

均生函数与 BP 神经网络耦合 模拟预测模型 (MGF-BP-I) 的建立与应用

刘丹辉¹, 马 龙¹, 刘廷玺¹, 杜志军², 王静茹¹, 黄 星¹

(1. 内蒙古农业大学水利与土木建筑工程学院, 内蒙古 呼和浩特 010018;
2. 内蒙古黄河工程局股份有限公司, 内蒙古 呼和浩特 010070)

摘 要: 水文气候因子模拟预测对气候变化研究、农业墒情预报、生态环境改善、水资源合理开发利用等具有一定参考意义。均生函数、BP 神经网络及其结合改进方式在模拟预测中各有优点, 被广泛应用, 但仍有进一步改进空间。针对 MGF、MGF-OSR、MGF-OSR-BP 等方法粗选因子集、粗选集组合筛选、收敛适应性、精度控制等可改进空间, 进一步发挥均生函数和 BP 神经网络优势, 建立了 MGF-BP-I 模拟预测模型。利用 MGF-OSR、MGF-OSR-BP、MGF-BP-I 对科尔沁沙地区域平均年降水进行了模拟预测。结果表明, 建模期 MGF-OSR-BP、MGF-BP-I 拟合效果均较好, MGF-BP-I 建模阶段最优模式精度优于 MGF-OSR-BP, MGF-BP-I 整体同时最优模式结果也非常好。检验期, MGF-BP-I 检验阶段最优及整体同时最优两种模式拟合效果最好, 相比其他模式精度有所提高。MGF-BP-I 考虑更加全面, 充分发挥了均生函数和 BP 神经网络优势, 精度远高于 MGF-OSR 和 MGF-OSR-BP, MGF-BP-I 整体同时最优模式更符合实际应用, 效果理想, 可用于水文气候因子模拟预测。

关键词: 均生函数; BP 神经网络; 耦合模拟; MGF-BP-I; 预测; 科尔沁沙地

中图分类号: P333 **文献标识码:** A **文章编号:** 1000-0852(2016)06-0007-09

1 引言

水文气候变化深刻影响着生态环境、农业发展、资源的更新变化与开发利用, 特别是干旱地区, 降水、气温的变化影响尤为直观。而对于部分寒冷地区, 冰雪变化引起的水文与农业效应日趋显著^[1]。对降水、径流等水文气候因子进行时间序列特征分析及模拟预测未来变化, 成为重要研究内容。这对水文气候变化研究、农业墒情预报、生态环境改善、水资源合理开发利用^[2-3]等都具有一定的参考意义。

目前, 对于水文气候因子时间序列进行模拟预测有时间序列^[4-5]、随机水文^[6]、奇异值分解(SVD)^[7-8]、降尺度^[9-10]、小波分析^[11-12]、均生函数^[13-14]、滑动平均-马尔科

夫模型^[15]、BP 神经网络^[16-19]等方法。提高模拟预测精度是各类模型的主要目标, 但大部分方法主要存在的仍然是精度问题。均生函数模拟预测模型^[20]优点是可改善其它模型对序列极值预测的不足, 进而提高极值拟合与预测效果, 已在水文气候预测中被广泛应用^[21]。

均生函数模拟预测模型在广泛应用的同时, 也不断被改进。一方面对其筛选预报序列的方式进行了改进, 对均生函数序列及其延拓序列, 采用以往方式筛选预报序列时, 可能存在筛选的不确定性, 有学者^[22-25]采用最优子集方法筛选预报序列, 较大程度地避免了这个缺陷, 对筛选出的最优子集序列, 建立多元线性回归预报方程, 形成了均生函数—最优子集回归 (MGF-OSR) 模型; 另外, 为了提高预报因子的全面性,

收稿日期: 2016-02-14

基金项目: 内蒙古自治区自然科学基金项目 (2014MS0407, 2010BS0608); 国家自然科学基金项目 (50869005)

作者简介: 刘丹辉 (1989-), 女, 内蒙古赤峰人, 硕士研究生, 主要从事水文及水资源, 环境演变、气候变化及二者响应关系研究。

E-mail: 1253642610@qq.com

通讯作者: 马龙 (1978-), 男, 内蒙古呼和浩特人, 教授, 博士, 主要从事水文及水资源, 环境演变、气候变化及二者响应关系研究。

E-mail: malong4444333@163.com

有学者^[26]在 MGF-OSR 模型的基础上,将均生函数序列与除预报因子外的其他影响因子序列相结合作为初选序列。另一方面,对筛选预报序列方式和预测方法同时进行了改进,有学者^[27-28]提出了主成分筛选序列的方法,对生成的均生函数序列利用主成分方法提取对因变量影响较强的序列,而 BP 神经网络具有通用性好、非线性映射能力强等优点,得到了广泛应用^[29-32],将筛选出的序列作为 BP 神经网络的输入,进而得到预测模型,其预报精度高于单一均生函数模拟预测模型。此外,也有学者^[33]将均生函数—最优子集回归方法和 BP 神经网络模型仅在筛选出预报因子后结合,以最优子集序列作为 BP 神经网络的输入,形成了模拟精度亦较高的 MGF-OSR-BP 预测模型。

由此可见,通过对均生函数与 BP 神经网络的结合,为神经网络提供了较好的输入基础,亦充分发挥了神经网络的模拟优势,大大提高了模拟与预测的精度。但仍有可改进之处,可进一步提高模拟预测精度。首先在均生函数序列及其延拓序列粗筛选方面,仍有部分较好的序列没有进入粗筛选后的序列中;另外,大多数方法是在确定组合序列的情况下,再使用 BP 神经网络进行参数优化,这就使得有可能没有选出与 BP 神经网络适宜的、有效结合的组合,没有充分发挥 BP 神经网络方法的优势。

因此,本次以均生函数为基础,利用 CSC 值与相关系数结合方式对均生函数及其延拓序列进行粗筛选,并将 BP 神经网络与均生函数在组合粗筛选、系数进一步优化进行耦合,形成了一种新的精度和效率均较高的模拟预测模型(MGF-BP-I),并以科尔沁沙地降水为例,进行了应用。

2 MGF-BP-I 模拟预测模型的建立

2.1 均生函数模拟预测模型的原理

均生函数模拟预测模型是基于系统状态前后记忆思想,由时间序列按照不同时间间隔计算均值,构建一组周期函数及其延拓序列,然后用原序列与这组函数建立回归预测方程,原理为^[20]:

