

大数据与现代农业

温孚江

山东农业大学 校长 教授
农业大数据联盟 理事长



一、大数据在农业现代化中的作用

● 1、生产环节更精准

● 2、产品流通更有序

● 3、管理决策更科学

通过预测：提高效益，改善农业的供给侧结构性改革



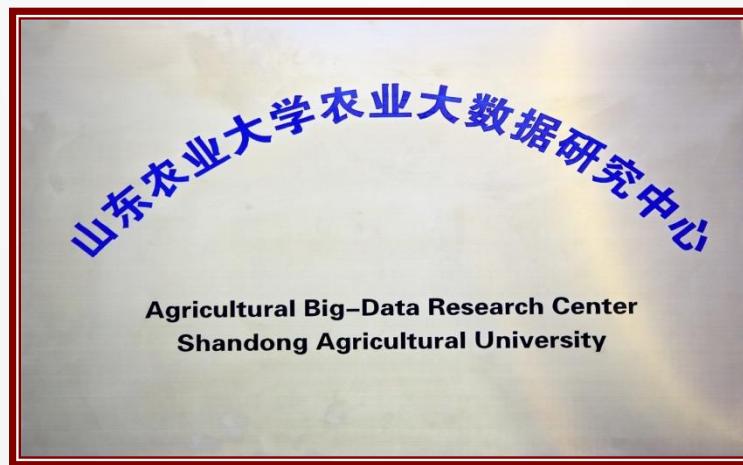
二、山东农业大学大数据工作



1、

搭建了研究平台

(1) 成立山东农业大学农业大数据研究中心



➤2013年5月16日，山东农业大学农业大数据研究中心挂牌成立。中心网址：
<http://www.nydata.com.cn>

➤农业大数据研究中心建设



1、

搭建了研究平台

(2) 成立农业大数据产业技术创新战略联盟



农业大数据产业技术创新战略联盟成立于2013年6月，是国内首家农业大数据联盟，现有联盟成员100余家，计划发展到1000家（2017年完成）。联系人：宋长青，13515486669

联盟网址：<http://www.nydata.com.cn>



(3) 建立了农业大数据田间采集定位站



2014年3月15日，针对“渤海粮仓”国家战略，山东农大农业大数据田间定位采集站在山东省无棣县挂牌成立；并安装了自主开发的物联网数据采集系统；2016年8月渤海粮仓数据采集综合试验站建成投入使用。



2、

农业大数据研究应用阶段成果

- 建立了生物信息数据库
- 构建了小麦、玉米病虫害预测预警系统
- 建立了智慧水利预测预警系统
- 搭建了渤海粮仓大数据平台
- 搭建了农业信息服务平台
- 奶牛数字化精细养殖系统

涵盖农业、畜牧业、水利、设施园艺、农产品价格预测预警等37个大数据平台或系统。



2.

农业大数据研究应用阶段成果

共发表论文：187篇
其中 SCI: 56篇
EI: 74篇
专利120项
专著: 4部



World J Microbiol Biotechnol
DOI 10.1007/s11274-015-1883-5



ORIGINAL PAPER

Genomic-associated Markers and comparative Genome Maps of *Xanthomonas oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola*

Wenjie Feng^{1,4} · Yi Wang² · Lisha Huang² · Chuanshun Feng^{1,4} · Zhaojun Chu^{1,4} · Xinhua Ding^{1,4} · Long Yang^{1,3}

Received: 13 April 2015 / Accepted: 3 June 2015
© Springer Science+Business Media Dordrecht 2015

