

一种时变耦合模型在油井产量预测中的应用

金宝强¹, 童凯军¹, 孙红杰², 轩红彦³

1. 中海石油(中国)有限公司天津分公司渤海油田勘探开发研究院, 天津 300452
2. 中海油田服务股份有限公司天津分公司, 天津 300452
3. 中国石油天然气股份有限公司青海油田分公司采油一厂, 青海敦煌 816400

摘要 油井产量随时间的变化关系受多种因素的影响, 它们之间是一种极其复杂的非线性关系, 常规油藏工程研究方法往往受相关参数的不确定性所限, 预测误差偏大。分析认为油井产量数据具有时间序列特征, 引入相空间重构技术, 利用 G-P 法求出最佳嵌入维数, 对油井产量构成的时间序列进行混沌特性识别。在此基础上, 利用支持向量机(SVM)方法, 构建具有时变特性的混沌-SVM 耦合模型, 该模型对中、短期的油井产量预测具有很高的精度, 在实际应用中, 可以不断补充新的历史数据, 进行滚动预测, 可靠性更高。实例 W8-5 井的应用效果也表明, 预测结果平均相对误差仅为 7%, 显示出模型具有较好的预测效果和实用价值。

关键词 时间序列; 相空间; G-P 法; 混沌-SVM 模型; 油井产量

中图分类号 TE328

文献标识码 A

文章编号 1000-7857(2010)17-0072-05

Application of Time Varying Coupling Model in Prediction of Well Production

JIN Baoqiang¹, TONG Kaijun¹, SUN Hongjie², XUAN Hongyan³

1. Bohai Oilfield Exploration and Development Institute, Tianjin Branch, CNOOC Limited, Tianjin 300452, China
2. Tianjin Branch, China Oilfield Services Limited, Tianjin 300452, China
3. First Production Plant, Qinghai Oilfield Company, China National Petroleum Corporation, Dunhuang 816400, Qinghai Province, China

Abstract The variation of the oil well production against time is controlled by several factors, in an extremely complicated nonlinear manner. The conventional petroleum reservoir engineering method suffers often from uncertainty of the correlation parameters and a large prediction error. Our analysis shows that the oil well production data have some features of time series. Therefore, the technique of the phase space reconstruction and the G-P method can be used to obtain the optimal embed dimension and then to identify the time series of the well oil production. On this basis, by using the support vector machine method, the chaos-SVM model with time varying character is built, with a very high precision for short-term well productions. In a real application, we can supplement the new historical data in real time to make the rolling prediction. The example W8-5 well application indicates that the relative error of the forecast results is only 7%, which shows that the coupling model has a good forecast ability and is of practical value.

Keywords time series; phase space; G-P method; Chaos-SVM model; oil well production

非线性科学是一门研究非线性共性的基础科学, 主体包括混沌(chaos)、分形(fractal)、孤子(soliton)。非线性科学的研

究不仅具有重大的科学意义, 还具有广泛的应用前景^[1]。其中, 混沌指在确定性系统中出现的一种貌似无规则、类似随

收稿日期: 2010-03-30; 修回日期: 2010-06-23

作者简介: 金宝强, 工程师, 研究方向为油气田开发地质, 电子信箱: jinbq@cnooc.com.cn

机的现象,没有明显的周期和对称,却有丰富的内部层次有序结构,是非线性系统中的一种新的存在形式。1980年, Packard 等提出重构相空间理论,开始从时间序列角度研究混沌,该理论对于描述混乱无规则的复杂事物颇有成效,从而获得了广泛应用^[2]。

混沌与分形具有很深的内在联系,混沌被认为是产生分形结构的根源之一。分形理论在储集层孔隙结构、裂缝溶洞发育程度研究和成矿预测等方面得到了广泛的应用^[3]。油气井产量是油气开采的重要技术经济指标,研究表明,某些油气井的产量变化具有分形特征。本文采用相空间重构和关联维数提取技术,证明了油井产量变化具有混沌特征;在此基础上,利用混沌-支持向量机耦合模型,用单变量方法分析多变量耦合关系导致的油井单井产量时间序列,从而得到对油井单井产量演变规律更详细的认识。

