

新型耦合数据驱动模型在降雨径流模拟中的应用研究

梁珂¹, 阚光远^{2,3}, 李致家¹

(1. 河海大学 水文水资源学院, 江苏 南京 210098; 2. 流域水循环模拟与调控国家重点实验室, 中国水利水电科学研究院, 水利部防洪抗旱减灾工程技术研究中心, 北京 100038; 3. 水利科学与工程国家重点实验室, 清华大学水利系, 北京 100038)

摘要:为解决传统数据驱动模型的不足,使其能实现降雨径流过程高精度连续模拟,提出新型耦合数据驱动模型——PEK,即:基于偏互信息的输入变量选择、基于新型集成神经网络的出流量预测和基于 K 最近邻模型的出流量误差预测。PEK模型具有以下特点:(1)提出了基于分离式选择策略和滑窗累积雨量的模型候选输入向量,并与基于偏互信息的输入变量选择方法联合使用,提高了输入信息的充分性和无冗余性,对建立精度高、泛化能力强的高质量模型意义重大;(2)提出了新型集成神经网络——EBPNN及其率定方法。联合使用NSGA-II多目标优化算法和早停止Levenberg-Marquardt算法,通过一次优化过程同时确定全局最优个体网络个数、各个体网络拓扑结构和网络参数。个体网络权重由基于AIC信息准则的权重优选方法确定。EBPNN在模拟精度和网络复杂度间取得了良好折衷,精度高、泛化能力强、率定结果客观;(3)PEK模型能够进行多步外推预报,实现了非实时校正模式下的高精度连续模拟,增长了预见期;(4)PEK模型不需要进行流域状态变量的计算,仅需初始出流量就可进行出流量的连续模拟。在呈村流域应用PEK和CLS两个数据驱动模型进行次洪降雨径流模拟及精度比较。结果表明PEK模型使用简便,模拟精度高于CLS模型,实现了多步外推的高精度连续模拟,增长了数据驱动模型的预见期。

关键词:降雨径流模拟;非实时校正;数据驱动模型;PEK模型;最优化方法

中图分类号:P338.1 文献标识码:A 文章编号:1000-0852(2016)04-0001-07

1 引言

经过多年发展,降雨径流模型日趋成熟,可以分为两类:概念性模型和数据驱动模型。其中,数据驱动模型因其简洁有效性广泛应用于降雨径流模拟中,取得了良好应用效果^[1-2]。常见的数据驱动模型有时间序列模型、神经网络模型和最近邻模型等^[3-4]。基于数据驱动模型的降雨径流模拟还存在一些问题,首先是建模方式和输入变量选择。以往大多数数据驱动模型采用的是实时校正模式,即用降雨量和实测前期流量(即预报时刻之前的实测出流量)作为模型输入对出流量进行预测。该模式只能进行单步外推预报,预见期短。此

外,数据驱动模型的降雨量最优输入向量通常筛选自大量离散单时刻降雨量,但实际上出流量往往与具有一定历时的累积降雨量密切相关,而非离散单时刻降雨量,这导致降雨径流关系模拟效果变差。水文模拟中应用最为广泛的数据驱动模型当推神经网络模型,但它在以下几个方面尚存在一些不足:输入变量选择、模型构建方式、拓扑结构和网络参数优化、网络集成方法。

本文对基于数据驱动模型的次洪降雨径流模拟进行了分析研究,针对上文提到的问题,以提高非实时校正降雨径流模拟精度为目的,建立了新型耦合数据驱动模型(PEK)。本文进行了以下改进和验证:(1)提出了基于分离式选择策略和滑窗累积雨量的模型候选输

收稿日期:2015-05-09

基金项目:国家自然科学基金项目(41130639,51179045);水利部公益项目(201501022);中国水科院科研专项(JZ0145B052016);中国水科院国际水利水电科技发展动态调研专项(JZ0145C102015)

作者简介:梁珂(1992-),女,河南新乡人,硕士研究生,主要从事水文模型和洪水预报研究。E-mail:liangkepapers@126.com

入向量,并与基于偏互信息的输入变量选择方法联合使用;(2)提出了新型集成神经网络——EBPNN及其率定方法。通过NSGA-II多目标优化算法和早停止Levenberg-Marquardt算法一次性确定全局最优个体网络个数、各个体网络拓扑结构和网络参数。个体网络权重由基于AIC信息准则的权重优化方法确定;(3)在呈村流域应用PEK和CLS两个数据驱动模型进行次洪降雨径流模拟和精度比较。研究结果表明PEK模型能够进行多步外推预报,实现了非实时校正模式下的高精度连续模拟,增长了数据驱动模型的预见期。PEK模型不需要进行流域状态变量的计算,仅需初始出流量就可进行出流量的连续模拟,非常实用方便。研究成果可应用于降雨径流模拟与预报,也可作为常用水文模拟方法的辅助和补充,为洪水预报、防洪规划决策和水资源评价管理等提供技术支撑,具有重要的现实意义。

