

doi:10.11835/j.issn.1674-4764.2017.04.016



局域法邻近点选取对供水量预测精度的影响

任刚红, 杜坤, 刘年东, 周明, 李诚

(昆明理工大学 建筑工程学院, 昆明 650500)

摘要:混沌局域法预测模型适用于非线性、非平稳的城市日供水量预测, 而邻近相点个数的选取对该模型预测精度有直接影响。传统方法通常以嵌入维 m 作为参考值, 凭经验选取 $m+1$ 个邻近相点, 且仅使用欧式距离法计算当前相点距离, 无法反映相点的运动趋势, 易引入伪邻近相点, 导致预测精度的降低。鉴于此, 将演化追踪法引入城市日供水量预测, 通过挖掘邻近相点的历史演化规律对参考样本进行优选, 以提高预测精度。最后, 采用实际日供水量数据验证所提出方法, 结果表明, 运用演化追踪法优选邻近相点能显著提高日供水量预测精度, 预测平均绝对误差由 2.501% 降低到 1.683%。

关键词:混沌理论; 局域法; 邻近点; 演化追踪法; 供水量预测

中图分类号: TP183 **文献标志码:** A **文章编号:** 1674-4764(2017)04-0102-05

Influence of select the local-region method nearest neighbours on water supply forecasting accuracy

Ren Ganghong, Du Kun, Liu Niandong, Zhou Ming, Li Cheng

(Faculty of Civil Engineering and Mechanics, Kunming University of Science and Technology, Kunming, 650500, P. R. China)

Abstract: The chaotic local-region forecasting model is suitable for nonlinear and non-stationary urban daily water supply forecast, and the neighbourhood selection has a direct impact on the model prediction accuracy. The traditional method usually takes the embedded dimension m as a reference, and selects $m+1$ nearest neighbours by experience. It usually introduces the pseudo nearest neighbours, which leads to the reduction of the prediction accuracy. Accordingly, the evolutionary tracing method is introduced into the prediction of urban daily water supply. By mining the historical evolution of nearest neighbours, the reference samples are optimized to improve the prediction accuracy. The proposed method is validated by the actual daily water supply data. The results show that the optimal approach is significantly improved by

收稿日期: 2016-12-03

基金项目:国家自然科学基金(51608242);昆明理工大学 2016 年学生课外学术科技创新基金(2015YB025);云南省人才培养计划(14118943)

作者简介:任刚红(1992-),女,主要从事市政工程研究,(E-mail)554769994@qq.com。

杜坤(通信作者),男,博士,(E-mail)250977426@qq.com。

Received: 2016-12-03

Foundation item: National Natural Science Foundation of China (No. 51608242); Kunming University of Science and Technology, 2016 Students Extracurricular Academic Science and Technology Innovation Fund (No. 2015YB025); Personnel Training Program of Yunnan Province (No. 14118943).

Author brief: Ren Ganghong(1992-), main research interest: municipal engineering, (E-mail)554769994@qq.com.

Du Kun(corresponding author), PhD, (E-mail)250977426@qq.com.

using evolutionary tracing method, and the average absolute error is reduced from 2.501% to 1.683%.

Keywords: chaos; local-region method; neighbourhood selection; evolutionary tracing method; water supply forecast

现今水资源极度缺乏,城市供水优化调度显得尤为重要,而供水量预测又是供水优化调度的重要组成部分,其预测准确与否直接影响到给水系统调度决策能否是有效制定^[1-3]。供水系统是一个具有时变特性的复杂系统,其本身影响因子表现出非线性、非平稳的特性。供水量预测模型一般分为传统预测模型和基于新技术的方法模型,传统预测模型的参数筛选复杂,且多基于数学理论假设的基础上,预测效果难以保证。例如多元回归分析、统计预测法、专家意见法等^[4]。混沌理论是一种新的预测技术,其通过历史观测数据,找出蕴含其中的内在演化规律,对整个系统进行描述和解释,进而对系统做出预测。它的出现为供水量的预测研究提供了新的思路,Chang等^[5]在相空间重构的基础上,对城市日用水量时序进行了分析,计算出符合案例实际情况的嵌入维数、嵌入延迟和Lyapunov指数。Jayawardena^[6]为优选局域预测模型的邻近点个数,提出一种基于广义自由度的新方法。张琴等^[7]针对时供水量短期高精度预测难题,提出一种基于横向分时段和纵向残差修正的联合时序短期混沌预测方法。高金良等^[8]利用Lyapunov指数构造供水预测模型,避免了主观因素的影响。总之,基于混沌理论的时序预测技术具有很好的应用前景。