1 油井产量变化的混沌特征检验原理

混沌系统最显著的特点是对初始条件非常敏感,系统初态的任何微小误差都将引起系统特征随时间急剧变化,但最终会收敛于相空间的一个有限区域,这一区域称为“吸引子”。混沌现象识别的基本出发点是识别相空间的吸引子是否为混沌吸引子^[4]。通过混沌系统的特性参数对其进行检验,比如通过关联维数、Lyapunov 指数、Kolmogrov 熵等检验具有时间序列特征的油井产量是否为混沌系统。本文主要通过分析关联维数来揭示油井单井产量变化的混沌特性。

1.1 重构相空间

混沌分析油井产量时间序列,首先要重构相空间。混沌系统的策动因素是相互影响的,因此在时间上先后产生的数据点也是相关的。Packard 等建议用原始系统中某变量的延迟坐标重构相空间, Takens 证明了可以找到一个合适的嵌入维数(延迟坐标的维数 $m \geq 2d$, d 为动力系统的维数)空间,把有规律的轨迹恢复出来^[5]。

对于油井单井产量构成的时间序列 (x_1, x_2, \dots, x_n) , 根据嵌入维数 m 和时间延迟 τ , 可重构出相空间, 相点为

$$y_i = (x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(m-1)\tau}) \quad (1)$$

y_i 为 m 维相空间矢量, 描述了油井产量构成的时间序列系统在 m 维相空间的演化轨迹, y_1, y_2, \dots, y_k 即为演化轨道。相空间重构技术关键在 m 和 τ 的确定, 二者均不宜过大或过小。

1.2 最佳时间延迟的确定方法

理论上讲, 数据点数无限时, 嵌入效果与 τ 无关, Takens 定理并没有对 τ 的要求。实际重建时, τ 对嵌入维的影响极大, τ 太小, 吸引子不能充分展开, 误差大; τ 太大, 不相关误差大且时间序列构建问题难以解决。获取时间序列的时间延迟比较成熟的计算方法: 自相关函数法和 C-C 法。其中, 时间序列的自相关函数为

$$\tau_T = \frac{\sum_{i=\tau+1}^n [(x_i - \bar{x})(x_{i-\tau} - \bar{x})]}{\sum_{i=1}^n [(x_i - \bar{x})^2]} \quad (2)$$

其中, $x_i (i=1, 2, \dots, n)$ 为变量时间序列, \bar{x} 为序列的均值, τ_T 为时间延迟为 τ 时的自相关系数。自相关系数值为初始值的 $1-1/e$ 时, 所对应的 τ 就是重构相空间的时间延迟。

1.3 嵌入维数的确定方法

实际应用中, 吸引子的维数未知, 从而无法确定吸引子的嵌入维数。1983年, Grassberger-Procaccia 提出了一种计算嵌入维数的方法——G-P 算法^[6]。定义关联积分函数 $C_m(r)$ 为

$$C_m(r) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i,j=1}^N H(r - \|Y_i - Y_j\|) \quad (3)$$

其中, Y_i, Y_j 为相空间中的相点; r 为给定的距离; H 为 Heaviside 函数

$$H(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (4)$$

给定 r 后, $C_m(r)$ 表示两点间距离小于 r 的概率。在适当范围内, 当 r 增加时, $C_m(r)$ 将以 r^{D_m} 的比率增加, 即

$$C_m(r) \propto r^{D_m} \quad (5)$$

两端同时取对数, 有

$$\ln C_m(r) = D_m \ln r + a \quad (6)$$

其中, a 为待定系数, D_m 为 $\ln C_m(r)$ 与 $\ln r$ 曲线上直线段的斜率。 m 增大到 m_b 后, D_m 的变化趋势急剧减缓, 在 $\ln C_m(r)$ 和 $\ln r$ 曲线上出现一簇近似平行的直线段。 m_b 为饱和嵌入维, 其对应的 D_{m_b} 就是实际系统的吸引子维数。 m_b 是否存在决定了系统的性质, m_b 存在, 表明系统的吸引子存在, 时间序列具有混沌特征; 反之, 时间序列为随机系统。