2 PEK模型

2.1 改进的输入变量选择方法

2.1.1 输入变量选择方法

从候选输入变量中挑选输入变量的方法称为输入变量选择方法。常见的输入变量选择方法有试算法、启发式算法、先验知识和统计分析等。统计分析方法适用性较广,效率高,结果稳定可靠,非常适合于数据驱动模型^[5]。因此,本文采用由Sharma提出的基于偏互信息的输入变量选择方法,是当前最好的基于统计分析方法的输入变量选择方法^[6-7]。

2.1.2 候选输入向量

(1)分离式选择策略。以往文献中,候选输入向量中的实测前期流量和降雨量通常同时参与选择,但数值试验表明如果实测前期流量和降雨量同时选择,则几乎没有降雨量输入变量能够被选入。这是由于实测前期流量与出流量间的关联性远大于降雨量与出流量间的关联性。因此,为了能够包含足够的降雨信息,实测前期流量与降雨量应分别经过不同的输入变量选择过程选入。本文提出了“分离式”候选输入向量表示方法:

$$X_t = (X_t^{(Q-OBS)}, X_t^{(P)})^T \quad (1)$$

$$X_t^{(Q-OBS)} = (Q_{i-1}^{(OBS)}, Q_{i-2}^{(OBS)}, \dots, Q_{i-n_Q}^{(OBS)})^T \quad (2)$$

$$X_t^{(P)} = (P_t, P_{t-1}, \dots, P_{t-n_P+1})^T \quad (3)$$

式中: $X_t^{(Q-OBS)}$ 表示 t 时刻实测前期流量候选输入向量; $X_t^{(P)}$ 表示 t 时刻降雨量候选输入向量。则 t 时刻最优输入向量可表示为:

$$X_t^{(S)} = (IVS_{Q-OBS}(X_t^{(Q-OBS)}), IVS_P(X_t^{(P)}))^T \quad (4)$$

式中: IVS_{Q-OBS} 和 IVS_P 分别表示实测前期流量候选输入向量和降雨量候选输入向量的基于偏互信息的输入变量选择。最优输入输出样本对可简写为 $X_t^{(S)} \sim Q_t^{(OBS)}$ 。

(2)滑窗累积雨量。通常情况下,降雨量候选输入向量的构成方式为式(3),但数值试验表明式(3)构成的降雨量最优输入向量无法包含足够的降雨信息,模拟效果不佳。原因在于式(3)中选出的降雨量输入变量通常是离散的、不连续的单时刻降雨量,事实上,与出流量密切相关的是具有一定历时和时延的累积降雨量。本文提出基于滑窗累积雨量的降雨量候选输入向量:

$$X_t^{(SWCR)} = (SWCRs_t^{(1)}, SWCRs_t^{(2)}, \dots, SWCRs_t^{(n_p)})^T \quad (5)$$

式中: $X_t^{(SWCR)}$ 表示 t 时刻滑窗累积雨量候选输入向量; $SWCRs_t^{(i)}$ 表示对应于 t 时刻的一系列滑窗累积雨量的集合,这些滑窗累积雨量具有相同的滑窗宽度(宽度为 i), $i=1, 2, \dots, n_p$ 。 $SWCRs_t^{(i)}$ 的组成方式见表1。

表1 $SWCRs_t^{(i)}$ 的组成方式
Table1 The configuration of $SWCRs_t^{(i)}$

| $SWCRs_t^{(i)}$ | 宽度 | 候选输入变量个数 | 候选输入变量 |
|-------------------|-------|----------|--|
| $SWCRs_t^{(1)}$ | 1 | n_p | $P_t, P_{t-1}, \dots, P_{t-n_p+1}$ |
| $SWCRs_t^{(2)}$ | 2 | n_p-1 | $P_t+P_{t-1}, P_{t-1}+P_{t-2}, \dots, P_{t-n_p+2}+P_{t-n_p+1}$ |
| $SWCRs_t^{(3)}$ | 3 | n_p-2 | $P_t+P_{t-1}+P_{t-2}, P_{t-1}+P_{t-2}+P_{t-3}, \dots, P_{t-n_p+3}+P_{t-n_p+2}+P_{t-n_p+1}$ |
| ... | ... | ... | ... |
| $SWCRs_t^{(n_p)}$ | n_p | 1 | $P_t+P_{t-1}+\dots+P_{t-n_p+1}$ |