Abstract *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) and *X. oryzae* pv. *oryzicola* (*Xoc*) cause two major seed quarantine diseases in rice, bacterial blight and bacterial leaf streak, respectively. *Xoo* and *Xoc* share high similarity in genomic sequence, which results in hard differentiation of the two pathogens. Genomic-associated Markers and comparative Genome Maps database (GMGM) is an integrated database providing comprehensive information including compared genome maps and full genomic-coverage molecular markers of *Xoo* and *Xoc*. This database was established based on bioinformatic analysis of complete sequenced genomes of several *X. oryzae* pathovars of which the similarity of the genomes was up to 91.39 %. The program was designed with a series of specific PCR primers, including 286 pairs of *Xoo* dominant markers, 288 pairs of *Xoc* dominant markers, and 288 pairs of *Xoo* and *Xoc* co-dominant markers, which were predicted to distinguish two pathovars. Test on a total of 40 donor pathogen strains using randomly selected 120 pairs of primers demonstrated that

Wenjie Feng and Yi Wang have contributed equally to this work.

✉ Xinhua Ding
xhding@sdu.edu.cn

✉ Long Yang
lyang@sdu.edu.cn

¹ College of Plant Protection, Shandong Agricultural University, Tai'an 271018, China

² College of Agriculture, Shandong Agricultural University, Tai'an 271018, China

³ Agricultural Big-Data Research Center, Shandong Agricultural University, Tai'an 271018, China

⁴ State Key Laboratory of Crop Biology/Shandong Provincial Key Laboratory of Agricultural Microbiology, Shandong Agricultural University, Tai'an 271018, China

over 52.5 % of the primers were efficacious. The GMGM web portal (<http://biodb.sdu.edu.cn/gmgm>) will be a powerful tool that can present highly specific diagnostic markers, and it also provides information about comparative genome maps of the two pathogens for future evolution study.

Keywords Genomic-associated markers · Comparative genome maps · Rice pathogens · Database

Introduction

Xanthomonas oryzae is the major phytopathogenic bacteria of rice (*Oryza sativa*), including *X. oryzae* pv. *oryzae* (*Xoo*) and *X. oryzae* pv. *oryzicola* (*Xoc*), which are the causal agent of bacterial blight (BB) and bacterial leaf streak (BLS), respectively. BB is the most serious bacterial disease of rice which is distributed in both temperate and tropical regions, and can decrease rice yields by up to 50 % (Mansfield et al. 2012). By contrast, BLS is prevalent primarily in tropical regions which can reduce yield by 30 % (Kang et al. 2008). Both BB and BLS also have been considered as quarantine diseases in some countries, including China and United States (Niño-Liu et al. 2006). Both *Xoo* and *Xoc* are not only important for food security, but also are models for understanding bacterial interactions with plants. Rapid detection and accurate identification of pathogens are important for both plant protection and regulatory reasons, and incorrect diagnoses can result in large financial losses. The traditional methods are based on isolation of pathogen, such as biochemical identification, pathogenicity tests, or serological tests, which were laborious because of the close relation of *Xoo* and *Xoc* (Niño-Liu et al. 2006). In recent years, molecular-based methods based on repeated elements, siderophore receptor gene, the family gene,





3、

学术交流与合作

交流内容	次数
邀请国内外专家来校学术交流	200余次
交流合作内容	合作单位
战略合作，联合培养研究生，培训青年教师、 联合申报课题	复旦大学
战略合作，联合培养研究生，培训青年教师	台湾铭传大学
战略合作，数据综合采集试验站建设	无棣县人民政府
战略合作	上海大数据产业技术创新战略联盟
战略合作	北京中关村大数据产业技术创新战略联盟
战略合作	上海市数据科学重点实验室