(1) 设原始的时间序列为:

$$x(t) = \{x(1), x(2), \dots, x(N)\} \quad (1)$$

式中: $t=1, 2, \dots, N$, N 为样本的大小。

为了拟合原序列中的高频部分,对原序列进行差分运算。这一运算实际上起着高通滤波的作用。利用

式(2)、(3)分别计算原始时间序列的一阶差分序列和二阶差分序列:

$$\Delta x(t) = x(t+1) - x(t) \quad (2)$$

式中: $t=1, 2, \dots, N-1$ 。

$$\Delta \Delta x(t) = \Delta^2 x(t) = \Delta x(t+1) - \Delta x(t) \quad (3)$$

式中: $t=1, 2, \dots, N-2$ 。

(2) 利用式(4)对原始时间序列、一阶差分序列和二阶差分序列做均值生成函数:

$$\bar{x}_i(i) = \frac{1}{n_i} \sum_{j=0}^{n_i-1} x(i+jl) \quad (4)$$

式中: $i=1, 2, \dots, l$, $1 \leq l \leq M$ 且 $n_i = \text{INT}(N/l)$, M 一般视样本数的大小取 $\text{INT}(N/3)$ 。

(3) 利用式(5)对一阶、二阶差分均生函数序列以及原序列均生函数序列作周期性延拓计算:

$$f_i(t) = \bar{x}_i[t - \text{INT}(\frac{t-1}{l})] \quad (5)$$

式中: $t=1, 2, \dots, N$ 。

(4) 为了拟合时间序列中向上递增和向下递减的趋势,进一步利用式(6)计算累加延拓序列,得到累加延拓序列矩阵:

$$f_i^{(3)}(t) = X(1) + \sum_{i=1}^{t-1} f_i^{(1)}(i+1) \quad (6)$$

式中: $t=2, 3, \dots, N$ 。

(5) 对 p 个周期延拓和累加延拓序列作 q 步预报。

2.2 BP 神经网络

本次采用三层 BP 神经网络模型,包括输入层、输出层和隐含层。其中,节点的作用函数采用 Sigmoid 型函数,其数学模型为^[16]:

$$f(x) = 1/(1+e^{-x}) \quad (7)$$

2.3 CSC 双评分准则原理

双评分准则(Couple Score Criterion, CSC)^[20]是以数量预报评分和趋势预报评分来权衡变量优劣,旨在使数量评分和趋势评分之和达到最大。用 S_1 表示数量评分,称为精评分; S_2 表示趋势评分,即粗评分。那么双评分准则表示为:

$$CSC = S_1 + S_2 \quad (8)$$

$$S_1 = nR^2 \quad (9)$$

$$S_2 = 2I = 2 \left[\sum_{i=1}^I \sum_{j=1}^I n_{ij} \ln n_{ij} + n \ln n - \left(\sum_{i=1}^I n_i \ln n_i + \sum_{j=1}^I n_j \ln n_j \right) \right] \quad (10)$$

式中: n 为样本数; R 为复相关系数; I 为趋势类别数; n_{ij} 为 i 类事件与 j 类估计事件列联表的个数。

2.4 均生函数-最优子集-BP神经网络 (MGF-OSR-BP)模拟模型预测过程

MGF-OSR-BP^[33]模拟模型预测具体过程如图1所示,步骤如下:

步骤1,通过最优子集回归方法对得到的均生函数及其延拓序列进行筛选,选出最优子集。

步骤2,将筛选出的最优子集作为输入训练集,利用BP神经网络对组合序列参数进行优化训练,得到BP神经网络模拟预测模型。

步骤3,利用优化参数后的BP神经网络模拟预测模型进行预测。

2.5 MGF、MGF-OSR、MGF-OSR-BP 模拟预测模型可改进空间

不论是均生函数(MGF),还是均生函数-最优子集回归(MGF-OSR),还是均生函数-最优子集-BP神经网络(MGF-OSR-BP),都存在可以进一步改进的空间。

第一,在使用均生函数及其延拓序列粗选因子集方面,使用了单一的方法,即CSC值较大的(在一定置

信区间)进入粗选集序列,而部分相关系数较大的序列没有进入,造成形成组合的粗选因子集不够全面。第二,对于粗选集序列组合的筛选,使用最优回归子集及CSC收敛准则选出最终使用的组合序列,按照CSC的定义和计算方式,可能造成真正模拟和检验均好的组合序列被遗漏。第三,将均生函数与BP神经网络进行结合,其间也有组合对方法收敛存在一定的适应性,前述方法只考虑了CSC值,没有考虑适应性。第四,各类模拟预测模型的目标就是最终在模拟和检验阶段的相对误差或绝对误差达到最小,而前述方法在计算过程中,均使用CSC值作为精度控制条件,这就遗漏了部分精度真正较高的组合序列。

2.6 均生函数-神经网络-改进(MGF-BP-I)模拟预测模型过程

针对均生函数、BP神经网络自身的优点及可进一步改进的空间与结合方式,建立了均生函数与BP神经网络耦合模拟预测模型MGF-BP-I,方法过程如图1所示,具体步骤如下:

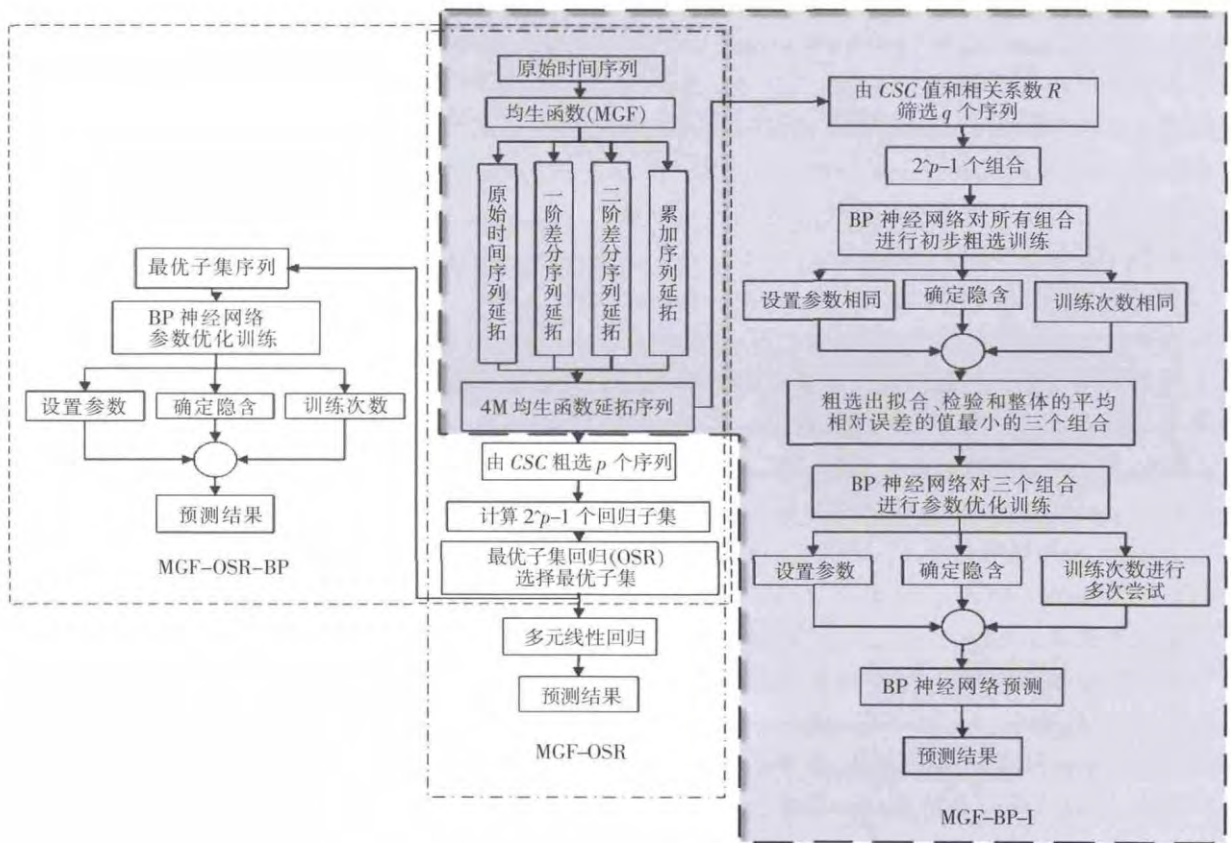


图1 MGF-OSR、MGF-OSR-BP和MGF-BP-I模拟预测模型流程 Fig.1 The simulation model flow of MGF-OSR, MGF-OSR-BP and MGF-BP-I

步骤 1,使用均生函数方法步骤(1)~(4)对原始时间序列进行延拓,得到均生函数及其延拓序列。

步骤 2,通过 CSC 值和相关系数两个参数按照一定的置信区间,对步骤 1 得到的均生函数及其延拓序列进行筛选,得出粗选因子集。

步骤 3,对步骤 2 得到的粗选因子集序列进行排列组合,形成 $2^q - 1$ 个组合。每一组合序列作为输入训练集,利用 BP 神经网络对所有组合进行初步粗选训练。训练时参数设置相同,得到每一组合下的模拟预测值。

步骤 4,对步骤 3 中各组合建模阶段和检验阶段的模拟预测值,分别与原序列计算平均相对误差 ($MAPE$),最后分别筛选出建模、检验和整体阶段平均相对误差最小的三个组合序列。

步骤 5,对步骤 4 得到的三个组合序列,再次利用 BP 神经网络对其分别进行优化训练,参数的设置与步骤 4 类似,只对训练次数进行调整,直到三个组合序列的 $MAPE$ 值分别达到最小值为止。

步骤 6,利用所建立的三种组合 BP 神经网络模拟预测模型进行预测。

3 模型的应用

应用 C 语言对 MGF-OSR、MGF-OSR-BP、MGF-BP-I 模拟预测模型过程进行了编程调试,应用于科尔沁沙地区域平均年降水的模拟预测中,并对三种方法计算结果进行了对比分析。

3.1 使用的数据及模拟阶段划分

科尔沁沙地及其周边地区分布的气象站点见图 2,共计分布有 27 个气象站点。本次收集了这些站点自建站以来至 2012 年的降水逐月数据,数据来源于气象局整编资料。经检验,各气象站点降水数据无明显的突变点和随机变化,数据变化相对均一和一致,数据可靠,可以代表区域降水状况。个别站点个别数据缺测,进行了插补,最终形成了各站点 1951~2012 年降水逐月数据序列。

对于科尔沁沙地区域平均年降水序列的求取,采用客观运行法,即将区域分成网格,形成格点,用研究区内外各气象站的降水数据,采用邻近各站权重(取距离因子)及综合权重确定各格点降水数据,取各格点数据的算术平均值,即为整个区域降水平均值序列。本次模拟预测采用的就是 1951~2012 年科尔沁沙地区域平均年降水序列。

模拟预测中,1951~2007 年为建模阶段,2008~2012 年为检验阶段,2013~2017 年为预测阶段。

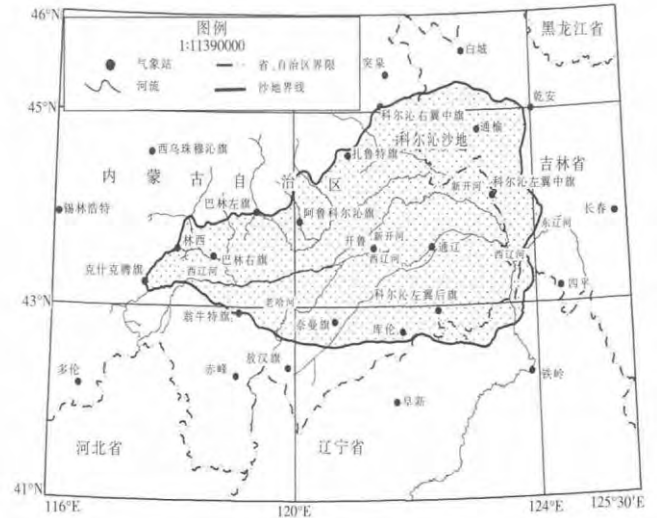


图 2 科尔沁沙地位置及气象站点分布
Fig.2 The position of Horqin sandy land and distribution of the meteorological stations