2.2 新型集成神经网络——EBPNN

2.2.1 EBPNN结构

本文提出一种新型集成神经网络——EBPNN(ensemble back propagation neural network),该模型采用三层反向传播神经网络作为个体网络,EBPNN模型的个体网络分别对输入进行计算,求得各自的输出,通过加权平均法求出最终集成神经网络的输出。个体

网络通过基于 AIC 信息准则的加权平均方法求得集成网络的输出。

2.2.2 EBPNN 率定方法

EBPNN 的特点在于提出了新型个体网络生成方法和个体网络权重生成方法。个体网络由基于 NSGA-II 算法和早停止 LM 算法的个体网络生成方法生成,该方法能自动的一次性确定个体网络的个数、各个体网络的最优拓扑结构和网络参数,生成的个体网络在保证模拟精度的前提下具有良好的多样性,提高了集成神经网络的泛化能力。使用基于 AIC 信息准则的个体网络权重生成方法对个体网络的模拟结果进行集成,在保证模拟精度的前提下提升了泛化能力。

(1)集成神经网络个体网络生成方法。新型集成神经网络模型通过对个体网络进行编码和优化来生成有限个个体网络,进而组成集成神经网络^[8-9]。为了提高优化效率,需要事先确定决策变量的上下限,尽量减小搜索空间的大小。优化中 NSGA-II 算法的目标函数有两个,分别为训练误差最小化和拓扑结构复杂度最小化。决策变量由 NSGA-II 算法进化,进化完毕后获得的一阶帕累托前沿中的各解即为各最优个体网络,经解码后生成所有最优个体网络。由于 NSGA-II 算法局部寻优能力较弱,故生成的个体网络的权值和偏置值还需要经过早停止 LM 算法进行精细训练。在本文中,大约率定期样本的 3/4 用来作为训练集,余下的率定期样本用来作为早停止算法的测试集。

(2)集成神经网络个体网络权重生成方法。个体网络生成完毕后,需要给每个个体网络赋予一个权重并根据此权重对个体网络输出进行加权平均求得集成神经网络总输出。本文采用基于 AIC 信息准则的个体网络权重生成方法,方法详见文献[10]。

2.3 新型耦合数据驱动模型——PEK 模型

2.3.1 PEK 模型结构

PEK 模型的全称为 PMI-based input variable selection, EPBNN-based discharge forecasting and K-nearest neighbor based discharge error forecasting,即:基于偏互信息的输入变量选择,基于新型集成神经网络的出流量预测和基于 K 最近邻算法的出流量误差预测。PEK 模型使用模拟前期流量替代实测前期流量,实现了非实时校正模式下的高精度连续模拟。PEK 模型通过 EBPNN,根据输入向量计算对应的预报输出,通过 K 最近邻算法,根据输入向量计算对应的预报误差,最后将预报输出与预报误差相加得到模拟输

出,PEK 模型结构图见图 1。

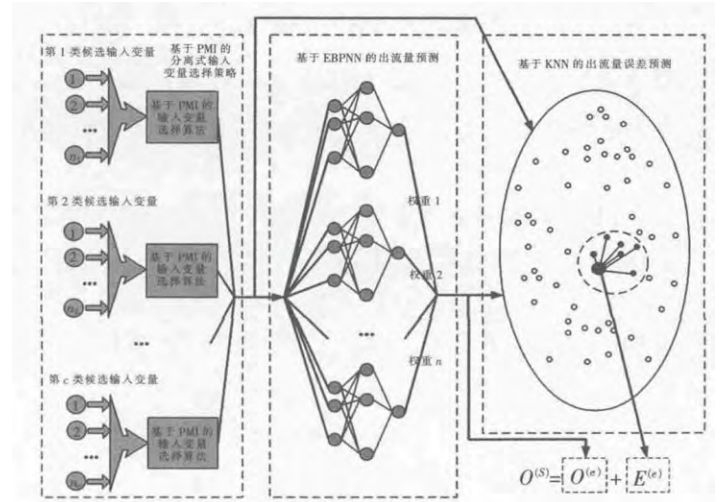


图 1 PEK 模型结构

Fig.1 The structure of PEK model

PEK 模型建模方式如下:

$$X_t = (X_t^{(Q_SIM)}, X_t^{(SWCR)})^T \quad (6)$$

$$X_t^{(Q_SIM)} = (Q_{t-1}^{(SIM)}, Q_{t-2}^{(SIM)}, \dots, Q_{t-n_Q}^{(SIM)})^T \quad (7)$$

$$X_t^{(SWCR)} = (SWCRs_t^{(1)}, SWCRs_t^{(2)}, \dots, SWCRs_t^{(n_s)})^T \quad (8)$$

$$X_t^{(S)} = (IVS_{Q_SIM}(X_t^{(Q_SIM)}), IVS_{SWCR}(X_t^{(SWCR)}))^T \quad (9)$$

$$Q_t^{(FOC)} = F_{EBPNN}(X_t^{(S)}) \quad (10)$$

$$E_t^{(FOC)} = F_{KNN}(X_t^{(S)}) \quad (11)$$

$$Q_t^{(SIM)} = Q_t^{(FOC)} + E_t^{(FOC)} \quad (12)$$

式中: $X_t^{(Q_SIM)}$ 表示 t 时刻模拟前期流量候选输入向量; $Q_{t-i}^{(SIM)}$ 表示 $t-i$ 时刻模拟前期流量, $i=1, 2, \dots, n_Q$; IVS_{Q_SIM} 表示模拟前期流量候选输入向量的基于偏互信息的输入变量选择; $IVS_{Q_SIM}(X_t^{(Q_SIM)})$ 表示 t 时刻模拟前期流量最优输入向量; $Q_t^{(FOC)}$ 表示 t 时刻预报出流量; $E_t^{(FOC)}$ 表示 t 时刻出流量误差; $Q_t^{(SIM)}$ 表示 t 时刻模拟出流量,即 t 时刻 PEK 模型的模拟结果; F_{EBPNN} 为 EBPNN 出流量预测; F_{KNN} 为 KNN 出流量误差预测。

对于单场次洪,PEK 模型的非实时校正模拟过程如下:为了求出 t 时刻的模拟出流量 $Q_t^{(SIM)}$,首先,根据式(7)和(8)生成 $X_t^{(Q_SIM)}$ 和 $X_t^{(SWCR)}$;其次, $X_t^{(Q_SIM)}$ 和 $X_t^{(SWCR)}$ 生成后,根据式(9)求出最优输入向量 $X_t^{(S)}$,此为

输入变量选择;第三步,根据式(10)和(11)计算 $Q_t^{(FOC)}$ 和 $E_t^{(FOC)}$,此为出流量预测和出流量误差预测;第四步,根据式(12)计算 $Q_t^{(SIM)}$ 。以上计算过程从 $t=1$ 迭代至 $t=T$ (T 为单场次洪的数据个数),所有迭代完成后就实现了非实时校正模拟。

2.3.2 PEK 模型率定方法

PEK 模型的率定涉及到 4 个映射关系的确定:(1) $IVS_{Q_{SIM}}$, 模拟前期流量候选输入向量的基于偏互信息的输入变量选择;(2) IVS_{SWCR} , 滑窗累积雨量候选输入向量的基于偏互信息的输入变量选择;(3) F_{EBPNN} , EBPNN 出流量预测;(4) F_{KNN} , KNN 出流量误差预测。

(1) $IVS_{Q_{SIM}}$:为了确定哪些模拟前期流量应被选入最优输入向量,首先,对于每场洪水,使用率定集生成候选输入输出样本对 $X_t^{(Q_{SIM})} \sim Q_t^{(OBS)}$ 。在率定过程中,使用实测前期流量来代替模拟前期流量生成模拟前期流量候选输入向量,记为 $\hat{X}_t^{(Q_{SIM})}$ 。各场洪水样本分别生成完毕后,将这些样本混合在一起组成总样本集,使用基于偏互信息的输入变量选择算法进行模拟前期流量输入变量的选择。

(2) IVS_{SWCR} :为了确定哪些滑窗累积雨量应被选入最优输入向量,首先,对于每场洪水,使用率定集生成候选输入输出样本对 $X_t^{(SWCR)} \sim Q_t^{(OBS)}$ 。各场洪水样本分别生成完毕后,将这些样本混合在一起并使用基于偏互信息的输入变量选择算法进行降雨量输入变量的选择。