2 基于相空间的时变耦合模型构建

20世纪90年代, 非线性方法开始在石油勘探开发中应用。其中, 支持向量机(Support Vector Machine, SVM)能较好地解决小样本、非线性、高维数和局部极小点等实际问题, 被广泛应用于地震、测井、油田开发等实际问题的研究中^[7]。SVM与相空间重构的基本思想有相同之处, 都是为了提取系统包含的信息及规律, 把输入向量扩展到高维空间。相空间重构只能扩展到有限维, 而SVM可以通过核函数自动将输入空间的数据映射到一个高维甚至无穷维特征的空间 F 中, 使得数据蕴含的信息自然地呈现出来(图1)。本文将混沌时间序列重构的相空间点作为SVM的输入参数, 时间序列点作为输出变量。从而将二者的耦合模型引入油气田油井产量预测中, 将随时间变化的产量数据利用混沌时间序列处理成一个相空间中的序列点, 基于混沌时间序列的支持向量机耦合模型构建流程如图2所示。

用重构的样本集对SVM进行训练, 得到SVM的任意时刻的预测模型^[8]

$$\hat{x}_{t+1} = \sum_{i=1}^{n-(m-1)\tau} (\alpha_i - \alpha_i^*) K(x_t + x_i) + b \quad (7)$$

其中, α_i 和 α_i^* 为拉格朗日乘子, 互为对偶; K 为核函数; b 为偏

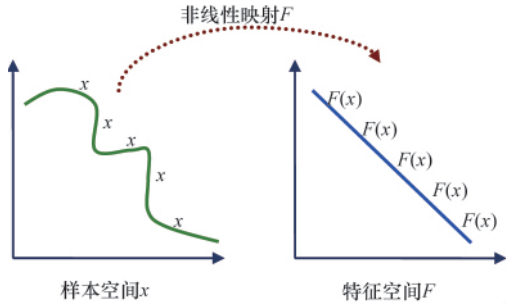


图 1 非线性映射关系示意图

Fig. 1 Nonlinear relationship diagram

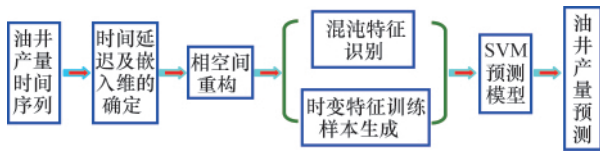


图 2 时变耦合模型预测油井产量流程

Fig. 2 The flow chart of time-varying coupled model to forecast oil production

差; $x_i = (x_{i\tau}, x_{i+2\tau}, x_{i+4\tau}, \dots, x_{i+(m-1)\tau})$ 。对于相空间的第 $t+1$ 点有 x_{i+1}

$l = (\hat{x}_{t+1}, x_{t+2\tau}, x_{t+4\tau}, \dots, x_{t+(m-2)\tau})$ 。由式(7)可得到对 $t+2$ 点预测为

$$\hat{x}_{t+2} = \sum_{i=1}^{n-(m-1)\tau} (\alpha_t - \alpha_i^*) K(x_t + x_{i+1}) + b \quad (8)$$

依次可得第 p 步的 SVM-混沌序列的预测模型为

$$\hat{x}_{t+p} = \sum_{i=1}^{n-(m-1)\tau} (\alpha_t - \alpha_i^*) K(x_t + x_{i+p}) + b \quad (9)$$

3 应用实例

锦州 9-3 油田是位于渤海辽东湾辽西凹陷北段的一个中型油田。自 1999 年 12 月油田主体区全面投入生产以来,油田主体区总体开发形势较好。生产动态资料表明油层连通性较好,平均单井日产液 $129\text{m}^3/\text{d}$,日产油 $40\text{m}^3/\text{d}$ 。目前,油田已进入中高含水阶段,综合含水 80%。以该油田 W8-5 井为例,采用具有时变特征的混沌-SVM 耦合模型分析预测了油井单井产量变化规律。