3.2 结果与分析

(1)MGF-OSR 方法过程

将科尔沁沙地 1951~2007 年区域平均年降水数据作为原始序列,按照均生函数步骤(1)~(4)得到 76 个序列。设定显著性水平 $\alpha=0.01$, $\chi_{0.01}^5=15.086$,将满足 $CSC>15.086$ 的序列粗选为预报因子,共计选出 15 个均生函数及其延拓序列,分别为 $f_4^{(0)}$ 、 $f_5^{(0)}$ 、 $f_6^{(0)}$ 、 $f_{12}^{(0)}$ 、 $f_{13}^{(0)}$ 、 $f_{14}^{(0)}$ 、 $f_{15}^{(0)}$ 、 $f_{16}^{(0)}$ 、 $f_{19}^{(0)}$ 、 $f_{12}^{(1)}$ 、 $f_{13}^{(1)}$ 、 $f_{17}^{(1)}$ 、 $f_{18}^{(1)}$ 、 $f_{19}^{(1)}$ 、 $f_{17}^{(2)}$,形成 524 287 个回归子集,选出 CSC 值最大的回归子集,做为预报方程:
$$y=-101.314-0.1366f_{12}^{(0)}-0.7279f_{17}^{(1)}+0.0443f_{12}^{(1)}+0.7573f_{13}^{(0)}+0.4812f_{19}^{(0)}+0.3964f_{16}^{(0)}-0.1655f_{17}^{(2)}+0.368f_{15}^{(0)}-0.2315f_{18}^{(1)}+0.1158f_{19}^{(1)}+0.2660f_{13}^{(1)}+0.5970f_4^{(0)}+0.7681f_5^{(0)}+0.3527f_6^{(0)}-0.0406f_{14}^{(0)}$$

(2)MGF-OSR-BP 方法过程

根据 MGF-OSR-BP 方法步骤(1)~(4),组合序列同上,BP 神经网络的输入层为 15,隐含层为 9,输出层为 1,学习速率为 0.9,动量因子为 0.7,误差的精度设置为 0.001,训练次数为 23 490 次后模拟值达到最优。

(3)MGF-BP-I 方法过程

根据 MGF-BP-I 方法步骤 (1)~(2), 计算得到

CSC>15.086 且 $R>0.34$ ($\alpha=0.01$) 的序列作为粗选序列, 共计 19 个序列。输入层为 1~19, 输出层为 1, 隐含层 N 根据经验公式来确定。学习速率为 0.9, 动量因子为 0.7, 误差的精度设置为 0.001, 训练次数为 2 000 次。选出建模阶段、检验阶段和整体阶段平均相对误差最小的三个组合序列。MGF-BP-I 建模最优模式最终选取的组合序列为 $f_{17}^{(0)}$ 、 $f_{12}^{(0)}$ 、 $f_{12}^{(1)}$ 、 $f_{16}^{(0)}$ 、 $f_{17}^{(2)}$ 、 $f_{15}^{(0)}$ 、 $f_{18}^{(1)}$ 、 $f_{13}^{(2)}$ 、 $f_{19}^{(1)}$ 、 $f_{13}^{(1)}$ 、 $f_4^{(0)}$ 、 $f_6^{(0)}$ 、 $f_{17}^{(3)}$ 、 $f_7^{(0)}$ 、 $f_{14}^{(0)}$, MGF-BP-I 检验最优模式最终选取的组合序列为 $f_{12}^{(1)}$ 、 $f_{17}^{(2)}$ 、 $f_{15}^{(0)}$ 、 $f_{13}^{(2)}$ 、 $f_{19}^{(1)}$ 、 $f_{13}^{(1)}$ 、 $f_7^{(0)}$, MGF-BP-I 整体最优模式最终选取的组合序列为 $f_{12}^{(1)}$ 、 $f_{15}^{(0)}$ 、 $f_{13}^{(2)}$ 、 $f_{19}^{(1)}$ 、 $f_{13}^{(1)}$ 、 $f_6^{(0)}$ 、 $f_{17}^{(3)}$ 、 $f_7^{(0)}$ 、 $f_{14}^{(0)}$ 。对这三个组合利用 BP 神经网络进行进一步的参数优化后, 得到最终的模拟预测模型及序列。

(4)结果分析

MGF-OSR、MGF-OSR-BP、MGF-BP-I 三种方法计算结果相对误差分布见图 3, 建模期模拟值和实测值的对比见图 4。其中 MGF-BP-I 又分建模阶段最优、检验阶段最优、整体同时最优三种方式, 三种方法共计五种计算模式。

对于建模阶段, 由图 3 发现, MGF-OSR、MGF-BP-I 检验阶段最优模式 90% 的年数相对误差分别集中于 5%~15% 和 0.5%~5% 之间, 误差相对较高。MGF-

BP-I 整体同时最优、MGF-OSR-BP、MGF-BP-I 建模阶段最优模式相对误差分别主要集中在 0~1%、0~0.05%、0~0.005% 的年数占建模阶段总年数的 94%、94%、95%, MGF-BP-I 建模阶段最优模式相对误差分布集中且精度高, 总体说明建模阶段相对误差最好的是 MGF-BP-I 建模阶段最优模式, 而 MGF-OSR-BP 和 MGF-BP-I 整体同时最优模式相对误差也较小, 均可以用于建模。

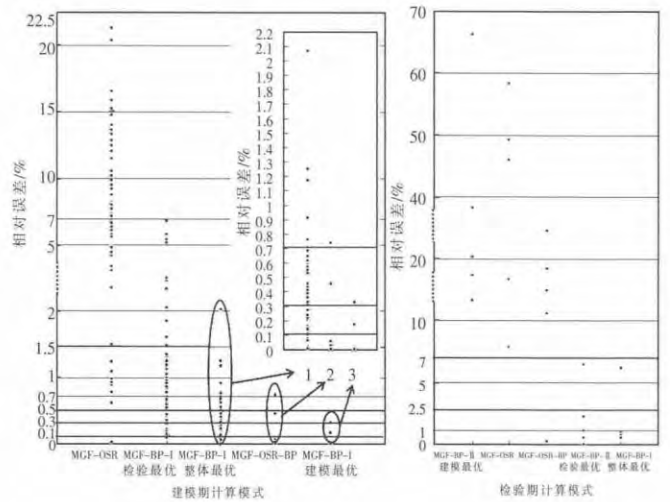


图 3 三种模拟模型建模阶段和检验阶段的相对误差分布图
Fig.3 Relative error distribution of the three kinds of simulation models for simulation and testing phase

