

信息熵在水文模型信息评价中的应用

王新民, 马长明, 刘波波, 杜鑫, 秦雪

(北京工业大学 建筑工程学院, 北京 100124)

摘要: 针对四川省安宁河流域安宁桥站集水区构建数字高程模型及水系划分, 分别以 MIKE11 的降雨径流模型 (RR 模型) 与新安江模型完成了流域日径流过程数值模拟, 选定确定性系数 DC 和相对径流深误差 BIAS 作为两个控制指标, 应用信息熵理论对两个模型的模拟结果进行比较分析。信息熵计算过程以观测日径流量为标准序列, 模拟计算的日径流量为观测序列计算残差, 进行归一化处理, 进而计算了潜在信息测度和信息熵相对测度。结果表明: 依据信息熵最大理论对水文模型的参数确定和模拟效果评价是有效的; 同时也须指出在对不同模型进行对比评价时, 信息熵的相对值也满足差异信息理论的非唯一性原理。

关键词: 日径流; 信息熵; RR 模型; 新安江模型; 水文模型

中图分类号: P33 文献标识码: A 文章编号: 1672-643X(2014)06-0056-05

Application of information entropy in information evaluation of hydrological model

WANG Xinmin, MA Changming, LIU Bobo, DU Xin, QIN Xue

(College of Architecture and Civil Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: The paper set up digital elevation watershed model and divided water system aimed at water collection area of Anningqiao hydrology station in Anning river basin, and simulated the daily runoff process of the watershed by means of two widely used rainfall-runoff software of MIKE11 and Xinanjiang hydrological model. It chose DC and BIAS as two control indexes and compared and analyzed the simulation results of RR model and Xinanjiang Model based on the theory of information entropy. Taking the day runoff as standard sequence and the computered day runoff as standard calculated residual, the paper normalized and calculated potential information measure and relative information entropy measure. The conclusions are that the maximum number of information entropy is practicable to determine the hydrological model's parameters and evaluate the simulating results. When carrying out comparative evaluation to difference models, the relative value of information entropy will also meet the principle of non-unique-ness of differential information theory.

Key words: day runoff; information entropy; RR model; xin'an river model; hydrological model

水文现象是由许多包含不确定性的因素相互作用而形成的复杂过程。流域水文模型是计算机科学和产汇流理论结合的产物, 是用数学手段描述和模拟水文现象的重要工具。

通过对同一流域适应性研究不同的模型, 找出模拟较好的模型, 以提高拟合效果和预测精度一直是水文科学工作者所关注、致力于研究解决的课题。基于概率密度函数而构建的信息熵理论, 提供了对流域水文模型的参数估计、不确定类型甄别与分析、

模拟效果评价等众多不确定性研究课题的一种思路^[1-3]。在有限的给定条件下, 在所有可行解中应该选择混乱程度最大, 其所包含的信息量最大, 也就是熵最大的一个, 这就是最大信息熵原理。

本文运用两种水文模型对四川省安宁河流域安宁桥站的日径流过程进行模拟计算, 利用信息熵理论, 基于归一化后的残差序列计算出信息熵, 进行水文模型信息评价, 给出了具体应用实例。

收稿日期: 2014-07-09; 修回日期: 2014-09-23

基金项目: 国家水体污染控制与治理科技重大专项 (2009ZX07104-006)

作者简介: 王新民 (1990-), 男, 陕西榆林人, 硕士研究生, 主要从事水文学及水资源方面的研究。

1 数字流域的建立

数字高程模型(DEM)是地理信息系统的基础数据,广泛应用在水文学研究和水资源管理中。本文选取四川省安宁河流域安宁桥站 $2\,316\text{ km}^2$ 的集水面积进行日径流过程模拟。所用DEM数据为位于地理位置北纬 $30.46^\circ \sim 31.52^\circ$,东经 $104.00^\circ \sim 105.18^\circ$ 的数据。该范围内设有水文站与蒸发站各1个,雨量站5个。

应用数字高程流域水系模型对凹陷区进行识别处理,按照斜坡最陡原则,采用高程增量叠加算法设定平坦区域网格单元流向,并确定全流域网格流向。在得到每个网格流向与水流累积值后,即可确定该流域的河网^[4]。首先以水流累积值最大的点作为流域的出口,标记为1号河网的起点,即精确的出口断面位置;给定最小水道给养面积阈值,生成河网并计算回水面积;最后子流域之间通过子流域的流向确定了拓扑关系^[5]。本文的研究区域中,最小水道给养面积阈值CSA和最小河长MSCL分别取 37 km^2 和 5 km 。

