

基于灰色-马尔柯夫链预测模型的 耕地需求量预测研究

刘耀林¹ 刘艳芳¹ 张玉梅²

(1 武汉大学资源与环境科学学院, 武汉市珞喻路 129 号, 430079)

(2 武汉市国土规划局, 汉口三阳路 1 号, 430010)

摘要: 讨论了耕地总量预测的方法, 在分析现有耕地预测方法的基础上, 针对灰色 GM(1, 1) 预测与马尔柯夫链预测的优点和不足, 提出了基于灰色-马尔柯夫链的耕地需求量预测模型, 并以湖北省耕地需求量为例, 对该模型进行了实例验证和应用。

关键词: 灰色系统; 马尔柯夫链; 耕地总量; 预测; 模型
中图法分类号: P273

耕地总量预测是土地利用规划的必要前提。早期的耕地总量预测研究是建立在需求决定性基础上的, 通过人口对粮食等作物的需求来控制耕地总量。随着耕地的减少和人地矛盾的加剧, 根据耕地保护和农业持续发展的要求, 耕地开发利用方式逐渐转变为土地供给制约及引导需求, 以保证耕地数量。目前, 基于这种思想一般采用线性回归模型、指数平滑法和模糊预测法等预测方法, 前两种方法是对时间序列的数据进行纯数学处理, 没有考虑到自然因素和社会因素对耕地总量影响的后效性; 后一种方法是将众多耕地的影响因素综合起来, 通过各因素对耕地总量的影响权重进行模糊预测, 这种方法的局限性在于影响耕地预测的因素和权重的确定较复杂, 同时预测本身对原始数据的精度要求较高。

灰色预测的优势在于短期预测, 缺点在于对长期预测和波动性较大数据列的拟合较差。马尔柯夫链预测的优势在于长期预测和对随机波动性较大数据列的预测问题, 但是, 马尔柯夫链预测对象要求具有平稳过程。为了提高预测精度, 把灰色 GM(1, 1) 预测与马尔柯夫链预测两者结合起来, 形成灰色-马尔柯夫链预测模型, 这样就能够用 GM(1, 1) 预测揭示耕地数据列的发展变化总趋势, 用马尔柯夫链预测来确定状态规律。马尔柯夫链预测方法对解决无后效性的预测有独到之

处, 尤其在进行耕地总量预测时, 两者的结合可以避免考虑其他多种影响因素, 同时又可以预测耕地总量减少率的不同状态, 具有较大的科学性和实用性。

1 灰色系统和马尔柯夫链原理

1.1 灰色系统 GM(1, 1) 模型

灰色系统理论提供了在贫信息情况下解决系统问题的途径, 一个贫信息系统或灰信息系统, 称为灰色系统。灰色系统将所有的随机变量都看作灰色变量, 将随机过程看作是灰色过程, 因而不需要大量的历史数据, 而只根据“近代”的一些数据就可以进行预测。这种方法对于规律性不很明显、影响因素多的数列进行预测具有明显的优势。灰色系统预测模型就是通过时序数据累加生成的模块建立起来的, 滤去原始序列中可能混入的随机量, 从上下波动的时间序列中寻找某种隐含的规律性, 得到随机性弱化而规律性强化了的新数列, 挖掘出原始序列的内在特征, 这从 GM(1, 1) 预测模型建立的过程可以看出。设 $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$ 为一组时间序列, 这些数据可能是无规律的, 但经过累加以后, 弱化了数据的随机性和波动性, 增加了信息的白化度而呈现出一定的规律。

1.2 马尔柯夫链

若随机过程 $\{x(t), t \in T\}$ 对任意有限时序 $t_1 < t_2 < \dots < t_n \in T, (x(t_1), x(t_2), \dots, x(t_n))$ 相应的状态 $a_1, a_2, \dots, a_n \in A$ (状态空间) 有:

$$p\{x(t_n) \leq a_n \mid x(t_{n-1}), \dots, x(t_1)\} = p\{x(t_n) \leq a_n \mid x(t_{n-1})\}$$