(3) F_{EBPNN} : $IVS_{Q_{SIM}}$ 和 IVS_{SWCR} 确定后可依此生成最优输入输出样本对 $\hat{X}_t^{(S)} = (IVS_{Q_{SIM}}(\hat{X}_t^{(Q_{SIM})}), IVS_{SWCR}(X_t^{(SWCR)})) \sim Q_t^{(OBS)}$,这些样本对作为 EBPNN 的率定样本对 EBPNN 进行率定。

(4) F_{KNN} : F_{EBPNN} 率定完毕后使用 $\hat{X}_t^{(S)}$ 和率定完毕的 EBPNN 计算预报出流量序列 $Q_t^{(FOC)}$,然后由 $E_t^{(FOC)} = Q_t^{(OBS)} - Q_t^{(FOC)}$ 计算出流量误差序列 $E_t^{(FOC)}$ 。最后,使用输入输出样本对 $\hat{X}_t^{(S)} \sim E_t^{(FOC)}$ 率定 K 最近邻算法的参数 K 。

3 CLS 模型

由 Natale 和 Todini 提出的约束线性系统(CLS)模型是总径流响应(TLR)模型的改进版,添加了非负响

应和水量平衡两个约束条件,此外,为了考虑雨强对产汇流过程的影响,CLS 模型依照不同的阈值将降雨量划分为几个部分,各部分分别用不同的响应函数进行产汇流计算。本文使用的 CLS 模型具有两个阈值。

4 模型应用与结果讨论

4.1 研究区域和资料概况

呈村流域位于安徽省的钱塘江流域,流域面积 290km^2 。该流域位于亚热带季风气候区,年平均温度 17°C ,年平均降水量 1600mm ,其中 4~6 月多雨,占 50%,易发生洪涝灾害,7~9 月占 20%,旱灾频繁。河川径流年内、年际变化较大,属于典型的湿润地区。研究区内有 10 个雨量站,将研究流域按照泰森多边形划分成 10 个子流域。选取 1990~1999 年间 20 场次洪资料进行次洪模型计算(其中 14 场用于模型率定,6 场用于模型检验)。蒸发站为呈村站。流域图见图 2。

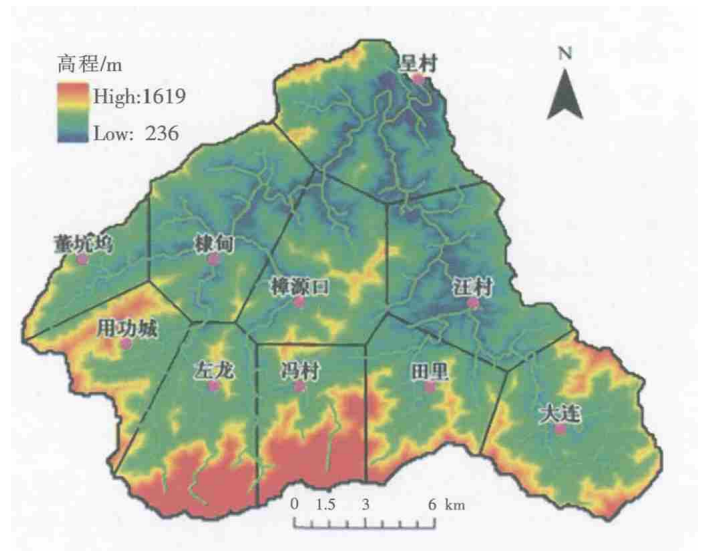


图 2 呈村流域图

Fig.2 The Chengcun watershed

4.2 PEK 模型率定

4.2.1 最优输入向量选择

呈村流域的次洪历时最小值为 80h,因此,为保证输入变量集能够包含充足的输入信息,首先尝试性地将呈村流域的阶数设置为 $n_I = n_O = 24$,使用基于偏互信息的输入变量选择方法进行输入变量选择。选择结果表明,用于生成最优滑窗累积雨量的降雨均介于 P_t 与 P_{t-23} 之间,最优预报前期流量和模拟前期流量均介于 $t-1$ 与 $t-24$ 时刻之间。因此,变量选择结果表明 $n_I = n_O = 24$ 对于呈村流域是足够大的,选择结果是合理的。