2 水文模型概述

2.1 RR模型

RR模型是MIKE11水环境模拟软件套装中NAM降雨径流模型的一个模块。MIKE11是由丹麦水力研究所(DHI)研制的一套一维水力学模型,用于河流系统模拟分析、设计、管理和调度。主要包括:水动力学模型(HD)、降雨径流模型(NAM)、洪水实时预报模型(FF)等。RR模型是NAM中较为经典的一个模块,在世界各地众多的水文模拟中得到广泛应用^[6]。

RR模型是一个集总参数的概念性水文模型,由水文循环中一系列定量的各种陆相特征所连接,分为融雪蓄水层、地表蓄水层、浅层蓄水层、地下蓄水层共4层蓄水体对研究流域进行产汇流模拟计算。RR模型模拟的是自然流域的降雨径流过程,其要求的资料输入为降水、蒸发,考虑融雪径流时还需要气温资料的输入。模型结果的输出包括径流、地下水位、土壤含水量和地下水的补给量^[7]。

2.2 新安江模型

新安江模型是河海大学赵人俊等研制的水文模型,在国内外湿润半湿润地区得到了广泛的应用。

新安江模型最初是二水源模型,包括直接径流和地下径流。产流计算用蓄满产流方法,流域蒸发

采用二层或三层蒸发,水源划分用的是稳定下渗法,直接径流坡面汇流采用串联线性水库概念提出的瞬时单位线法,地下径流坡面汇流用离散线性水库法,河道汇流采用马斯京根河段演算法。

新安江模型已发展为三水源——地面径流、壤中流、地下径流以及其他多水源模型。三水源新安江模型结构与二水源相比,主要有水源划分和产流计算单元划分两方面的改进^[8]。

3 水文模型序列信息熵的应用

水文现象的复杂性和水文输入信息的不完整,往往导致水文模型不唯一,甚至辨识参数的不唯一。对同一个研究对象(诸如流域),如何甄别模型参数,比较不同水文模型所包含的信息量,是十分重要的。

从水文信息论的观点来看,有多种途径可以解决此类问题。例如可从模型的水文物理概念上进行评述;也可以比较水文模型与实测序列误差的信息序列,即从系统输出的观测来进行评述^[9]。相较而言,前者属于语义信息的评述范围,量化相对困难,定性相对较多。而后者容易定量化表达,其内在信息量丰富,是综合了多种因素作用反应后的结果。

3.1 信息熵函数的基本性质

C. Shannon给出的信息熵测度定义可以按照如下形式表示:

$$I(Y) = -K \sum_{j=1}^s y_j \ln y_j \quad (j \in J, k = 1/\ln 2) \quad (1)$$

式中: Y 为 s 维(s 为有限集或无限集)随机变量序列; y_j 为 Y 中第 j 维分量得的概率分布密度,满足: $0 \leq y_j \leq 1, K$ 为对数转化常数, $K = 1/\ln 2$ 。

式(1)具有如下基本性质^[3]:1.非负性;2.连续性;3.可加性;4.单调性;5.有极值。当概率分布为均匀分布时,信息熵取极值,即:

当取 $y_1 = y_2 = \dots = y_s = 1/s$ 时,有最大信息熵测度:

$$I_{\max}(X) = K \ln s \quad (2)$$

由(1)式定义可知,当序列 X 的元素反映两次测试结果之差,则可得出:

(1) X 各分量值差异性越大,计算信息熵测度值越大,序列所包含的信息量就越大;

(2) X 各分量无差异性时,序列所含的信息熵测度为零,故信息量为零。

3.2 流域径流信息熵的计算方法

对流域径流计算结果与实测结果的吻合度越好,说明模型模拟计算越接近实际水文过程,模型评

价值就应越高。在构建流域径流信息熵计算的归一化残差序列时,需要考虑该点与信息熵序列差异越大值越大的不同。

对流域年径流观测值和计算值序列,可按直方图构建日残差序列如下:

选取流域的实测序列为 $\{Q(k)\}$, $k=1,2,\dots,n$, n 为参考标准序列,对水文模型 i 有计算值:

$$Q_i(k) = M_i(P(k), \theta_i) \quad (3)$$

式中: $P(k)$ 为模型的输入信息; k 为离散时间; θ_i 为模型 i 的参数; M_i 为模型 i 的结构函数。结构函数反映了模型对产汇流机理的认识程度。结构函数确定后,其模拟计算所得的径流量由输入信息和模型参数所确定。由此可见,模型结构、输入信息以及模型参数的不确定性可由 Q_i 模拟的精确程度所评定。