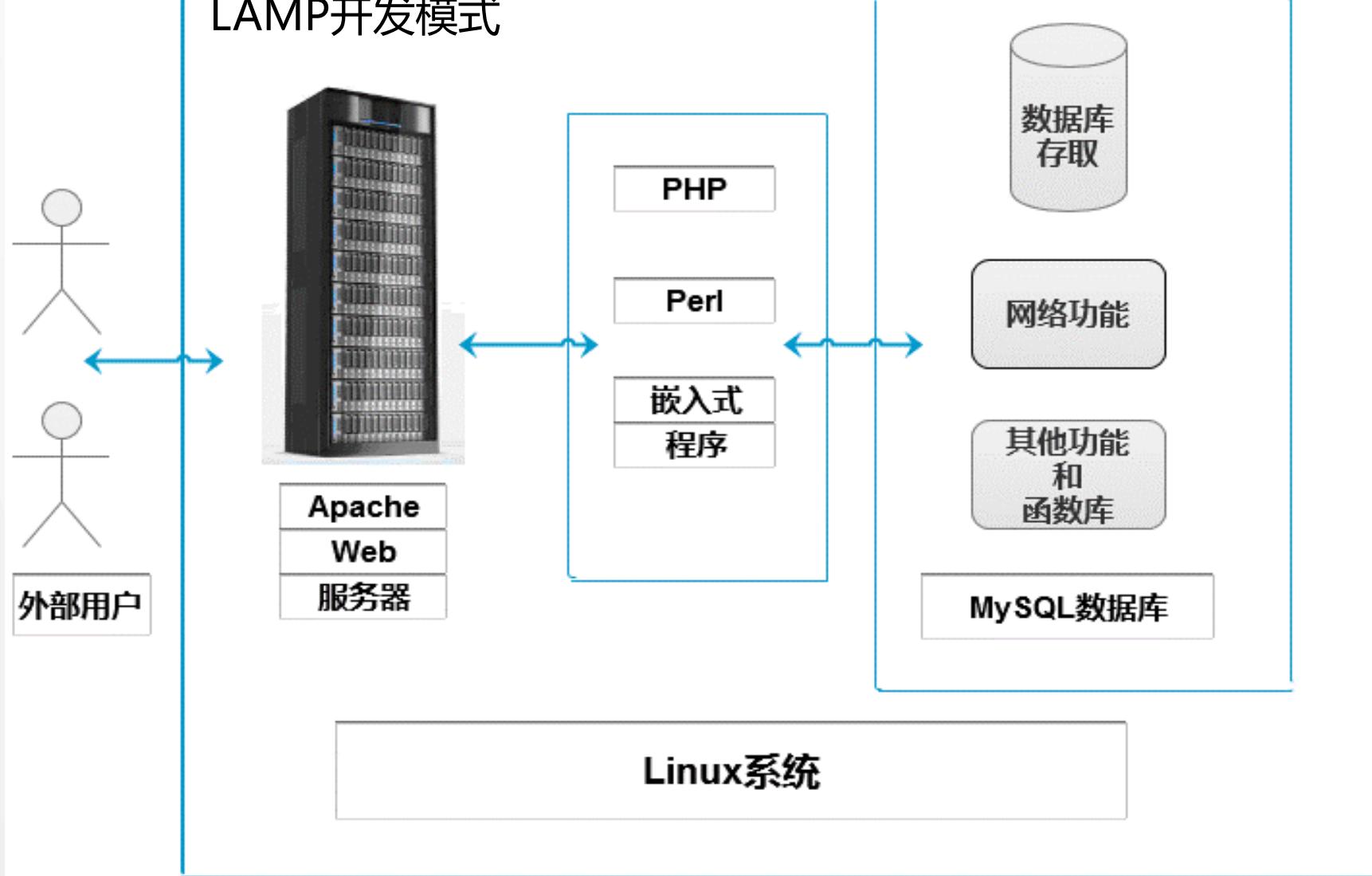
三、典型案例



案例1

生物数据库的构建与应用

LAMP开发模式





案例1

生物数据库的构建与应用

生物大数据基础

数据类型	物种/份	数据大小	描述
种质资源	1472	50G	烟草生物性状，照片信息
植物基因组	60	56.55G	植物基因组序列信息
植物转录组	15	60G	罗非鱼、玉米、小麦等RNAseq
表达谱	212	6.47G	EST、CDS等编码序列信息
病原序列	2346	5.6G	病毒，细菌，真菌，线虫，寄生植物
苹果重测序	100	2T	100个全球苹果重测序序列信息
.....