一般预测模型可分为3个步骤:1)为刻画实测数据蕴含的信息该采用何种样本形式;2)选择多少个样本作为参考样本;3)选择预测模型。混沌局域法:1)相空间重构;2)确定参考邻近点数;3)局域模型进行预测。通过对文献的总结,笔者发现绝大多数运用混沌局域法对供水量预测的研究都侧重于相空间重构及预测模型的选择,即第1和3步,很少考虑第2步“选择多少个样本作为参考样本”,即对邻近点选取问题的研究较少。实际上,混沌局域法的预测精度和计算量是由邻近点来决定的,邻近点数选取过少,大量有效信息可能被忽略,导致系统历史运动趋势无法被充分反映。邻近点数选取过多,易引入无效参考样本,而且,局域法的线性假设条件也可能不满足。因此,在使用混沌局域法对供水量进行预测时,研究邻近点的选取对预测精度的影响是十分有必要的。本文提出基于演化追踪法来确定局域法邻近点数的定量方法,最后利用实测日供水

量数据对该方法进行了分析验证。

1 混沌局域法多步预测模型

研究表明,混沌局域模型对噪声较大的复杂系统的预测较其他模型更具优势^[9],它不是用所有状态向量来拟合重构函数,而是选定相空间最后的一状态点作为基准点 X_M^r ,再挑选邻近 X_M^r 的 K 个状态点来拟合,故其拟合速度快,预测精度高。

1)相空间重构^[10]。其基本思路是将一维混沌时间序列映射到高维的空间,目的是恢复有规律的吸引子,从而使蕴藏在时间序列中的信息显露。设混沌时间序列 $\{x_1, x_2, \dots, x_N\}$ 则相空间重构为

$$X(t_i) = [x(t_i), x(t_i + 2\tau), \dots, x(t_i + (m-1)\tau)] \\ (i = 1, 2, \dots, N - (m-1)\tau) \quad (1)$$

式中: m 为嵌入维数; τ 为嵌入延迟。二者共同决定了相空间形态,也即拟合样本形式。常用计算 m 与 τ 的方法有:邻近点维数法、 $G-P$ 法、高阶相关法、互信息法以及 $C-C$ 法等^[11]。

2)选取邻近点。在相空间中找出基准点 X_M^r 的参考邻近点集 $\{Y_i^r, i = 1, 2, \dots, K\}$ (传统邻近点数确定方法为 $K = m + 1$),计算出各邻近点到 X_M^r 之间的欧式距离为 d_i ,设 d_{\min} 是 d_i 中的最小值,定义 $\{Y_i^r\}$ 的权重系数矩阵 W_i ,得

$$d_i = Y_i^r - X_{M2}^r \quad (2)$$

$$W_i = \frac{e^{(-\alpha(d_i - d_{\min}))}}{\sum_{i=1}^K e^{(-\alpha(d_i - d_{\min}))}} \quad (3)$$

式中: $i = 1, 2, \dots, K$; α 为常数,一般取 $\alpha = 1$ 。

3)计算拟合参数。设 X_M^r 的 S 步演化相点为 $X_{M+\theta}^r$,对一阶多步局域预测模型进行线性拟合,得

$$\begin{bmatrix} X_{M+\theta}^1 \\ X_{M+\theta}^2 \\ \vdots \\ X_{M+\theta}^q \end{bmatrix} = \begin{bmatrix} e & X_M^1 \\ e & X_M^2 \\ \vdots & \vdots \\ e & X_M^q \end{bmatrix} \begin{bmatrix} a_s \\ b_s \end{bmatrix} \quad (4)$$