模型 i 与实测序列的残差平方和 $\Delta_i^2 Q(k)$ 为:

$$\Delta_i^2 Q(k) = [Q_i(k) - Q(k)]^2 \quad (4)$$

利用差异信息序列的信息测度对模型信息作出评价,定义归一化的残差序列为:

$$y_i(k) = \left(\frac{1}{1 + \Delta_i^2 Q(k)} \right) / \left(\sum_{i=1}^n \frac{1}{1 + \Delta_i^2 Q(k)} \right) \quad (5)$$

故模型 i 的信息熵潜在信息测度和相对测度计算公式分别为:

$$I(Q_i) = -K \sum_{k=1}^n y_i(k) \ln y_i(k) \quad (6)$$

$$I_e(Q_i) = \frac{I(Q_i)}{I_{\max}(Q_i)} \times 100\% \quad (7)$$

同理,也可由上述方式计算出模型 j 的残差平方序列 $\Delta_j^2 Q(k)$ 和归一化序列 $y_j(k)$,信息熵潜在信息测度 $I(Q_j)$ 和相对测度 $I_e(Q_j)$ 。

当输入信息确定后,对水文模型比较,反映了模型结构及参数选定的合理性。如对水文模型 i 与 j 比较,则可通过水文信息测度 $J(Q_i)$ 与 $I(Q_j)$ 的对比来说明。信息熵越大,则模型对原型的模拟效果越好。

反映两个模型信息量的差异定义如下^[10]:

$$\Delta I_{ij}(Q) = \frac{I(Q_j) - I(Q_i)}{I(Q_i)} \times 100\% \quad (8)$$

该定义也可评价相同信息输入来源的同一模型不同参数下的信息量差异。

3.3 RR模型和新安江模型的应用实例

本文选取1992、1995、1997、1999年作为率定期,2000、2001作为验证期进行日径流模拟。

关于模型参数的率定,RR模型采用SCE-UA优化算法自动率定。新安江模型采用人工试错法,

以确定性系数DC和径流深相对误差BIAS作为目标函数进行率定。日径流模拟参数率定结果如表1、2所示。

表1 RR模型日径流模拟参数表

U_{\max}	L_{\max}	$CQOF$	$CKIF$	$CK_{1,2}$
10.2	100	0.998	532.8	35.1
TOF	TIF	TG	$CKBF$	
0.0932	0.169	0.942	1707	

表2 新安江模型日径流模拟参数表

IM	WM	WUM	WLM	B
0.01	135	80	20	0.4
C	FC	K	KKG	U
0.18	110	0.2	0.99	33

模拟结果如图1~6所示:

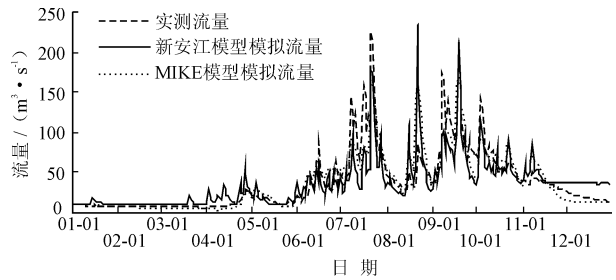


图1 1992年日径流过程

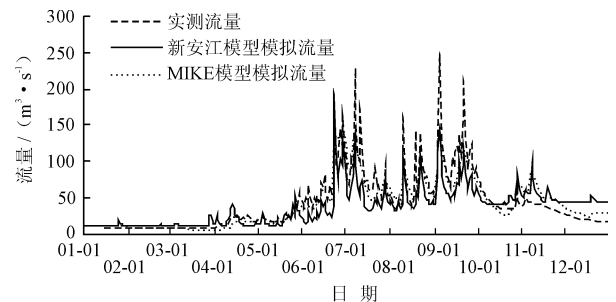


图2 1995年日径流过程

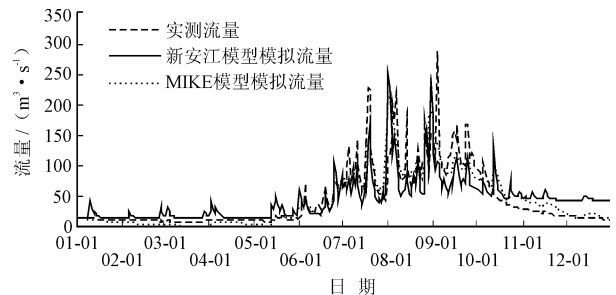


图3 1997年日径流过程

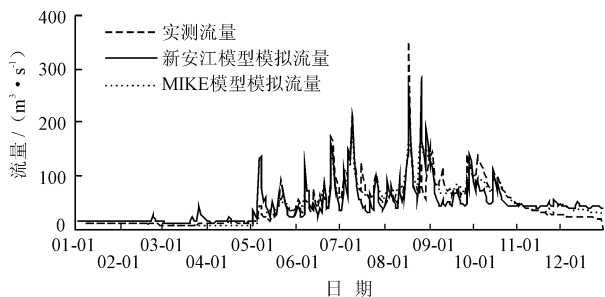


图 4 1999 年日径流过程

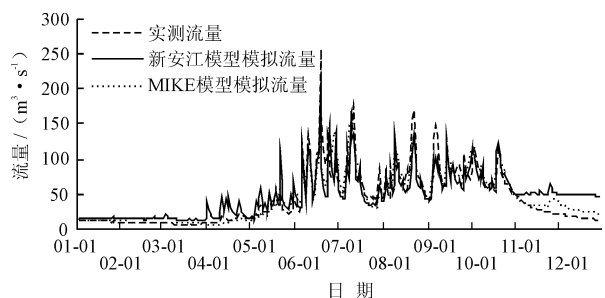


图 5 2000 年日径流过程

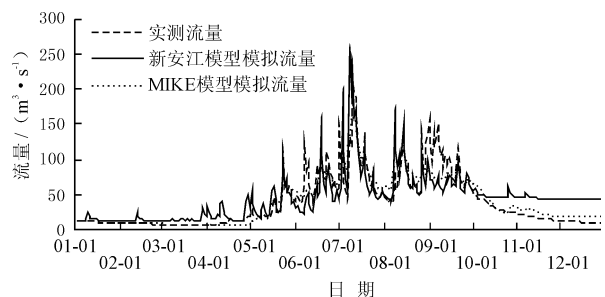


图 6 2001 年日径流过程

由图 1~6 可以看出,RR 模型与新安江模型都较好地模拟了四川省安宁河流域安宁桥站的日径流过程。RR 模型模拟的洪峰过程比较矮胖,新安江模型模拟的洪峰过程比较尖瘦。

两个模型模拟效果的比较结果参见表 3 与表 4。可见两个模型信息熵的信息测度均接近于 7.9。两个模型相比,RR 模型每一年的确定性系数均大于新安江模型,1992、1997、1999 年 3 年 RR 模型的径流深相对误差较新安江模型的要大,其中 1992 年 RR 模型与新安江模型的信息熵基本相等,另外两年 RR 模型的信息熵相对较大。1995、2000、2001 年 3 年的径流深相对误差较新安江模型的要小,其中 2000 年 RR 模型与新安江模型的信息熵基本相等,另外两年 RR 模型的信息熵相对较大。

由测度计算结果可以看出,水文模型序列的确定性系数和相对径流深误差与信息熵之间不具有相关性。这符合差异信息理论的非唯一性原理,即灰色理论的研究对象,由于信息的不完全性、不确定性

和其他因素,将不可避免地造成多准则,具体表现为前因后果往往是多对多的映射,即过程的非唯一性。模型不唯一、辨识参数不唯一、解不唯一、决策方法和结果不唯一是其具体体现。

表 3 RR 模型与新安江模型日径流模拟效果比较 %

年份	确定性系数		径流深相对误差		
	RR 模型	新安江模型	RR 模型	新安江模型	
率定期	1992	0.77	0.68	0	-0.03
	1995	0.81	0.68	-0.02	-0.12
	1997	0.79	0.61	-0.02	0.03
验证期	1999	0.75	0.64	0.01	-0.02
	2000	0.77	0.67	0	0.06
	2001	0.81	0.60	0	0.12