案例1

生物数据库的构建与应用

<http://biodb.sdau.edu.cn/database/index.html>

类型	数量
专业数据库	9
在线实用软件	7
涵盖物种	>259
种质资源	1,472
SSR标记	6,799,910
联合标记	3,559,291
SNP位点	20,602,973
选择性剪切	45,326,068

山东农业大学生物数据库
Biodatabase of SDAU

植物两种标记互作数据库

烟草遗传育种数据库

单核苷酸多态性扫描软件

水稻两种病原鉴定



案例1

生物数据库的构建与应用

生物数据库应用成果

- 已经定位和克隆了一些与高产、优质、高抗等主要农艺性状**相关的基因**；
- 构建了快速鉴定田间病害的**试剂盒**；利用数据库中的分子标记，**筛选了大量种质资源**，为种质资源的调查、收集、分类、保存和育种利用提供了物质基础和技术支持。



案例2

农作物虫害发生的大数据预测

基于SVR的玉米田四代棉铃虫发生量的预测模型构建

研究背景



近年来，随 Bt 棉大范围种植，第四代棉铃虫对玉米叶片和果穗，特别是果穗的危害逐渐加重。

而棉铃虫的发生具有非线性、不稳定、相关变量多的特点。将支持向量机回归(SVR)用于棉铃虫发生量的建模和预测分析，为防控服务。





案例2

农作物虫害发生的大数据预测

基于SVR的玉米田四代棉铃虫发生量的预测模型构建

SVR模型建立

输入

变量

气象变量 “**20-20时降水量**”、“**极大风速**”、“**极大风速的风向**”、“**平均本站气压**”等**16**个气象变量

输出变量

玉米田四代棉铃虫发生量



案例2

农作物虫害发生的大数据预测

基于SVR的玉米田四代棉铃虫发生量的预测模型构建

数据采集与标准化

20-20时降水	极大风速	极大风速的风	平均本站气压	平均风速	平均气温	平均水汽压	平均相对湿度	日照时数	日最低本站气	日最低气温
90.9	72.6	6.7	10034.4	23.2	25	275.5	85.9	44.7	10015.8	218.9
14.8	81.3	6.5	10077.8	20.6	262.	274.2	81	76.7	10049.2	215.6
19	71.1818182	5.54545455	10048.5455	23.2727273	26	282	83.8181818	60.7272727	10033.0909	224.272727
19.3	99.7	8.1	10049.4	25.9	265.	258.6	73.6	98.7	10028.5	216.9
9028.63636	86.8181818	7.81818182	10026.0909	24.0909091	271.27272	307.272727	86.4545455	42.4545455	10006.7273	232.636364
9848.5	66.9	10.7	10022.3	18.1	263.	291.7	85.8	56.3	10003.2	227.7
8985.63636	95.3636364	9.18181818	10050.9091	27.9090909	271.18181	285.181818	80.0909091	55.8181818	10024.5455	236.272727
54.8	78.3	7.5	10040.6	18.8	26	288.9	84.6	68.2	10020	227.4
93.4545455	69.3636364	7.36363636	10026	21.1818182	271.63636	285.636364	80.7272727	54.3636364	10009.6364	236.454545
6626	103.6	8.5	10015.2	23.8	256.	287.1	87.7	49.2	9993.3	226.5
3022.81818	81.5454545	8.90909091	10049.1818	25.5454545	262.45454	267.636364	79.2727273	56.9090909	10026.3636	221.909091
3771.29455	83.4072727	8.01181818	10040.6027	22.92	264.9945	282.822727	82.3063636	61.9372727	10019.4564	226.564545
5972.72727	63.1818182	8.90909091	10053.9091	20	265.	268.363636	77	53.8181818	10034.2727	224
45.4	78	9.2	10029	19.4	261.	276.1	82	60.1	10007.6	226.9
24.5454545	75.4545455	6.09090909	10041.2727	18.4545455	267.54545	278.181818	80.3636364	63.2727273	10024.2727	227.272727
8.7	60	9.5	10052.6	16.7	280.	285	76.5	60.4	10035.6	241.8
32.2727273	71.1818182	8.63636364	10028.3636	18.5454545	257.09090	285	84.4545455	48	10008.1818	220.545455
19.6	55.9	5.2	10054.6	13.4	258.	276.272727	87.1	31.3	10039.5	224.5
0.90909091	54.4545455	8.36363636	10039	17.1818182	296.90909	286.4	79.8181818	89.9090909	10018.5455	261.090909
3513.2	63.3	7.3	10049.7	18.5	26	328.181818	289	84	49.8	10017.1
										235.6