4.2.2 EBPNN 率定

为了提高优化效果和效率,NSGA-II 算法参数设置如下:种群数 $pop=100$,总进化代数 $gen=1\ 000$,交叉概率 $pc=0.9$,变异概率 $pm=0.1$ 。LM 算法参数设置如下:最小梯度 $min_grad=1E-10$,初始 μ 值 $\mu=0.001$, μ 减少因子 $\mu_dec=0.1$, μ 增加因子 $\mu_inc=10$, μ 最大值 $\mu_max=1E-10$ 。对于早停止策略,均匀选取大约 3/4 的率定样本作为训练集,其余率定样本作为测试集,测试失败次数为 5。个体网络最优拓扑结构和权重见表 2。

表2 PEK模型个体网络最优拓扑结构和权重
Table2 The optimal topology and parameter settings of the component network for PEK model

| 隐含层神经元最大个数 | 拓扑结构 | 权重 |
|------------|--------------|---------------|
| 17 | 8-2-1 | 0.1439 |
| | 8-3-1 | 0.2156 |
| | 8-4-1 | 0.2147 |
| | 8-6-1 | 0.1405 |
| | 8-7-1 | 0.0957 |
| | 8-8-1 | 0.0804 |
| | 8-9-1 | 0.1093 |

过多的隐含层神经元会导致过拟合,降低个体网络的泛化能力。如表 2 所示,最优隐含层神经元个数大部分都没有超过输入层神经元个数。仅有 1 个个体网络的隐含层神经元个数稍大于输入层神经元个数(8-9-1)。表 2 中加粗的拓扑结构具有最大权重,在某种程度上可认为是帕累托最优集中最优的拓扑结构。这些最优拓扑结构的隐含层神经元个数大部分很小且均未超过输入层神经元个数。以上这些结果表明,优化出的个体网络在满足模拟精度的前提下,其网络规模均较小,具有更好的泛化能力。

4.3 CLS 模型率定

本文的 CLS 模型采用两个雨量阈值,即模型有三个子响应函数。每个子响应函数的宽度(即 k_1 、 k_2 和 k_3)通过试算法优化。根据流域洪水历时的最小值,子响应函数宽度的范围预先设置为[1, 24]。下文将通过优化结果的分析来说明这些设置的合理性。对于每组候选的 k_1 、 k_2 和 k_3 值,由二次规划算法率定 CLS 模型,同时记录模拟误差。当所有候选的 k_1 、 k_2 和 k_3 值均试算完毕,选择具有最小预测误差的宽度值作为最优解。CLS 模型率定结果(即响应函数)见图 3。如图 3 所示,呈村流域响应函数的宽度大部分都小于上

限 24,这说明响应函数宽度上限设置为 24 是合理的。

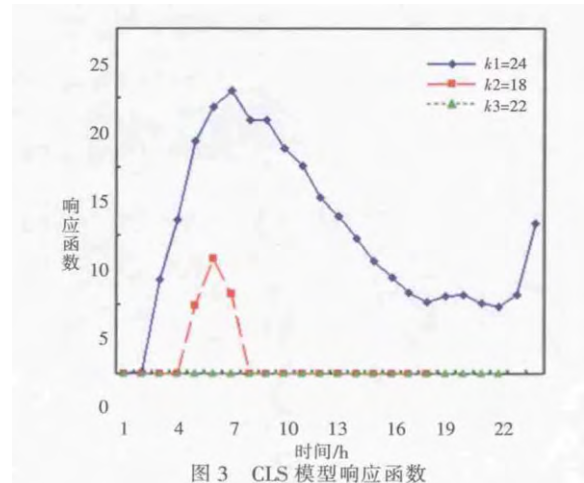


图3 CLS 模型响应函数
Fig.3 The response function of CLS model

4.4 误差评定准则

本文次洪模拟结果的评价基于纳须效率系数(CE)、均方根误差(RMSE)和平均绝对值误差(MAE)三个准则。

4.5 模拟精度比较

PEK 与 CLS 模型模拟结果误差统计表见表 3。对于呈村流域率定期和检验期,PEK 模型模拟效果均明显好于 CLS 模型。PEK 模型针对传统数据驱动模型存在的不足进行了多项改进,优化了输入变量选择、集成神经网络设计与训练、添加模拟前期流量和 KNN 出流量误差预测的新型建模方式,取得了最好的模拟效果。CLS 模型是线性模型,非线性模拟能力较差,因此模拟结果不理想。