表 4 RR 模型与新安江模型径流过程的序列信息测度比较 %

年份	信息熵的信息测度		信息熵的相对测度		信息量差异	
	RR 模型	新安江模型	RR 模型	新安江模型		
率定期	1992	7.97	7.91	99.90	99.17	0.73
	1995	7.96	7.83	99.69	98.18	1.51
	1997	7.98	7.79	100.00	97.67	2.33
验证期	1999	7.96	7.85	99.76	98.35	1.41
	2000	7.97	7.96	99.87	99.74	0.14
	2001	7.89	7.78	98.87	97.55	1.32

水文模型残差序列信息熵反映了模型计算序列与实测序列过程线的垂向相似性。信息熵较大时,水文模型 i 的模拟值 $Q_i(k)$ 与观测值 $Q(k)$ 之差为常数 C ,即模型模拟序列的过程线垂向平移 C 个单位后与观测序列过程线相等。水文模型残差序列的信息熵越大,残差序列趋向于常数 C 的不确定性越高,故水文模型序列信息熵与确定性系数和相对径流深误差之间并无特定关系。

4 结 语

水文序列的信息是可以度量,并且可以比较信息量的大小和顺序的。研究表明依据信息熵最大理论对水文模型的参数确定和模拟效果评价是有效的,针对不同水文模型所具有的水文序列信息,可以基于水文差异序列信息熵和差异信息熵的测度公式进行,它们为信息量大小评价和模型选择提供了准则和依据。信息熵越大,模型模拟序列包含的信

息量越大。但同时也须指出在对不同模型进行对比评价时,信息熵的相对值也满足差异信息理论的非唯一性原理。序列的确定性系数和径流深相对误差与信息熵相互独立。因此在信息熵的比较应用过程中,需要结合确定性系数和径流深相对误差等指标,通过对其信息熵的比较,进而判断不同水文模型模拟结果的优劣,才具有研究价值。

参考文献:

- [1] 徐宗学. 流域水文模型[M]. 北京:科学出版社,2009:1-16.
- [2] 刘洋. 山溪洪水风险分析方法研究[D]. 北京:北京工业大学,2013.
- [3] 龚伟. 基于信息熵和互信息的流域水文模型不确定性分析[D]. 北京:清华大学,2012.
- [4] 王中根,刘昌明,等. 基于DEM的分布式水文模型研究综述[J]. 自然资源学报,2003,18(2):168-173.
- [5] 李巧玲. 基于DEM的水文模型及流域汇流非线性问题研究[D]. 南京:河海大学,2007.
- [6] Rahman M M, Goel N K, Arya D S. Development of the jamuneswari flood forecasting system: case study in angladesh[J]. Journal Of Hydrologic Engineering, 2012, 17: 1123-1140.
- [7] Rahman M M, Arya D S, Goel N K, et al. Design flow and stage computations in the teesta river, bangladesh, using frequency analysis and MIKE 11 modeling[J]. Journal of Hydrologic Engineering, 2011, 16: 176-186.
- [8] 包为民. 水文预报[M]. 北京:中国水利水电出版社,2009:143-167.
- [9] 王文圣,张翔,等. 水文不确定性分析方法[M]. 北京:科学出版社,2011:243-262.
- [10] 夏军. 灰色系统水文学——理论、方法及应用[M]. 武汉:华中理工大学出版社,2000:82-96.
- [9] Samad F A. An evaluation technique for water resources in a complex river system[D]. Sydney: University of New South Wales,1982.
- [10] Kondo K. Economic analysis of water resources in Japan: using factor decomposition analysis based on input-output tables[J]. Environmental economics and Policy Studies,2005,7(2):109-129.
- [11] Erik D, Esther V. Analysing andalusian virtual water trade in an input-output framework[J]. Regional studies,2007,41(2):185-196.
- [12] 李林红. 滇池流域可持续发展动态投入产出最优控制模型[J]. 控制与决策,2001,16(Z1):685-688.
- [13] 方国华,钟淋娟,吴学文,等. 水资源利用和水污染防治投入产出最优控制模型研究[J]水利学报,2010,41(9):1128-1134.
- [14] 江苏省统计局,国家统计局江苏调查总队. 江苏统计年鉴[M]. 北京:中国统计出版社,2011.
- [15] 江苏水利厅. 江苏省水资源公报[Z]. 2010.
- [16] 雒晓娜. 投入产出系数的修订及其多目标优化模型应用研究[D]. 大连:大连理工大学,2006.
- [17] 丁莉娅. 火电用水占工业用水1/6,火电节水需做好[DB/OL]. (2011-09-09)[2014-06-15]. <http://finance.sina.com.cn/roll/20110909/174910462210.shtml>.

(上接第55页)