3.1 油井产量时间序列的相空间重构

对于 W8-5 井,采用自相关函数法确定了重构相空间的最佳时间延迟 τ 为 6(图 3)。

3.2 油井产量变化的混沌特征识别

确定了最佳时间延迟 $\tau=6$,取嵌入维数 $m=1,2, \dots, 15$ 。对于每个 m 值,使用 G-P 算法计算得到一系列 $\ln C_m(r)$ 、 $\ln r$ 值,绘制 $\ln C_m(r) - \ln r$ 关系图(图 4);选取线性相关程度高的区间作为无标度区间,在区间内绘制 $D_m - m$ 关系曲线(图 5)。由图 4、图 5 可得,随着相空间嵌入维数 m 的增加, D_m 值增大;当 m 增大到 9 时, D_m 值增幅变缓,趋于一定值 1.0。说明 W8-5 井的油井产量时间序列具有混沌性质, $m=9$ 可满足计算要求,可选择维数 $m=9$ 作为最佳相空间嵌入维数。

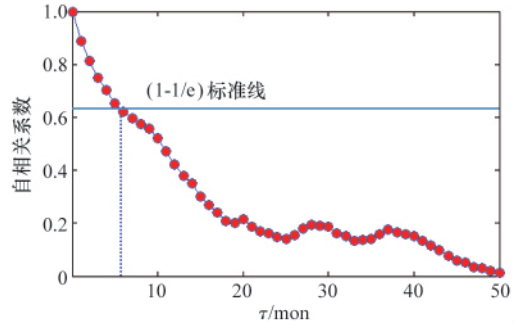


图 3 τ 与自相关系数的关系曲线

Fig. 3 Diagram for the time delay and the correlation coefficient

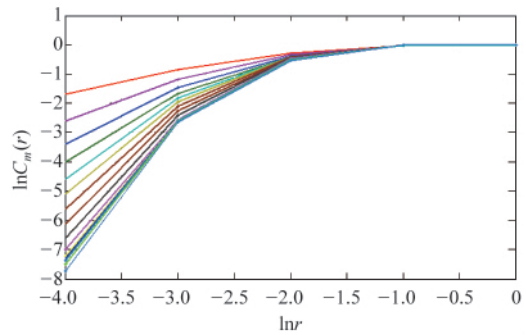


图 4 油井产量 $\ln C_m(r) - \ln r$ 关系曲线

Fig. 4 Diagram of $\ln C_m(r)$ and $\ln r$ for oil production

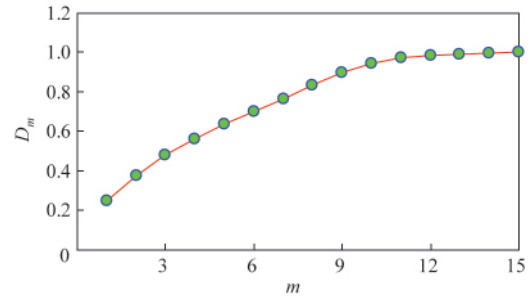


图 5 油井产量 $D_m - m$ 关系曲线

Fig. 5 Diagram of D_m and m for oil production

3.3 具有时变特征的 SVM 耦合模型的构建

1) 油井产量时间序列样本集的构建。确定了时间延迟、嵌入维数后,整理 184 个数据点(以 2002 年 1 月至 2009 年 10 月的半月累产数据为分析数据点),前 150 个数据点为训练样本点,后 34 个为检验样本点。由重构相空间嵌入相点构成训练样本集,得到用于 SVM 的学习样本为

$$X = \begin{bmatrix} x_1 & x_7 & \dots & x_{49} \\ x_2 & x_8 & \dots & x_{50} \\ \vdots & \vdots & \ddots & \vdots \\ x_{135} & x_{141} & \dots & x_{183} \end{bmatrix} \quad Y = \begin{bmatrix} x_{50} \\ x_{51} \\ \vdots \\ x_{184} \end{bmatrix} \quad (10)$$