该过程称为具有马尔柯夫性(或无后效性), 具有此性质的 $x(t)$ 称为马尔柯夫过程(马氏过程)。

假定马氏过程 $\{x_n, n \in T\}$ 的参数集 T 是离散的时间序列, 即 $T = \{0, 1, \dots, n\}$, 则相应 x_n 可能取值的全体组成的状态空间是离散的状态集 $A = \{a_1, a_2, \dots, a_n\}$ 。

设有随机过程 $\{x_n, n \in T\}$, 若对于任意的 $n \in T$ 和任意 $a_1, a_2, \dots, a_n \in A$, 条件概率满足:

$$p\{x_n = a_n \mid x_1 = a_1, x_2 = a_2, \dots, x_{n-1} = a_{n-1}\} = p\{x_n = a_n \mid x_{n-1} = a_{n-1}\}$$

则称 $\{x_n, n \in T\}$ 为马尔柯夫链(马氏链)。马氏链是马氏过程的一种特殊情况, 它表明事物的状态由过去转变到现在, 由现在转变到将来, 且无后效性。马尔柯夫链的统计特性决定了条件概率 $p\{x_n = a_n \mid x_{n-1} = a_{n-1}\}$ 。

在条件概率 $p\{x_n = j \mid x_{n-1} = i\}$ 中, $x_n = j$ 表示在 n 时刻系统(或过程)处于状态 j , 故条件概率 $p\{x_n = j \mid x_{n-1} = i\}$ 表示系统在时刻 $n-1$ 处于状态 i 条件下, 时刻 n 系统转移到状态 j 的概率。它相当于随机移动的质点在时刻 $n-1$ 处于状态 i 条件下, 经过一步随机移动转移到状态 j 的概率, 记此条件概率为 $p_{ij}(n)$:

$$p_{ij}(n) = p\{x_n = j \mid x_{n-1} = i\}$$

一般地, 转移概率 $p_{ij}(n)$ 不仅与 i, j 有关, 也与 n 有关。由系统中各种可能状态的转移概率 p_{ij} 构成的矩阵为马氏链的一步转移概率矩阵。其特点是矩阵中的每个元素 p_{ij} 均为非负, 且每行元素之和为 1, 即 $\sum_j p_{ij} = 1$ 。

k 步转移概率由 C-K 方程 $P_{ij}(k) = \sum_{l \in I} P_{il}(k-1) \cdot P_{lj}(k-1)$ 确定, k 步转移概率矩阵为:

$$P^k = \begin{bmatrix} P_{11}(k) & \dots & P_{1N}(k) \\ \vdots & & \vdots \\ P_{N1}(k) & \dots & P_{NN}(k) \end{bmatrix}$$

反复用 C-K 方程, k 步转移概率可以通过一步转移概率 P_{ij} 得到:

$$P(k) = P^k$$

2 灰色-马尔柯夫链预测模型

灰色系统预测主要用于时间短、数据资料少、

波动不大的预测问题, 其预测的几何图形是一条单调递增或单调递减的平滑曲线。进行长期预测时, 预测值常偏高或偏低, 因而对随机波动性较大的数据序列拟合较差, 预测精度较低。马尔柯夫链预测的对象是一个随机变化的动态系统, 具有以下特点: 一个 n 阶马尔柯夫链由 n 个状态的集合 $\{E_1, E_2, \dots, E_n\}$ 和一组转移概率 $p_{ij} (i, j = 1, 2, \dots, n)$ 所确定, 该过程在任一时刻只能处于一个状态, 如果在时刻 k 过程处在 E_i , 则在时刻 $k+1$, 它将以概率 p_{ij} 处于状态 E_j 。马尔柯夫链的上述特点决定了其预测是根据状态之间的转移概率来推测系统未来发展变化的。转移概率 p_{ij} 反映了各种随机因素的影响程度, 因而马尔柯夫链适合于随机波动性较大的数据预测问题, 在这一点上恰好可以弥补灰色预测的局限。但是马尔柯夫链预测对象还要求具有平稳过程等均值, 然而耕地总量的预测问题是随时间变化而呈现某种变化趋势的非平稳随机过程。采用灰色 GM(1, 1) 模型对耕地总量的时序数据进行拟合, 找出其变化趋势, 则可以弥补马尔柯夫链预测的局限性。

