教育出版社, 2011: 116-180.
- [19] 黄朝章, 蔡国华, 赵艺强, 等. 单料烟主流烟气 HCN 与烟叶常规化学成分的相关性[J]. 烟草科技, 2013(2): 62-64.
- [20] 郑宏伟, 刘新建, 崔伟, 等. 天门冬氨酸热裂解行为对卷烟烟气成分的影响[J]. 湖北农业科学, 2014(9): 2149-2152.

责任编辑: 罗慧敏

英文编辑: 罗维

简 讯

“十二·五”国家科技支撑计划“农村物联网综合信息服务科技工程”项目完成验收

2015 年 5 月 29 日,“十二·五”国家科技支撑计划“农村物联网综合信息服务科技工程”项目验收会在湖南农业大学召开。

“农村物联网综合信息服务科技工程”项目主要是解决湖南、湖北、广东和重庆等国家农村农业信息化示范省“三农”问题的重大需求,以信息技术改造传统农业为出发点,紧密结合农业生产、经营、流通等环节,力求攻克一批对我国农业未来发展有重大影响和重大应用前景的现代农业信息关键技术。重点研发农村信息资源可视化和智能化体验式公共服务平台,研制适合农业产业化生产过程的关键农业信息感知与传输关键技术及低成本终端;构建具备科学性、准确性、实时性的综合、专业的农村农业信息服务平台。面向大田种植、设施园艺、畜禽养殖、水产养殖等应用领域开展农业产业的物联网关键技术示范与应用。

项目于 2012 年经国家科技部批准立项,下设 8 个课题,项目组织单位为湖南省科技厅,有 11 个省的大学、科研院所、企业参加了项目研究。经过 3 年的实施,各课题组已经完成了课题的任务指标,共发表论文 153 篇,其中被 SCI/EI 收录 80 篇,获得国家发明专利 31 项、实用新型专利 10 项,申请国家发明专利 40 项、实用新型专利 5 项,获得软件著作权 111 项,培养硕士、博士研究生 151 名,项目还在全国 11 个省市建立了物联网示范基地 94 个。

湖南作为国家农村农业信息化示范省,由湖南农业大学、湖南大学等单位承担了“农村物联网基础平台共性关键技术研究”、“农村物联网信息服务系统集成与应用研究”、“农村物联网行业信息系统关键技术研究”3 个课题的研究任务。研制了一批农用信息感知传感器及数据采集终端,开发了基于云服务的水产、大田、茶园、生猪、温室大棚物联网监控系统、建立了生猪、茶叶和食用菌全产业链质量安全生产管理与追溯系统。课题技术产品在望城国家农业科技园和岳阳国家农业科技园区内进行了推广应用,带动了农业产业与信息化的融合,促进了农业产业发展。