为了避免 SVM 核函数在计算内积时计算困难,需要对产量数据进行归一化预处理,使其处于[0,1]之间。

$$X = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (X \in (0,1)) \quad (11)$$

其中, x, X 分别为归一化前后的变量值, x_{\min}, x_{\max} 分别为变量极小值、极大值。

2) RBF 核函数选择和 SVM 参数优化。SVM 预测模型中核函数的形式很多,本文采用能够很好解决复杂非线性问题的径向基 RBF 核函数。

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (12)$$

得到 SVM 的估计函数,形式为

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) + b \quad (13)$$

SVM 模型的参数控制主要核函数 RBF 的参数基宽 γ 、损失函数 e -insensitive 参数 p 及惩罚参数 C 。这 3 个关键参数的确定方法简述如下:首先不做任何优化,由系统自动计算出一个大概初值。以均方误差(Mean Squared Error, MSE)值大小为选择标准,首先保持 γ, p 值不变,由于 C 经常变化范围较大,以指数形式搜寻最优的 C 值;由图 6(a)可得,随着 C 值增大, MSE 值逐渐减小;当 C 增大到 1 之后, MSE 值变化微小;由此可确定 C 值为 1。保持 C, p 值不变,以线性增长搜寻

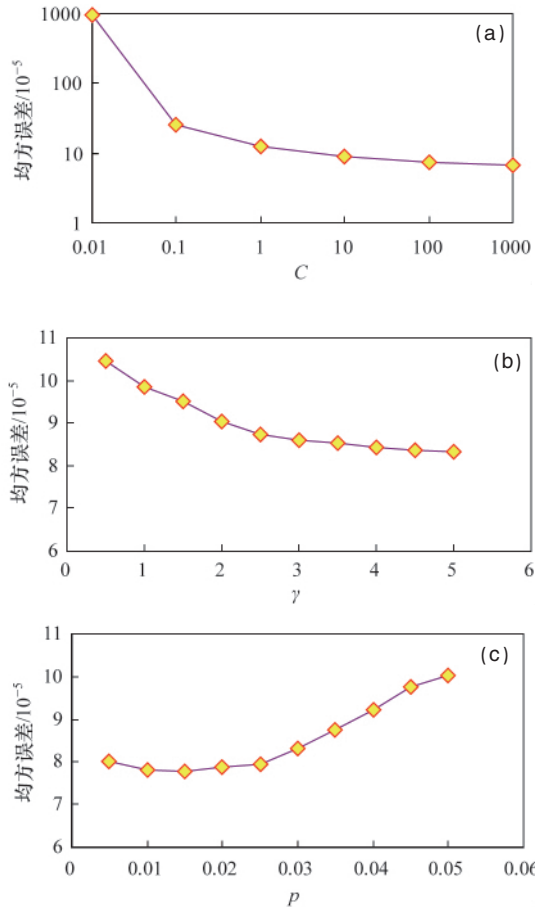


图 6 SVM 模型参数随 MSE 变化曲线

Fig. 6 Diagram for model parameters of SVM and MES

最优的 γ 值;由图 6(b)可知,随着 γ 值增大, MSE 值逐渐减小;当 γ 增大到 3 之后, MSE 变化微小;由此可确定 γ 值为 3。保持 C, γ 值不变,以线性增长搜寻最优的 p 值;由图 6(c)可得,随着 p 值增大, MSE 值先减小后增大;在 $p=0.015$ 处 MSE 取得极小值;由此可确定 p 为 0.015。最后得到最优参数组合 ($C=1, \gamma=3, p=0.015$)。