应用灰色-马尔柯夫链方法进行预测的基本思路是, 先建立灰色 GM(1, 1) 模型, 求出其预测曲线 $\hat{Y}(k)$; 再以平滑的预测曲线 $\hat{Y}(k)$ 为基准, 划分若干动态的状态区间; 根据落入各状态区间的点计算出马尔柯夫转移概率矩阵, 预测未来状态, 从而得出预测值区间; 取区间中点, 最终得到精度较高的预测值。

2.1 建立 GM(1, 1) 模型

根据灰色系统理论, GM(1, 1) 模型的 $\hat{Y}(k)$ 曲线为:

$$\hat{Y}(k) = \hat{x}^{(0)}(k+1)$$

2.2 状态划分

以 $\hat{Y}(k)$ 曲线为基准, 根据每个预测对象的具体情况, 划分成与 $\hat{Y}(k)$ 曲线平行的若干条形区域, 每个条形区域构成一个状态, 任一状态区间 Q_i 表达为:

$$Q_i = [Q_{1i}, Q_{2i}], i = 1, 2, \dots, n$$

式中, $Q_{1i} = \hat{Y}(k) + A_i; Q_{2i} = \hat{Y}(k) + B_i; A_i, B_i$ 可由用户根据具体情况自行确定。

由于 $\hat{Y}(k)$ 是时间 k 的函数, 因而 Q_{1i}, Q_{2i} 也随时序变化, 即状态 Q_i 具有动态性。

若假设划分为 3 个状态区间, 则可将原始数据列 $x^{(0)}(k)$ 、GM(1, 1) 预测曲线 $\hat{Y}(k)$ 及划分状态表示成图 1。

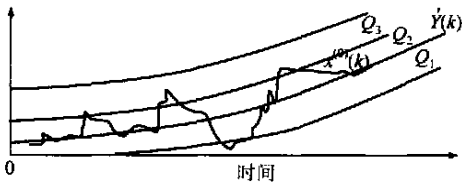


图 1 状态划分图

Fig. 1 State Division

2.3 计算转移概率矩阵

转移概率矩阵公式为:

$$P_{ij}^{(m)} = M_{ij}^{(m)} / M_i$$

式中, $P_{ij}^{(m)}$ 为由状态 Q_i 经过 m 步转移到 Q_j 的概率 ($i, j = 1, 2, \dots, n$); n 为划分的状态数目; M_i 为原始数据按一定的概率落入状态 Q_i 的样本数; $M_{ij}^{(m)}$ 为由状态 Q_i 经 m 步转移到 Q_j 的原始数据样本数。

当状态划分不太适合, 以致某一状态中无原始数据落入时, 则可令 $P_{ij}^{(m)} = P_{ji}^{(m)} = 0$ 。由此, 可求出状态转移概率矩阵 $P^{(m)}$ 来预测未来状态的转向:

$$P^{(m)} = \begin{bmatrix} P_{11}^{(m)} & \dots & P_{1N}^{(m)} \\ \vdots & & \vdots \\ P_{N1}^{(m)} & \dots & P_{NN}^{(m)} \end{bmatrix}$$

一般只需考察一步转移概率矩阵 $P^{(1)}$, 但当状态的将来转向难以确定时, 则需要考察多步转移概率矩阵 $P^{(m)}$ 。设预测对象基期处于 Q_k 状态, 考察 $P^{(1)}$, 则预测下一步状态有如下两种方法。