0.722905	0.0536131	0.281947	0.07116651	1	1	1	0.0950324	0.10097
1	1	1	1	0.358942	0.508028	0.257298	0.235421	0.71844
0.277654	0.785548	1	0.150404	0.38665	0.567116	0.319756	1	1
0.282123	0.62704	0.724138	0.10785	0.47733	0.39499	0.584521	0.606911	0.53203
0.141899	0.491841	0.736308	0.731475	0.177582	0.28966	0.095723	0.168467	0.07184
0.636872	1	0.926978	0.941306	0.629723	0.695568	0.57909	0.237581	0.27378
0.372067	0.044289	0.0791075	0.442406	0.0188917	0.0160565	0.000678887	0.511879	0.55145
0.75419	0.575758	0.46856	0.517241	0.7267	0.857418	0.665988	0.12527	0.17864
0.783799	0.347319	0.0141988	0.358767	0.463476	0.586384	0.399864	0.593952	0.53980
0.608939	0.002331	0.395538	1	0.770781	0.69043	0.835709	1	1
0.85419	0.00699301	0.156187	0.300073	0.148615	0.278099	0.00746775	0.231102	0.18252
0.228492	0.827506	0.69574	0.687454	0.0843829	0.0867052	0.0794297	0.185745	0.16116
1	0.230769	0.144016	0.665444	1	1	1	0.425486	0.34368



案例2

农作物虫害发生的大数据预测

基于SVR的玉米田四代棉铃虫发生量的预测模型构建

输出结果与检验

支持向量机回归拟合结果与实际值对比

年份	真实值	拟合值	绝对误差	相对误差(%)
1999	15.5	15.5998	- 0.0998	0.64
2000	49	48.8998	0.1002	0.20
2001	14	14.0998	- 0.0998	0.71
2002	24.5	24.3999	0.1001	0.41
2003	5	5.1001	- 0.1001	2.00
2004	19	18.5749	0.4251	2.24
2005	16	16.1003	- 0.1003	0.63
2006	39.3	37.6993	1.6007	4.07
2007	42.5	42.3998	0.1002	0.24
2008	47	47.1003	0.1003	0.21
2009	34.5	34.6003	- 0.1003	0.29
2010	38	38.0999	- 0.1	0.26