将上述最优参数组合代入 RBF 核函数,求解一个二次规划问题得出 α_i, α_i^*, b 之后,即可得到需要的 SVM 模型。

3.4 模型应用结果

利用构建好的具有时变特性的 SVM 模型对 W8-5 井单井产油量的历史拟合及预测,结果如图 7 所示。数据总体上可分为 3 部分:① 历史拟合部分 (2004 年 1 月 31 日—2008 年 6 月 30 日),由相空间重构技术构建出的训练样本经过参数优化之后对自身数据进行历史拟合,结果令人比较满意。② 验证样本部分 (2008 年 6 月 30 日—2009 年 10 月 31 日),利用训练好的内核对该部分的检验样本进行预测,预测结果相对误差如图 8 所示。相对误差处于 0~10% 之间,平均值为 7%,主要受产油量时间序列混沌性、构造状态空间的维数 m 、序列长度 N 和预测长度 L 等影响。结果表明该误差在允许范围之内,能够满足生产现场的实际需求。③ 预测样本部分 (2009 年 10 月 31 日—2011 年 11 月 30 日),这部分预测结果不久将会得到证实。利用具有时变特性的 SVM 模型对油气井产量进行预测,结果可靠,具有一定的合理性,同时也为生产现场提供了一种有别于常规油藏工程方法的新途径。

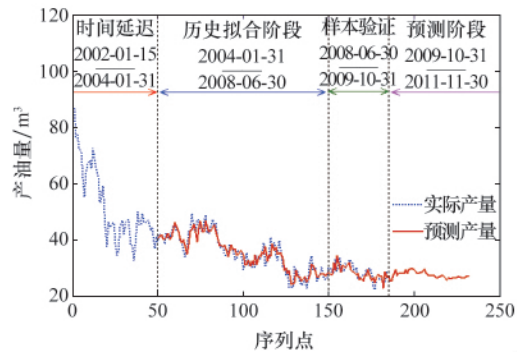


图 7 W8-5 井单井产油量的历史拟合和预测

Fig. 7 Oil production history matching and prediction of single well W8-5

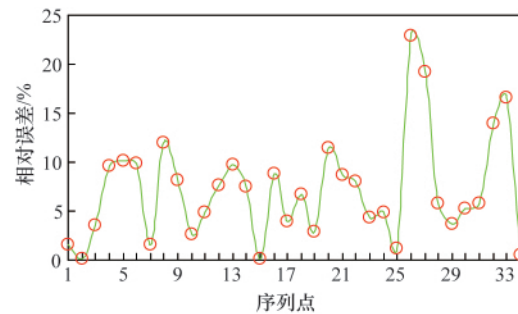


图 8 样本验证点相对误差关系

Fig. 8 Validation relative error of the sample points

4 结论

1) 单井产油量构成的时间序列具有混沌特征,说明产量随时间的变化关系为一个非线性混沌动力学系统。

2) 在结构风险最小化原则的基础上,由混沌理论及SVM技术构建的耦合模型考虑了训练样本的拟合性和复杂性,检验样本的相对误差处于0~10%,模型预测结果能够满足现场实际生产需求。

3) 矿场实际生产中油井生产动态是一个受多因素影响的复杂变量,本文提出的方法对无重大措施的油井生产动态预测效果较好。

参考文献 (References)