1) 伪随机数模拟法。考察 $P^{(1)}$ 中的第 k 行, 产生 $(0, 1)$ 均匀分布的伪随机数 r , 选取 d , 看是否满足不等式:

$$\sum_{j=1}^{d-1} P_{kj} \leq r \leq \sum_{j=1}^d P_{kj}$$

如果满足, 则下一时刻最有可能由 Q_k 状态转向 Q_d 状态。

2) 比较法。考察 $P^{(1)}$ 中的第 k 行, 若 $\max P_{kj} = P_{kd}$, 则下一时刻转移到 Q_d 状态。

2.4 确定预测值

当未来的转移状态 Q_d 确定后, 也就确定了预测值的变动区间 $[Q_{1d}, Q_{2d}]$, 预测值 $\hat{Y}(k)$ 取该区间的中点, 则有:

$$\hat{Y}(k) = \frac{A_i + B_i}{2}$$

2.5 模型精度检验

模型精度检验公式为:

$$q^{(0)}(k) = x^{(0)}(k) - \hat{x}^{(0)}(k+1)$$

$$\varepsilon(k) = q^{(0)}(k) / x^{(0)}(k)$$

式中, $q^{(0)}(k)$ 为残差; $\varepsilon(k)$ 为相对误差。用后验差检验, 方法如下。

1) 求 $x^{(0)}(k)$ 的平均值 \bar{x} :

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x^{(0)}(k)$$

2) 求 $x^{(0)}(k)$ 的方差 S_1 :

$$S_1^2 = \frac{1}{n} \sum_{k=1}^n (x^{(0)}(k) - \bar{x})^2$$

3) 求残差 $q^{(0)}(k)$ 的平均值 \bar{q} :

$$\bar{q} = \frac{1}{n} \sum_{k=1}^n q^{(0)}(k)$$

4) 求 $q^{(0)}(k)$ 的方差 S_2 :

$$S_2^2 = \frac{1}{n} \sum_{k=1}^n (q^{(0)}(k) - \bar{q})^2$$

5) 求后验差比值 C :

$$C = S_2 / S_1$$

6) 求小误差频率 P :

$$P = \{ |q^{(0)}(k) - \bar{q}| < 0.674 5 S_1 \}$$

一个好的预测, 要求 C 越小越好, 一般要求 $C < 0.35$, 最大不超过 0.65。预测好坏的另一个指标是小误差频率要大。所谓小误差是指绝对偏差 $|q^{(0)}(k) - \bar{q}| < 0.674 5 S_1$, 或者说相对误差 $(|q^{(0)}(k) - \bar{q}| / S_1) < 0.674 5$ 。一般要求 $P > 0.95$, 不得小于 0.7。按照 P 与 C 的大小, 可将预测精度分为 4 个等级, 见表 1。

表 1 预测精度等级

等级	P	C
好	> 0.95	< 0.35
合格	> 0.80	< 0.45
勉强	> 0.70	< 0.50
不合格	≤ 0.70	≥ 0.65

3 湖北省耕地总量的波动情况及现状

本文选取 1987 年到 1998 年的湖北省耕地总量作为研究数据, 具体见表 2。

3.1 湖北省耕地总量的预测

考虑到耕地总量逐年下降, 如果用它作为原始时间序列进行预测, 误差会较大, 而耕地减少率有大有小, 更能反映耕地的变化情况, 同时考虑到计算的简便性, 本文选取耕地减少率作为原始时间序列 $X^{(0)}$ 。

1) 输入原始数据。

2) 计算并绘制 $Y(k)$ 曲线, 如图 2。

表 2 1987~1998 年湖北省耕地总量/($10^3 \cdot \text{ha}$)

Tab. 2 Gross Arable Land in Hubei During 1987~1998/($10^3 \cdot \text{ha}$)