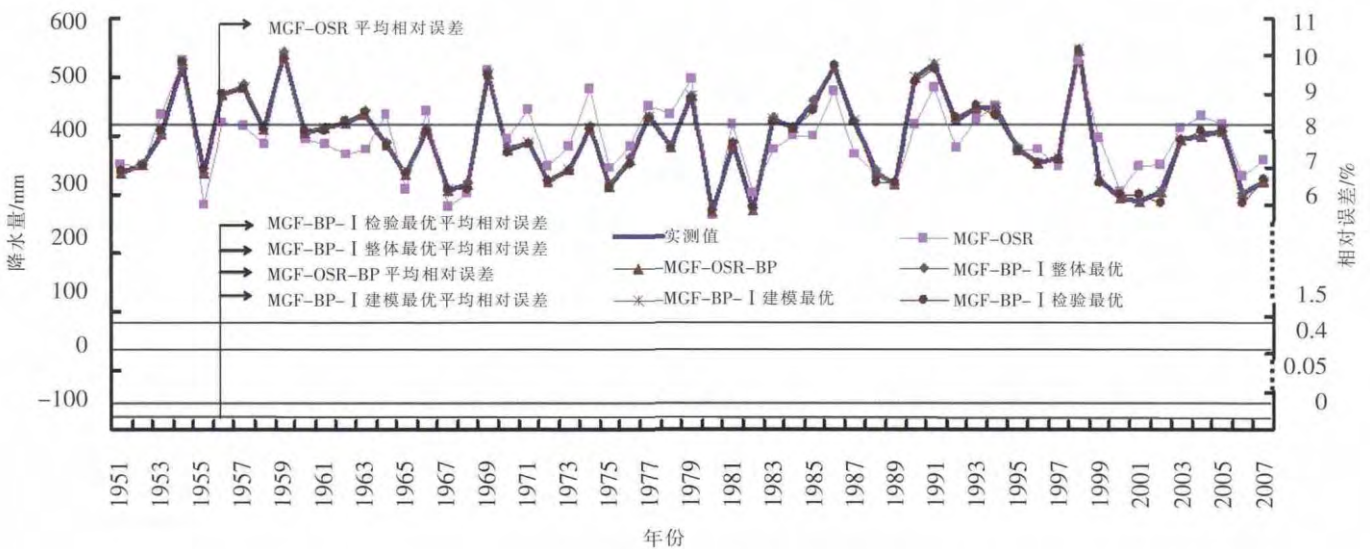


图 4 建模期实测值和模拟值及平均相对误差对比图
Fig.4 The measured values and simulated values and relative error comparison

由图4可以看出,除MGF-OSR方法拟合效果不太理想外,其他四种计算模式拟合效果均较好。MGF-OSR、MGF-OSR-BP、MGF-BP-I建模阶段最优、MGF-BP-I检验阶段最优、MGF-BP-I整体同时最优五种模式建模阶段的平均相对误差分别为8.1407%、0.0325%、0.0099%、1.3196%、0.4220%,五种模式建模序列的均方根误差(MSE)分别为37.6170、0.4275、0.1680、6.5476、2.1225,平均绝对误差(MAE)分别为31.54mm、0.12mm、0.04mm、4.92mm、1.65mm,建模期合格率^[34]分别为96%、100%、100%、100%和100%。MGF-BP-I建模阶段最优模式优于MGF-OSR-BP模式;而MGF-BP-I整体同时最优模式结果也非常好。此外,MGF-OSR、MGF-OSR-BP方法均以CSC值判定

模拟的好坏,CSC值越大,模拟越好。从这个角度,五种模式的CSC值分别为94.89、170.388、170.39、139.34、170.35,MGF-BP-I的CSC值都远远高于MGF-OSR,而MGF-BP-I建模阶段最优的CSC值优于MGF-OSR-BP,MGF-BP-I整体同时最优的CSC值接近于MGF-OSR-BP,可以说明MGF-BP-I方法优势明显。

对于检验阶段,由图3可以看出,MGF-OSR模式4年的相对误差集中于10%~60%之间。MGF-BP-I建模阶段最优、MGF-OSR-BP模式4年的相对误差集中于10%~40%之间。MGF-BP-I检验阶段最优模型4年的相对误差集中于0~2.3%,最大值为6.7%。MGF-BP-I整体同时最优模式4年的相对误差集中于0~1%之间,最大值为6.5%。

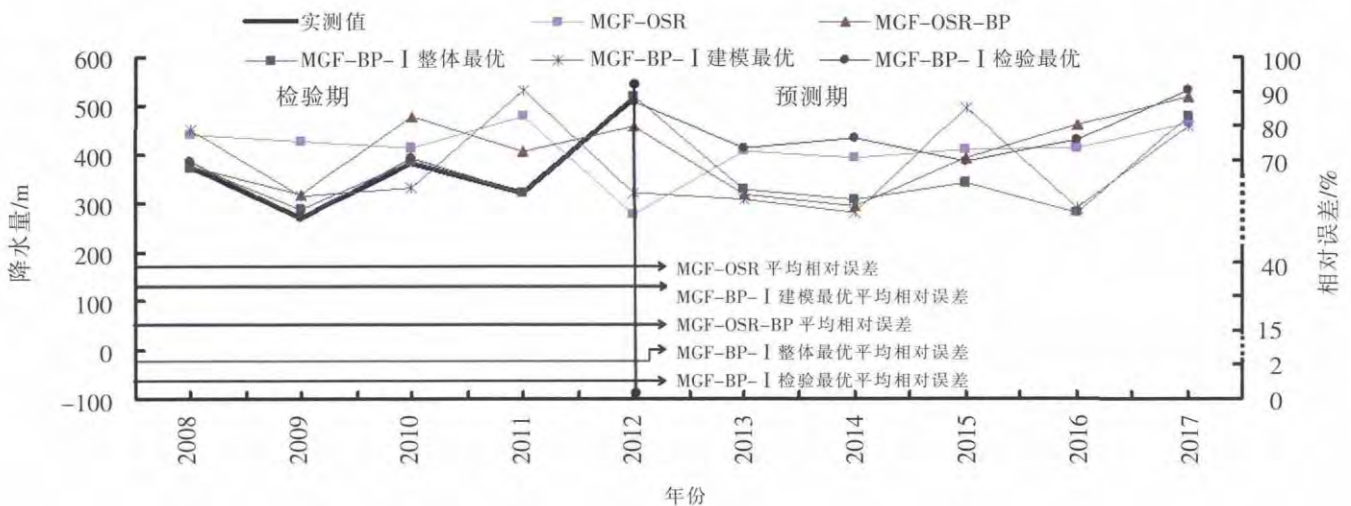


图5 检验期实测值与模拟值、平均相对误差对比及预测图

Fig.5 Comparison between the measured and simulated values, relative error and forecast chart in the inspection period