支持向量机回归预测结果与实际值对比

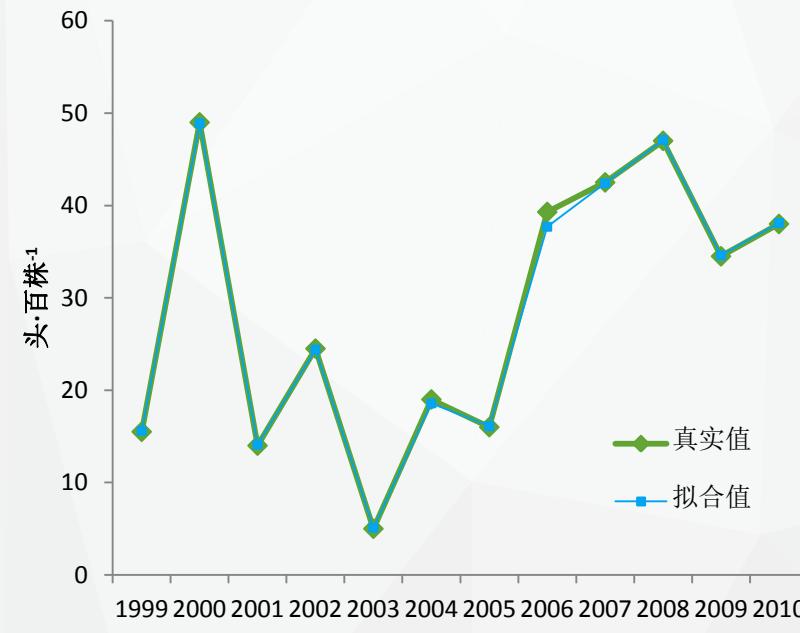
年份	真实值	预测值	绝对误差	相对误差(%)
2011	52	52.9923	- 0.9923	1.90
2012	42.6	41.2432	1.3568	3.17
2013	32.5	35.1017	- 2.6018	8.01