年份	耕地总量	耕地减少量	耕地减少率/(%)
1986	3 545.00	39.61	11.05
1987	3 517.99	27.01	7.62
1988	3 498.47	19.52	5.55
1989	3 486.57	11.90	3.40
1990	3 476.77	9.80	2.80
1991	3 458.46	18.31	5.27
1992	3 421.57	36.89	10.67
1993	3 392.74	28.83	8.43
1994	3 375.60	17.14	5.05
1995	3 358.01	17.59	5.21
1996	3 349.25	8.76	2.61
1997	3 342.45	6.80	2.03
1998	3 327.16	15.29	4.57

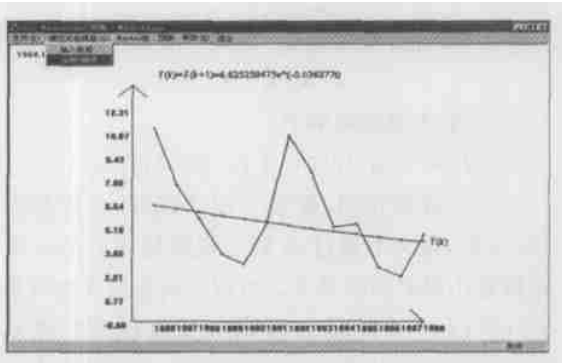


图 2 $Y(k)$ 曲线图

Fig. 2 $Y(k)$ Curve

3) 划分状态。根据原始数据序列图和 $Y(k)$ 曲线图, 按照实际情况划分状态, 在程序中设置了人工划分和计算机自动划分两种情况, 根据具体情况, 选择一种。本文选用人工划分状态, 划分为 Q_1 、 Q_2 、 Q_3 、 Q_4 四种状态, 见图 3。

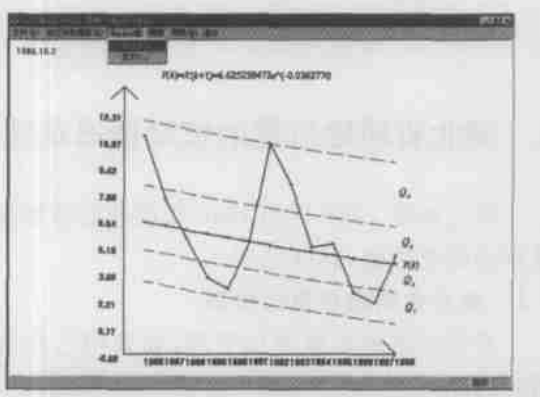


图 3 状态划分曲线图

Fig. 3 State Division

4) 确定转移概率矩阵。根据状态图, 求转移

概率矩阵, 因为最后一个点的转移状态未知, 因此, 只考虑前面 $n-1$ 个点的情况。现以 P_{ij} 的第一行为例, 具体说明 P_{ij} 的求法。 P_{ij} 的第一行即第一状态, 由图 4 可以看出, 共有 4 个点处于第一状态, 其中有两个点经过一步向后转移仍处于第一状态, 有一个点经过一步向后转移可到达第二状态, 有一个点经过一步向后转移可到达第三状态, 没有点能经过一步转移到达第四状态, 因此, 矩阵 P_{ij} 的第一行为 $2/4$ 、 $1/4$ 、 $1/4$ 、 0 。依此类推, 可求出矩阵 P_{ij} 的值。已求得的 P_{ij} 如图 4 所示。

5) 预测。下面预测 1999 年耕地减少率。已知 1998 年处于 Q_3 , 由矩阵第三行可知, 1998 年可向 Q_1 、 Q_2 或 Q_3 转移, 因此, 1999 年可处于状态 Q_1 、 Q_2 或 Q_3 , 根据前面介绍的伪随机数法或比较法可以得出, 1999 年处于 Q_3 状态, 进而可得 GM(1, 1) 的预测值和灰色-马尔柯夫模型的预测值。