检验期实测值与预测值对比见图5。检验阶段五种模式平均相对误差分别为35.5473%、16.2509%、30.9184%、1.3262%、1.8652%,均方根误差分别为148.6060、66.5610、136.3810、6.5123、8.1918,平均绝对误差分别为128.71mm、57.75mm、115.85mm、5.22mm、5.73mm,CSC值分别为73.14、135.42、135.50、167.23、160.41,检验期合格率^[34]分别为40%、80%、80%、100%和100%。可以看出,检验阶段MGF-BP-I检验阶段最优、MGF-BP-I整体同时最优两种模式拟合结果最好,相比其他三种模式结果,精度有所提高。

综上所述,建模阶段精度最高的是MGF-BP-I建模阶段最优模式,而检验阶段是MGF-BP-I检验阶段

最优和MGF-BP-I整体同时最优模式。从整体看,MGF-BP-I整体同时最优模式建模阶段和检验阶段整体相对误差均较小,完全符合《水文情报预报规范》^[34]对中长期序列预报的评定要求,在实际应用中选用其更符合实际。

使用五种模式对研究区2013~2017年5年的降水进行预测,结果见图5所示。可以看出,其他模式预测结果均高于MGF-BP-I整体同时最优模式10~180mm左右。以MGF-BP-I整体同时最优模式结果为准可知,2015年和未来2017年降水相对丰富,其他为平水年份,这与已经过去的2013、2014和2015年的实际相符。

4 结论与讨论

4.1 结论

(1)针对均生函数(MGF)、均生函数—最优子集回归(MGF-OSR)和均生函数—最优子集—BP神经网络(MGF-OSR-BP)中均生函数及其延拓序列粗选因子集、粗选集组合筛选、收敛适应性、精度控制等可以进一步改进的方面,充分发挥均生函数与BP神经网络各自优势,建立了MGF-BP-I模拟预测模型。

(2)通过应用对比发现,建模期除MGF-OSR方法拟合效果不太理想外,MGF-OSR-BP、MGF-BP-I两种方法四种模式拟合效果均较好。MGF-BP-I建模阶段最优模式优于MGF-OSR-BP模式;而MGF-BP-I整体同时最优模式结果也非常好。检验阶段,MGF-BP-I检验阶段最优、MGF-BP-I整体同时最优两种模式拟合结果最好,相比其他三种模式结果,精度有所提高。

(3)MGF-BP-I模拟预测模型不论是拟合阶段最优模式,还是检验阶段最优模式,其精度均远远高于MGF-OSR和MGF-OSR-BP方法,MGF-BP-I整体同时最优模式两阶段结果精度均较高,在实际应用中更符合实际。

(4)通过预测发现,除个别年份和个别模式外,其他模式预测结果均高于MGF-BP-I整体同时最优模式,偏高10~180mm左右。2015年和未来2017年降水相对丰富,其他为平水年份,与已经过去的年份相符。

(5)总体来看,MGF-BP-I方法考虑更全面,并充分发挥了均生函数和神经网络的优势,效果理想,可以用于水文气候因子的模拟预测。

4.2 讨论

本文利用MGF-BP-I模型对科尔沁沙地区域平均年降水进行了模拟和预测,拟合阶段最优模式、检验阶段最优模式精度均远远高于MGF-OSR和MGF-OSR-BP方法,MGF-BP-I整体同时最优模式两阶段结果精度均较高。对比其他相关研究^[33],其MGF-OSR-BP模型建模期平均相对误差为0.23%,检验期为4.45%;MGF-OSR模型建模期为6.10%,检验期为13.75%;均生函数逐步回归模型建模期为6.68%,检验期为17.6%。另有与MGF-OSR-BP基本类似改进方法^[27]研究结果表明,建模期平均相对误差为1.45%,检验期为5.65%;BP神经网络模型建模期为1.40%,检

验期为7.30%;均生函数建模期为4.65%,检验期为12.08%。本次所建立的MGF-BP-I模型,各阶段精度均远远高于相关研究结果精度。我们也尝试利用单独均生函数序列(模式1)和利用均生函数及其一二阶差分序列(模式2)两种模式进行模拟预测,可以很明显的看出,模式1模拟结果差于模式2,模式2很好的拟合了原序列中的高频部分。

虽然MGF-BP-I模型实现了较高的模拟预测精度,但其后一阶段(即BP嵌入耦合模拟阶段)中的BP算法本质为梯度下降法,降低了算法学习效率,加之组合的增多与BP的多次耦合嵌入,其运行效率相比MGF-OSR和MGF-OSR-BP有所降低。此外,模型后一阶段是半黑箱式模型,模型参数较多,但是没有具体物理意义,模型中一系列参数的选取,主要依靠经验和试算,同样对精度和效率产生一定影响。这些方面对于精度和效率的提升均有一定改进空间,需要今后研究中深入考虑这些因素,进一步改进模型,提高精度和效率。

参考文献:

- [1] 沈永平,苏宏超,王国亚,等. 新疆冰川、积雪对气候变化的响应(): 水文效应[J]. 冰川冻土, 2013,35(3):513-527. (SHEN Yongping, SU Hongchao, WANG Guoya, et al. The responses of glaciers and snow cover to climate change in Xinjiang (I): hydrological effects[J]. Journal of Glaciology and Geocryology, 2013,35 (3):513-527. (in Chinese))
- [2] Toda Y, Abe F. Prediction of precipitation sequences within grains in 18Cr-8Ni austenitic steel by using system free energy method[J]. ISIJ International, 2009,49(3):439-445.
- [3] 谢炯光,曾琼,纪忠萍. 中国近30年来气象统计预报进展[J]. 气象科技, 2003,31(2):67-79. (XIE Jiongguang, ZENG Cong, JI Zhongping. Statistical forecast of meteorology for the last 30 years in China[J]. Meteorological Science and Technology, 2003,31(2):67-79. (in Chinese))
- [4] 桑燕芳,王中根,刘昌明,等. 水文时间序列分析方法研究进展[J]. 地理科学进展, 2013,32(1):20-30. (SANG Yanfang, WANG Zhonggen, LIU Changming, et al. Research progress on the time series analysis methods in hydrology[J]. Progress in Geography, 2013,32(1): 20-30. (in Chinese))
- [5] 吕志涛. 时间序列分析方法在郑州市降水量预报中的应用[J]. 南水北调与水利科技, 2014,12 (4):35-37. (LV Zhitao. Application of time series analysis on the annual precipitation of Zhengzhou city[J]. South-to-North Water Transfers and Water Science & Technology, 2014,12(4):35-37. (in Chinese))
- [6] 陈桂英. 我国现有短期气候业务预测方法综述 [J]. 应用气象学报, 2000,11(Z1):11-20. (CHEN Guiying. Review of the methods for operational short-range climate prediction in China [J]. Quarterly Journal of Applied Meteorology, 2000,11(Z1):11-20. (in Chinese))

- [7] 程慧萍,贾晓静. 印度洋冬季海温异常对中国春季降水的影响[J]. 高原气象, 2014,33(3):733-742. (CHENG Huiping, JIA Xiaojing. Influence of sea surface temperature anomalies of Indian Ocean in winter on precipitation over China in spring[J]. Plateau Meteorology, 2014,33(3):733-742. (in Chinese))
- [8] 刘抗,李照荣,张宇. 1971-2010年中国西北地区秋季降水变化特征[J]. 中国沙漠, 2014,34(3):805-813. (LIU Kang, LI Zhaorong, ZHANG Yu. Variation of autumn precipitation in northwest China in 1971-2010[J]. Journal of Desert Research, 2014,34(3):805-813. (in Chinese))
- [9] 李毅,周牡丹. 新疆地区棉花和甜菜需水量的统计降尺度模型预测[J]. 农业工程学报, 2014,30(22):70-79. (LI Yi, ZHOU Mudan. Projections of water requirements of cotton and sugar beet in Xinjiang downscaling model [J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014,30(22):70-79. (in Chinese))
- [10] 刘永和,郭维栋,冯锦明,等. 气象资料的统计降尺度方法综述[J]. 地球科学进展, 2011,26(8):837-847. (LIU Yonghe, GUO Weidong, FENG Jinming, et al. A summary of methods for statistical downscaling of meteorological data [J]. Advances in Earth Science, 2011,26(8):837-847. (in Chinese))
- [11] 李淼,夏军,陈社明,等. 北京地区近300年降水变化的小波分析[J]. 自然资源学报, 2011,26(6):1001-1011. (LI Miao, XIA Jun, CHEN Sheming, et al. Wavelet analysis on annual precipitation around 300 years in Beijing area[J]. Journal of Natural Resources, 2011,26(6):1001-1011. (in Chinese))
- [12] 桑燕芳,王中根,刘昌明. 小波分析方法在水文学研究中的应用现状及展望[J]. 地理科学进展, 2013,32(9):1413-1422. (SANG Yanfang, WANG Zhonggen, LIU Changming. Applications of wavelet analysis to hydrology: status and prospects [J]. Progress in Geography, 2013,32(9):1413-1422. (in Chinese))
- [13] 刘洪兰,白虎志,张俊国. 河西走廊中部近53年降水变化及未来趋势预测[J]. 干旱区研究, 2011,28(1):146-150. (LIU Honglan, BAI Huzhi, ZHANG Junguo. Change and trend prediction of precipitation in the central Hexi corridor in Gansu province in recent 53 years[J]. Arid Zone Research, 2011,28(1):146-150. (in Chinese))
- [14] 黄颖,金龙. 华南前汛期降水预测模型及其预测试验[J]. 热带气象学报, 2011,27(5):753-757. (HUANG Ying, JIN Long. Prediction model for annually first rainy season precipitation in south china and prediction tests [J]. Journal of Tropical Meteorology, 2011,27(5):753-757. (in Chinese))
- [15] 孙鹏,张强,白云岗,等. 基于马尔科夫模型的新疆水文气象干旱研究[J]. 地理研究, 2014,33(9):1647-1657. (SUN Peng, ZHANG Qiang, BAI Yungang, et al. Transitional behaviors of hydrometeorological droughts in Xinjiang using the Markov chain model[J]. Geographical Research, 2014,33(9):1647-1657. (in Chinese))
- [16] 金龙. 神经网络气象预报建模理论与应用 [M]. 北京: 气象出版社, 2004:75-80. (JIN Long. Theory Method and Application of Neural Network Weather Forecast [M]. Beijing: China Meteorological Press, 2004:75-80. (in Chinese))
- [17] 农孟松,黄海洪,孙崇智,等. 基于主分量神经网络的降水集成预报方法研究 [J]. 气象, 2011,37(3):354-355. (NONG Mengsong, HUANG Haihong, SUN Chongzhi, et al. A neural network model based on principal component analysis for ensemble precipitation prediction[J]. Meteorological Monthly, 2011,37(3):352-355. (in Chinese))
- [18] Bashir Z A, El-Hawary M, Applying E. Wavelets to short-term load forecasting using PSO-based neural networks [J]. Ieee Transactions on Power Systems, 2009,24(1):20-27.
- [19] Li L Y, Chen Y, Xu T B. Super-resolution mapping of wetland inundation from remote sensing imagery based on integration of back-propagation neural network and genetic algorithm[J]. Remote Sensing of Environment, 2015,164:142-154.
- [20] 魏凤英. 现代气候统计诊断与预测技术 [M]. 北京: 气象出版社, 1999:239-247. (WEI Fengying. Modern Climate Statistical Diagnosis and Prediction Technology [M]. Beijing: China Meteorological Press, 1999:239-247. (in Chinese))
- [21] 郭建平,陈玥煜. 新疆棉花热量指数的均生函数预测模型[J]. 干旱区资源与环境, 2010,24(8):175-179. (GUO Jianping, CHEN Yueyi. Mean generating function model for cotton heat index forecasting in Xinjiang [J]. Journal of Arid Land Resources and Environment, 2010,24(8):175-179. (in Chinese))
- [22] 马龙,刘廷玺,冀鸿兰,等. 基于气候重建资料及均生函数—最优子集回归模型的降水预测[J]. 水文, 2013,33(1):63-67. (MA Long, LIU Tingxi, JI Honglan, et al. Precipitation short-term forecasting in the Khorchin sand based on climate reconstruction data and mean generating function—optimal subset regression (MGF-OSR) prediction model[J]. Journal of China Hydrology, 2013,33(1):63-67. (in Chinese))
- [23] 窦浩洋,邓航,孙小明,等. 基于均生函数最优子集回归预测模型的青藏高原气温和降水短期预测 [J]. 北京大学学报(自然科学版), 2010,46(4):643-648. (DOU Haoyang, DENG Hang, SUN Xiaoming, et al. Short-term temperature and precipitation forecast over Tibetan plateau using mean generating function—optimal subset regression[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2010,46(4):643-648. (in Chinese))
- [24] 赵雪花,陈旭. 经验模态分解与均生函数—最优子集耦合模型在年径流预测中的应用[J]. 资源科学, 2015,37(6):1173-1180. (ZHAO Xuehua, CHEN Xu. Application of empirical mode decomposition and mean generating function—optimum subset regression hybrid model for annual runoff forecasting[J]. Resources Science, 2015,37(6):1173-1180. (in Chinese))
- [25] 文绮新,孙国栋,张传江,等. 基于均生函数的最优子集回归方法在江西雨季降水预测中的应用[J]. 灾害学, 2000,15(2):11-16. (WEN Qixin, SUN Guodong, ZHANG Chuanjiang, et al. Optimal subset regression method based on mean-value generation function and