- [1] 吕金虎, 陆君安, 陈士华. 混沌时间序列分析及其应用[M]. 武汉: 武汉大学出版社, 2002.
Lu Jinhua, Lu Junan, Chen Shihua, et al. Wuhan: Wuhan University Press, 2002.
- [2] 程健, 白静宜, 钱建生, 等. 基于混沌时间序列的煤矿瓦斯浓度短期预测[J]. 中国矿业大学学报, 2008, 37(2): 231-235.
Cheng Jian, Bai Jingyi, Qian Jiansheng, et al. Journal of China University of Mining & Technology, 2008, 37(2): 231-235.
- [3] 陈志海, 戴勇, 郎兆新. 缝洞性油藏油井产量变化分形特征及其意义[J]. 石油勘探与开发, 2006, 33(1): 95-98.
Chen Zhihai, Dai Yong, Lang Zhaoxin. Petroleum Exploration and Development, 2006, 33(1): 95-98.
- [4] 李小刚, 杨兆中, 伍晓妮, 等. 油井水力压裂后产量变化的混沌特征分析[J]. 断块油气田, 2009, 16(1): 63-65.
Li Xiaogang, Yang Zhaozhong, Wu Xiaoni, et al. Fault-Block Oil & Gas Field, 2009, 16(1): 63-65.
- [5] 陈玉华, 杨永国, 彭高辉. 矿井涌水量混沌时间序列分析与预测[J]. 煤田地质与勘探, 2008, 36(4): 34-36.
Chen Yuhua, Yang Yongguo, Peng Gaohui. Coal Geology & Exploration, 2008, 36(4): 34-36.
- [6] 杨继鏊, 召利民, 杨绍清, 等. 基于混沌时间序列法的平潭浪高预测研究[J]. 海洋预报, 2008, 25(4): 63-69.
Yang Jili, Shao Limin, Yang Shaoqing, et al. Marine Forecasts, 2008, 25(4): 63-69.
- [7] 赵金省, 李兆敏, 李宾飞, 等. 基于支持向量机的蒸汽吞吐井调剖参数优化[J]. 西南石油大学学报, 2007, 29(5): 53-56.
Zhao Jinsheng, Li Zhaomin, Li Binfei, et al. Journal of Southwest Petroleum University, 2007, 29(5): 53-56.
- [8] 童凯军, 单钰铭, 李海鹏. 支持向量回归机在气井产能预测评价中的应用[J]. 新疆石油地质, 2008, 29(3): 382-384.
Tong Kaijun, Shan Yuming, Li Haipeng. Xinjiang Petroleum Geology, 2008, 29(3): 382-384.

(责任编辑 刘志远)

·学术动态·



“2011 农业工程新技术国际学术会议”征文

中国农业工程学会将于2011年5月27日在山东省淄博市召开“2011农业工程新技术国际会议”。

征文内容:

1) 农业工程新技术、新进展;保护性耕作技术及机具研究;变量施肥及精确喷药技术;复式中耕作业及除草机械;联合收获机减振、优化、可靠性及智能化设计技术;经济类作物联合收获技术与机具;林果采集、秸秆收集及高效生态化利用技术;设施农业控制、高产新技术;种子清选、加工、干燥、贮藏新技术;生物系统虚拟建模与可视化仿真研究;农业装备虚拟样机、虚拟试验及可靠性设计方法;生物芯片、生物传感器和生物仪器的智能设计;精准农业技术、装备与仪器;大中型拖拉机及配套机具的设计、制造及运用;汽车、发动机和农业装备中的节能及智能计算。

2) 支撑低碳农业及可持续发展的农业工程新技术;土壤有机碳库组分、有机碳的稳定性与农田生产力;循环型农业工程技术;节水灌溉及农业节能减排新技术;提高农业装备作业效率的技术、措施及模式;生态农业、绿色农业、有机农业、观光农业等新模式。

3) 太阳能、风能、生物质能利用新技术;太阳能建筑、温室及光伏发电技术;风光互补发电系统及风能独立发电系统;大型风力发电机设计、控制及并网技术;生物质能利用新技术(秸秆、能源作物、林木、海藻等生物质能源转化应用原理与技术)。

4) 食品及农产品品质分析及质量安全检测控制新技术;色谱技术在食品、农产品检测分析中的应用;质谱技术在食品、农产品检测分析中的应用;光谱技术在食品、农产品检测分析中的应用;食品、农产品的生物检测方法;食品、农产品快速检测技术及可追溯系统研究与应用;食品、农产品加工新技术。

征文截止时间:2011年2月20日。

联系方式:山东省淄博市张周路12号山东理工大学农业工程与食品科学学院(255049);传真:0533-2786896;电子信箱:aent2011@gmail.com。

会议网址:http://icae2011.sdut.edu.cn。