6) 结果对比。将 1999 年实际耕地减少率与由灰色 GM(1, 1) 直接预测出的结果和由灰色-马尔柯夫链模型预测出的结果进行比较。

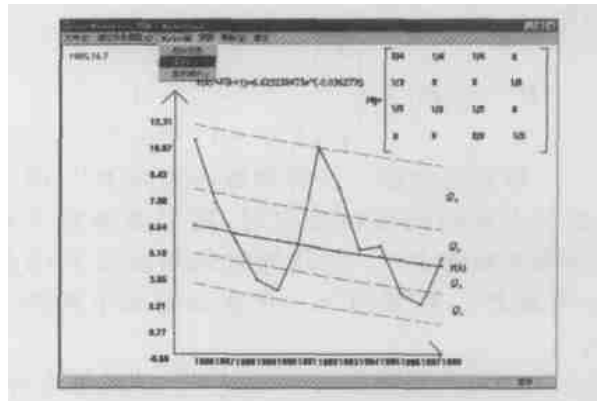


图 4 P_{ij} 显示图

Fig. 4 P_{ij} Display

由计算结果可知, 灰色-马尔柯夫链模型预测出来的值比 GM(1, 1) 的预测值更接近实际值。由灰色-马尔柯夫链模型预测出来的耕地减少率得 1999 年的耕地总量为 $3.310 08 \times 10^6 \text{ha}$, 与实际的耕地总量 $3.310 40 \times 10^6 \text{ha}$ 相比, 误差为 9.6×10^{-5} 。

下面由 1998 年的状态预测 2000 年的耕地减少率。从 1998 年到 2000 年要考察 P_{ij} 的二阶矩阵。同样采用伪随机数法或比较法, 得出 2000 年处于 Q_2 。取灰色-马尔柯夫链预测出来的值 3.236 877 作为 2000 年的耕地减少率, 可得 2000 年耕地总量为 $3.299 68 \times 10^3 \text{ha}$ 。

同理可预测其他未来年份的耕地减少率和耕

地总量。本文采用同样方法预测了 2005 年和 2010 年的耕地减少率和所处的状态, 见表 3。

表 3 未来几年耕地减少率和所处状态/(%)

Tab. 3 Deduction Rate of Arable Land in Future Years and Their States/(%)

年份	所处状态	GM(1, 1)预测值	灰色-马尔柯夫链预测值
2000	Q_2	3.986 877	3.236 877
2005	Q_3	3.448 365	4.448 365
2010	Q_1	2.876 330	0.526 330

7) 精度检验。模型精度检验最后的检验结果为 $P=1.0$, $C=0.30$ 。

由表 1 可得, 该灰色-马尔柯夫链预测模型精度为“好”。如果遇到精度不高的情况, 可在程序的该界面中点“重新划分状态”按钮, 重新回到第三步划分状态, 此模型的精度高低最主要在划分状态, 需要经过反复比较、划分, 才能达到最佳效果。

3.2 预测结果分析

对对近期和远期的年份预测可以看出, 耕地减少率将会是处于一个平稳状态, 耕地总量减少的速度逐渐减慢, 直至趋近于一个常数, 在直角坐标中表现为与 y 为常数的直线无限接近, 这个常数就是耕地总量达到平衡的数。

4 结 语

1) 应用灰色-马尔柯夫链预测方法来预测耕地总量, 比直接应用灰色系统预测更能反映耕地总量变化所处的状态, 能较精确地预测未来几年的耕地总量。

2) 灰色-马尔柯夫链预测法能充分利用时间序列数据给予的信息进行分析预测, 与其他常用的时间序列预测方法比较, 该方法计算简便, 精度较高。

3) 灰色-马尔柯夫链方法是灰色系统理论和马尔柯夫链方法的结合, 应用于耕地总量预测时, 可以避免自然环境、生态环境、政策导向等因素, 同时又考虑了耕地总量在不同年份下的状态和数量, 具有较大的科学性和实用性。