5 结论

本文建立了新型耦合数据驱动模型——PEK 模型并提出了模型率定方法。在呈村流域将 PEK 模型与 CLS 模型进行了应用比较。PEK 模型使用基于偏互信息的输入变量选择方法从滑窗累积雨量候选输入向量和模拟前期流量候选输入向量中提取充足且无冗余的输入信息,通过基于新型集成神经网络模型的出流量预测和基于 K 最近邻算法的出流量误差预测进行出流量的高精度连续模拟。PEK 模型采用了全新的非实时校正建模方式,不需要实时信息,实现了数据驱动模型的高精度多步外推预报。PEK 模型不需要进行流域状态变量的计算,仅需初始出流量就可进行出流量的连续模拟。PEK 模型显著提高了非实时校正数据驱动模型的模拟精度,增长了预见期。本文提出的模型率定方法较为客

表3 模拟结果误差统计
Table3 The error statistics of the simulation results

| | 洪水序号 | CE | | RMSE | | MAE | |
|-----|------|------|------|-------|--------|-------|-------|
| | | PEK | CLS | PEK | CLS | PEK | CLS |
| 率定期 | 1 | 0.95 | 0.76 | 24.73 | 56.22 | 13.32 | 45.27 |
| | 2 | 0.86 | 0.65 | 37.11 | 59.47 | 30.03 | 44.95 |
| | 3 | 0.84 | 0.67 | 34.01 | 48.25 | 16.28 | 37.38 |
| | 4 | 0.87 | 0.74 | 33.62 | 48.23 | 18.80 | 39.26 |
| | 5 | 0.97 | 0.68 | 22.40 | 76.77 | 9.63 | 41.53 |
| | 6 | 0.90 | 0.66 | 32.60 | 59.71 | 19.48 | 43.66 |
| | 7 | 0.97 | 0.67 | 29.90 | 99.84 | 16.55 | 48.80 |
| | 8 | 0.87 | 0.77 | 59.00 | 78.75 | 33.11 | 58.44 |
| | 9 | 0.96 | 0.81 | 20.17 | 46.07 | 12.90 | 37.55 |
| | 10 | 0.90 | 0.55 | 24.00 | 51.78 | 17.18 | 33.91 |
| | 11 | 0.93 | 0.64 | 25.61 | 58.64 | 16.37 | 48.84 |
| | 12 | 0.93 | 0.68 | 62.99 | 131.13 | 35.33 | 82.04 |
| | 13 | 0.93 | 0.50 | 16.44 | 44.83 | 12.96 | 28.63 |
| | 14 | 0.97 | 0.88 | 18.53 | 36.88 | 9.58 | 27.58 |
| 检验期 | 1 | 0.98 | 0.85 | 32.30 | 89.31 | 27.29 | 73.86 |
| | 2 | 0.94 | 0.49 | 51.27 | 149.91 | 35.06 | 91.29 |
| | 3 | 0.94 | 0.76 | 17.44 | 33.82 | 11.86 | 26.56 |
| | 4 | 0.95 | 0.80 | 27.45 | 53.65 | 16.75 | 35.71 |
| | 5 | 0.97 | 0.88 | 28.13 | 58.30 | 19.35 | 40.54 |
| | 6 | 0.98 | 0.60 | 24.57 | 103.51 | 10.95 | 57.52 |

观,易于使用,受人为因素影响小,率定出的模型具有良好的模拟精度和预报能力。主要结论如下:

(1) 提出了基于滑窗累积雨量的降雨量候选输入向量及输入变量的分离式选择策略,并与基于偏互信息的输入变量选择方法联合使用,确保了输入信息的充足性和无冗余性,对建立精度高、泛化能力强的高质量数据驱动模型意义重大。

(2) 提出了新型集成神经网络——EBPNN 及其率定方法。通过 NSGA-II 多目标优化算法和早停止 Levenberg-Marquardt 算法确定全局最优个体网络个数、各个体网络拓扑结构和网络参数。个体网络权重由基于 AIC 信息准则的权重确定方法确定。EBPNN 在模拟精度和网络复杂度间取得了良好折衷,精度高、泛化能力强、率定结果客观、受人为因素影响小。

(3) 基于新型输入变量选择方法、EBPNN 模型和 K 最近邻算法,构建了新型非实时校正降雨径流模拟模型——PEK 模型,提出了 PEK 模型的率定方法。PEK 模型能够进行高精度多步外推预报,增长了数据驱动模型的预见期。此外,PEK 模型不需要进行流域状态变量的计算,仅需初始出流量就可进行出流量的

连续模拟。

(4) 在呈村流域将 PEK 模型与 CLS 模型进行了应用比较。结果表明 PEK 模型应用效果明显优于 CLS 模型,具有较好的实用价值和应用前景。

参考文献:

- [1] Guanyuan Kan, Cheng Yao. Improving event-based rainfall-runoff simulation using an ensemble artificial neural network based hybrid data-driven model[J]. Stochastic Environmental Research and Risk Assessment, 2015,29(5):1345-1370.
- [2] Phanida Phukoetphim, Asaad Y. Knowledge extraction from artificial neural networks for rainfall-runoff model combination systems[J]. Journal of Hydrologic Engineering, 2014,19(7):1422-1429.
- [3] LU Chen, Vijay P. Singh. Copula entropy coupled with artificial neural network for rainfall-runoff simulation [J]. Stochastic Environmental Research and Risk Assessment, 2014,28(7):1755-1767.
- [4] LI Zhijia, KAN Guanyuan. Improved neural network model and its application in hydrological simulation [J]. Journal of Hydrologic Engineering, 2014,19(10).
- [5] Maier HR. Application of natural computing methods to water resources and environmental modeling[J]. Mathematical and Computer Modelling, 2006,44(5-6):413-414.
- [6] Bowden GJ, Dandy GC, Maier HR. Input determination for neural network models in water resources applications. Part 1-background

- and methodology [J]. *Journal of Hydrology*, 2005,301(1-4):75-92.
- [7] Sharma A. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: part 1—a strategy for system predictor identification[J]. *Journal of Hydrology*, 2000,239:232-239.
- [8] Bowden GJ, Dandy GC, Maier HR. Input determination for neural network models in water resources applications. part 1—background and methodology[J]. *Journal of Hydrology*, 2005a,301:75-92.
- [9] Bowden GJ, Maier HR, Dandy GC. Input determination for neural network models in water resources applications. part 2—case study: forecasting salinity in a river [J]. *Journal of Hydrology*, 2005b,301(1-4):93-107.
- [10] Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, seconded[M]. New York: Springer, 2002.

Application of A New Coupled Data-driven Model in Rainfall-Runoff Simulation

LIANG Ke¹, KAN Guangyuan^{2,3}, LI Zhijia¹

(1. *College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China*; 2. *State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, Research Center on Flood & Drought Disaster Reduction of the Ministry of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing 100038, China*; 3. *State Key Laboratory of Hydrosience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing 100038, China*)

Abstract: For the purpose of overcoming the disadvantages of the traditional data-driven model and implementing the high accuracy rainfall-runoff simulation by data-driven model, a new coupled data-driven model named PEK has been proposed in this paper. The PEK model was developed by coupling partial mutual information based on the input variable selection, novel ensemble back-propagation neural network based on discharge forecasting and K-nearest neighbor based on discharge error forecasting. The PEK model has the following characteristics: (1) The separate IVS strategy and the sliding window accumulative rainfall based on model candidate input vector. These two methods combined with the PMI-based IVS approach ensure the sufficiency and parsimony of the input information and is very important to the construction of the high quality data-driven model; (2) The novel ensemble back-propagation neural network and the methodology of calibration were proposed in this paper. The global optimal number of component networks, topology and network parameters were obtained simultaneously by using the NSGA-II multi-objective optimization algorithm and the early stopping LM algorithm. The combination weights of the component networks were obtained by the AIC-based component network weights assignment approach. The EBPNN model can make a good compromise between simulation accuracy and network complexity; (3) The PEK model implemented multi-step forecasting and high accuracy simulation under the non-updating mode. The forecast period was also increased; (4) The PEK model doesn't need to compute the catchment state variables and implements continuous simulation only by using the initial discharge. In this paper, the PEK and CLS models were applied in hourly rainfall-runoff simulation in the Chengcun watershed and the results were compared. The simulation results indicate that the PEK model is easy to use, is better than CLS in simulation accuracy, can realize multi-step high accuracy simulation and increase the forecasting period of the data-driven model.

Key words: rainfall-runoff simulation; non-updating; data-driven model; PEK model; optimization method

(上接第 52 页)

Abstract: After the soil moisture sensor being installed, it is necessary to make calibration on the soil conditions in the field for verifying the rating relation. In practice, field calibration of soil moisture sensor is a complicated step, during which high quality operation is needed, on-site taking soil is difficult, and it involves working long hours. In this study, the undisturbed soil columns were collected in field to send to laboratory, the soil was taken out after saturation, let a natural loss of soil moisture, then regularly weighting was made while the original values from the soil moisture sensor were recorded, so as to study the calibration of the soil moisture sensor, and discuss the ways of improving efficiency of installing soil moisture sensors.

Key words: undisturbed soil column; soil moisture sensor; calibration method