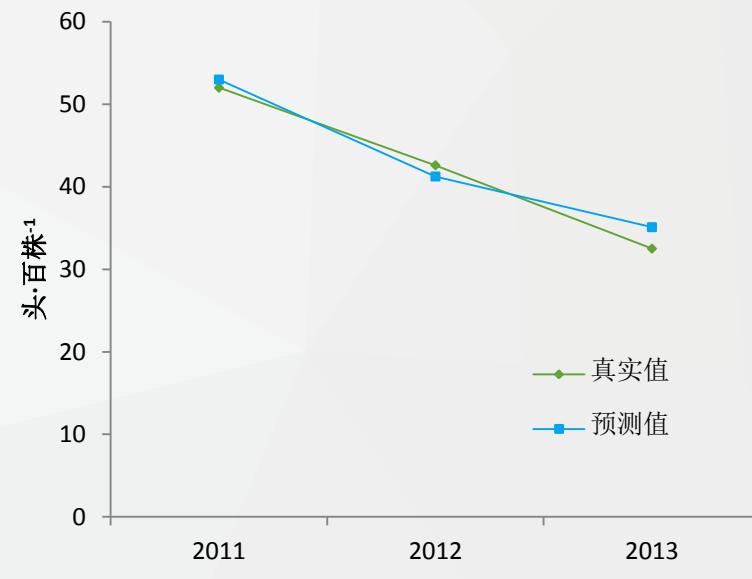
案例2

农作物虫害发生的大数据预测

基于SVR的玉米田四代棉铃虫发生量的预测模型构建



训练集拟合结果（12年的数据）



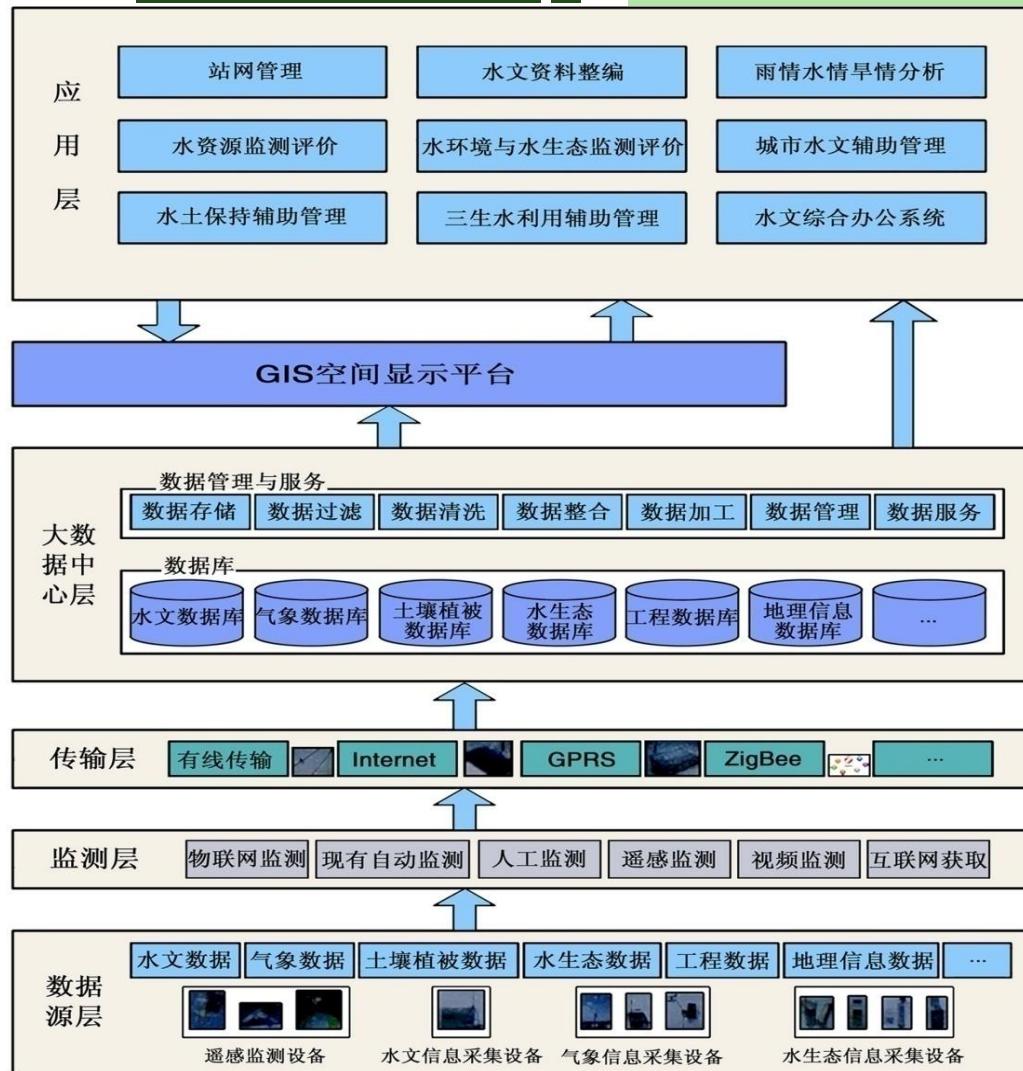
样本的预测结果（3年的数据）

通过回归模型得到的训练集样本的拟合值与实际值的相关系数为0.99，而测试集样本的预测值与实际值的相关系数为0.96。



案例3

大数据在水利水文中的应用

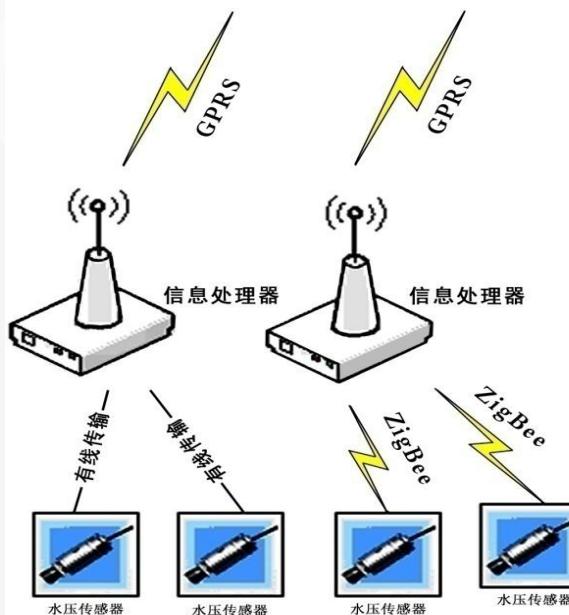


- ◆ 水资源短缺是未来的长期挑战
- ◆ 统筹安排水利设施建设
- ◆ 积极保护水资源及水环境
- ◆ 提升雨洪存蓄利用能力
- ◆ 增强应对干旱洪涝能力
- ◆ 科学全局规划水资源调配
- ◆ 保障社会健康持续发展