参 考 文 献

- 1 王万茂. 土地利用规划学. 北京: 中国大地出版社, 1996
- 2 张立才. 土地管理论文集. 北京: 中国大地出版社, 1997
- 3 郑人权. 预测学原理. 北京: 中国统计出版社, 1988
- 4 张有为. 预测的数学方法. 北京: 国防工业出版社, 1991
- 5 黄梦平, 林志扬. 管理预测学. 福州: 福建科学技术出版社, 1999
- 6 姜进树. 预测与规划. 上海: 上海科学技术出版社, 1989
- 7 冯文权. 经济预测与决策技术. 武汉: 武汉大学出版社, 1989
- 8 唐小毅. 经济预测方法. 成都: 电子科技大学出版社, 1989
- 9 巴鲁查-赖特 A T. 马尔柯夫过程论初步及其应用. 上海: 上海科学技术出版社, 1979
- 10 侯振挺, 郭青峰. 齐次马尔柯夫过程. 北京: 科学出版社, 1978
- 11 胡迪鹤. 随机过程概论. 武汉: 武汉大学出版社, 1986
- 12 邓聚龙. 灰色预测与决策. 武汉: 华中理工大学出版社, 1992
- 13 杭 斌. 利用回归模型估计马尔柯夫转移概率. 山西财经学院学报, 1992(2)
- 14 Liu S F, Deng J L. The Rang Suitable for GM(1, 1). Journal of Grey System, 1999, 11(1)
- 15 Zhu K L. The Calculating Pattern and User of the Grey Markov Chain Analysis. System Engineering-Theory & Practice, 1988
- 16 Duncan T E, Pasik D B, Stettner L. Adaptive Control of a Partially Observed Discrete Time Markov Process. Appl. Math. Optim., 1998, 37(3)
- 17 Bluementhal R M, Getoor R K. Markov Process and Potential Theory. New York: Academy Press, 1986
- 18 Graversen S E, Vuolle-Apiala J. α -self-similar Markov Process. Probability Theory Related Fields, 1986(71)

第一作者简介: 刘耀林, 教授, 博士, 博士生导师。主要从事地理信息分析和应用模型研究。

E-mail: yaolin610@163.com

(下转第 596 页)

Oil-Gas Reservoir Evaluation Based on GIS

SHI Dong^{1,2} CHEN Jun³ ZHU Qing¹

(1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2 Department of Geoscience, Yangtze University, 1 Nanhuan Road, Jingzhou 434023, China)

(3 National Geomatics Center of China, 1 Baishengcun, Zizhuyuan, Beijing 100044, China)

Abstract: Conventional reservoir evaluation method and its existing problems are analyzed through studying commercial GIS. The spatial database of oil-gas reservoir evaluation is established. All kinds of multidisciplinary oilfield graphic information and its correlating attribute data relating to reservoir are integrated. By means of integrating reservoir evaluation model with GIS, oil-gas reservoir evaluation problem is figured out. Thus the spatial analysis function of GIS is extended. An instance of a western oilfield in our country is cited to prove that the new method makes the evaluation more accurate and reliable.

Key words: GIS; oil-gas reservoir evaluation; spatial analysis; integration mode

About the first author: SHI Dong, associate professor, Ph. D. Ph. D candidate. She is engaged in the research on GIS and oil-gas reservoir. E mail: sdd129@263.net

(责任编辑: 涓涓)

(上接第 579 页)

Prediction of Gross Arable Land Based on Grey-Markov Model

LIU Yaolin¹ LIU Yanfang¹ ZHANG Yumei²

(1 School of Resource and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2 Wuhan Bureau of Land Management, 1 Sanyang Road, Wuhan 430010, China)

Abstract: This paper discusses the prediction methods of gross arable land based on grey-markov model. The advantages and disadvantages of grey GM (1, 1) prediction and Markov prediction model have been analyzed. A new method, which is called grey Markov model, is presented based on analyzing current methods for arable land prediction. This prediction model is identified by taking prediction of arable land in Hubei province.

Key words: grey system; Markov chain; gross arable land; prediction; model

About the first author: LIU Yaolin, professor, Ph. D. Ph. D supervisor. his researches include the analysis and application of geographical information modelling.

E mail: yaolin610@163.com

(责任编辑: 涓涓)