- its application in precipitation forecast in rainy season in Jiangxi province[J]. Journal of Catastrophology, 2000,15(2):11-16. (in Chinese))
- [26] 何小娟. 用多因子筛选和均生函数序列双重最优子集作降水量预报[J]. 广西气象, 2002,23(4):39-42. (HE Xiaojuan. Establishing the precipitation forecast by utilizing the double optimized subsets of multi-factors screening and mean-value generated function serial[J]. Journal of Guangxi Meteorology, 2002,23(4):39-42. (in Chinese))
- [27] 李阳,纪昌明,李克飞,等. 结合均生函数的神经网络在中长期水文预报中的应用[J]. 水电能源科学, 2013,31(2):19-22. (LI Yang, JI Changming, LI Kefei, et al. Application of neural network combined with mean generation function to mid long term hydrological forecasting [J]. Water Resources and Power, 2013,31(2):19-22. (in Chinese))
- [28] 农吉夫,金龙. 基于MATLAB的主成分RBF神经网络降水预报模型[J]. 热带气象学报, 2008,24(6):713-717. (NONG Jifu, JIN Long. The rainfall forecast model of PCA-RBF neural networks based on MATLAB[J]. Journal of Tropical Meteorology, 2008,24(6):713-717. (in Chinese))
- [29] 余世鹏,杨劲松,刘广明,等. 基于模糊神经算法的区域地下水盐分动态预测[J]. 农业工程学报, 2014,30(18):142-150. (YU Shipeng, YANG Jingsong, LIU Guangming, et al. Regional groundwater salinity dynamics forecasting based on neuro-fuzzy algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014,30(18):142-150. (in Chinese))
- [30] Nguyen Ngoc Son, Ho Pham Huy Anh. Adaptive displacement online control of shape memory alloys actuator based on neural networks and hybrid differential evolution algorithm [J]. Neurocomputing, 2015,166:464-474.
- [31] Zhang J K, Ji M, Zhang Y. Tourism sustainability in Tibet - forward planning using a systems approach [J]. Ecological Indicators, 2015,56:218-228.
- [32] Guo X X, Liebgott Herve, Friboulet Denis. Back-propagation beam former design for motion estimation in echocardiography[J]. Ultrasonic Imaging, 2015,37(3):179-204.
- [33] 马利,李雪莲,张波,等. 一种基于时间序列的预测方法[P]. 中国专利: CN 102495937 A. 2012. (MA Li, LI Xuelian, ZHANG Bo, et al. Prediction Method Based on Time Series[P]. CN 102495937 A. 2012. (in Chinese))
- [34] GB/T 22482-2008, 水文情报预报规范 [S]. (GB/T 22482-2008, Hydrological Information Forecasting Specification[S]. (in Chinese))

A Coupled Simulation and Forecast Model of Mean Generating Function and BP Neural Network (MGF-BP-I)

LIU Danhui¹, MA Long¹, LIU Tingxi¹, DU Zhijun², WANG Jingru¹, HUANG Xing¹

(1. College of Water Conservancy and Civil Engineering College, Inner Mongolia Agricultural University, Hohhot 010018, China;

2. Inner Mongolia Yellow Rive Construction Co.,Ltd., Hohhot 010070, China)

Abstract: Simulation and prediction of hydrological and climate factors is very significant for climate change research, soil moisture forecasting, ecological environment improvement, reasonable development and utilization of water resources. Mean generating function method, BP neural network method and their combination are widely used in simulation and prediction. Each of these methods have their own advantages, but there is still room for further improvement. As for rough selection of factor set, selection of the factor set combination and accuracy control of MGF, MGF-OSR, MGF-OSR-BP, a simulation and forecast model MGF-BP-I was built for taking the advantages of mean generating function and BP neural network. The mean annual precipitation in the Horqin Sandy Land was simulated and forecasted by using MGF, MGF-OSR-BP, MGF-BP-I. The results show that, in the modeling period, MGF-BP-I and MGF-OSR-BP have better fitting effect, optimization mode accuracy of MGF-BP-I is better than that of MGF-OSR-BP, and global optimization mode of MGF-BP-I is very good. In the verification period, MGF-BP-I verification phase optimization and MGF-BP-I global optimization mode of simulation results are best. MGF-BP-I takes the advantages of mean generating function and BP neural network, its accuracy is much higher than those of MGF-OSR and MGF-OSR-BP. MGF-BP-I global optimization model is more consistent with the practical application, the effect is ideal, can be used in the simulation and forecast of hydrological and climate factors.

Key words: mean generating function; BP neural network; coupled simulation; MGF-BP-I; prediction; Horqin Sandy Land