平台设计图

案例3

数据采集：历史数据和自动采集

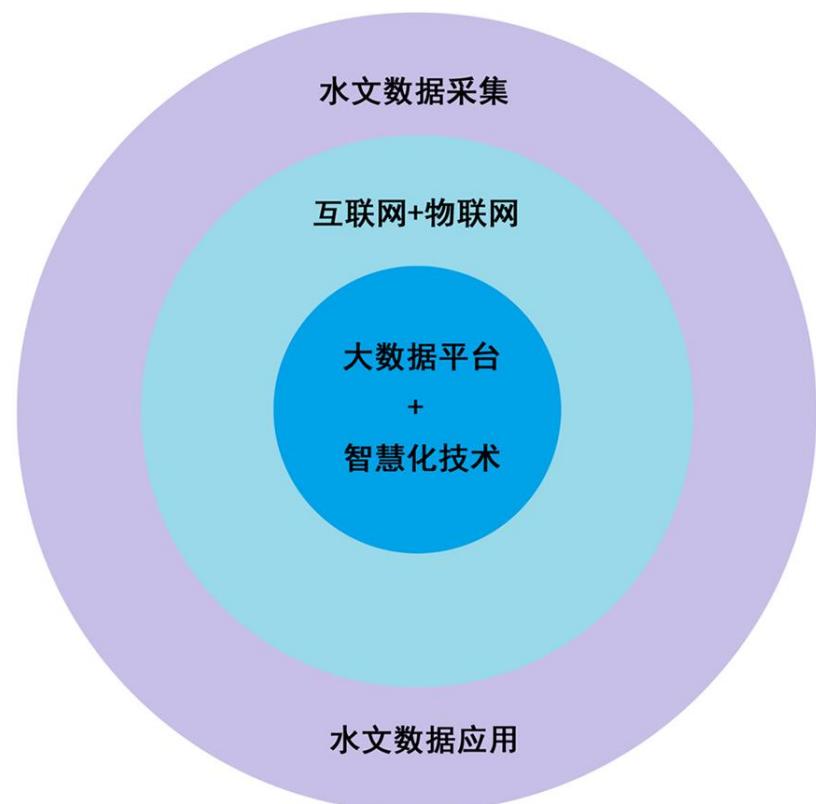


口预测水资源短缺状况，统筹安排水利设施建设，增强全局规划水资源能力，增强应对干旱洪涝灾害能力等



案例3

大数据在水利水文中的应用



◆ 成果

初步建立起山东水文
大数据应用平台系统。

- 1、为城镇化人口、工业、农业种植结构调整等提供服务；
- 2、为地方水利建设提供服务。

案例4



奶牛数字化精细养殖系统的建立与推广

李松和董事长将详细讲解（产学研结合项目）





四、展望

农业大数据产业的“春天”来了

国家战略：促进大数据发展行动纲要

农业部：推动农业农村大数据发展的实施意见

机遇：

我们将面临大数据驱动农业产业发展的良好机遇，在全国将形成立体的“农业大数据产业”

。



四、展望

建设泰山神农智谷

2015年11月，由我校倡导并参与的“泰山神农智谷”农业大数据产业园区落户泰安国家高新技术产业开发区。总投资35亿元，将分三期建设完成预计引入500家企业，实现年产值100亿元。

我校是国家农业农村信息化示范基地，以此为契机努力打造大数据技术创新和产业孵化基地，推动农业农村信息化与农业现代化的快速融合与发展。

And Then,
ARE YOU READY?
WHAT ARE YOU WAITING FOR?
NOW, ACTION!



欢迎感兴趣的企业入驻

“泰山神农智谷”农业大数据产业园在山东省泰安高新区奠基



谢谢！