式中: a_s 和 b_s 为待求的拟合参数。鉴于大多数文献中 a_s 、 b_s 计算式冗长、不易理解,本文给出 a_s 、 b_s 的矩阵表达式,使其便于理解且易于编程^[12]。设线性方程组(4)的矩阵表达式为

$$Y = Ax \quad (5)$$

$$\text{式中: } Y = \begin{bmatrix} X_{M+\vartheta}^1 \\ X_{M+\vartheta}^2 \\ \vdots \\ X_{M+\vartheta}^q \end{bmatrix}, A = \begin{bmatrix} e & X_M^1 \\ e & X_M^2 \\ \vdots & \vdots \\ e & X_M^q \end{bmatrix}, x = \begin{bmatrix} a_s \\ b_s \end{bmatrix}.$$

这里定义目标函数为

$$f(x) = (Ax - Y)^T W_i (Ax - Y) \quad (6)$$

依据多元函数的极值理论,要使公式(6)取得最小解则有

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2A^T W_i (Ax - Y) = 0 \\ \Rightarrow A^T W_i A x &= A^T W_i Y \end{aligned} \quad (7)$$

即得参数 a_s, b_s 的矩阵表达式

$$\begin{bmatrix} a_s \\ b_s \end{bmatrix} = (A^T W_i A)^{-1} A^T W_i Y \quad (8)$$

4) 预测计算。将 a_s, b_s , 带入 ϑ 步预测公式(4), 即可得到演化 ϑ 步后的预测值 $X_{M+\vartheta}^r$, 其最后一个分量即为原时序的 ϑ 步预测值 $x_{N+\vartheta}$ 。

2 演化追踪法原理

多数学者在运用混沌局域模型预测供水量时, 往往忽略邻近点的选取问题, 一般认为当基准点与邻近点间的欧式距离小于某阈值则为有效邻近点, 或简单地取邻近点数为 $K = m + 1$, 并不考虑邻近点的运动与演化趋势, 易引入“伪邻近点”或“弱相关点”, 这些无效参考点可能会导致模型预测效果不佳甚至失效^[13]。

局域预测法认为混沌吸引子外的任一点都与其邻近点域有相同的演化行为, 因此通过拟合各邻近点运动趋势就可推断出预测点的运动趋势。如图 1 所示, 设到基准点 X_M^r 的欧式距离满足某一阈值的邻近点有 $\{Y_i^r, i = I, J, K, L\}$, $\{Y_i^r\}$ 上一步历史相点则为 $\{Y_j^r, j = I - 1, J - 1, K - 1, L - 1\}$ 。同理, $\{Y_i^r\}$ 的下一步未来相点则为 $\{Y_k^r, k = I + 1, J + 1, K + 1, L + 1\}$ 。邻近点的 4 种可能的演化情况如图 1 所示: ① 线 Y_J^r 与 X_M^r 的历史相点 Y_{J-1}^r 与 X_{M-1}^r 很接近, 但未来相点 Y_{J+1}^r 与 X_{M+1}^r 却相距甚远, 说明两个相点的演化趋势差逐渐偏离, X_M^r 的演化趋势不能 Y_J^r 预示, Y_J^r 是伪邻近点; ② 线 Y_I^r 与 X_M^r 相应的历史相点、未来相点皆接近, Y_I^r 为理想的参考邻近点; ③ 线 Y_L^r 和 X_M^r 的历史相点 Y_{L-1}^r 与 X_{M-1}^r 距离虽较远, 但未来相点 Y_{L+1}^r 与 X_{M+1}^r 却慢慢靠近, 说明其与基准点的演化行为趋于一致, Y_L^r 是一般的参考邻近点; ④ 线 Y_K^r 与 X_M^r 的历史相点、未来相点都相距较远, 虽满足邻近点定义, 但与 X_{M+1}^r 的演化趋势无相关性, Y_K^r 是伪邻

近相点。

若基准点 X_M^r 及其历史相点 X_{M-n}^r 与对应的邻近点“一直邻近”, 即说明基准点与邻近点的演化行为相同, 进而可推出预测点的演化趋势, 这就是演化追踪法的基本原理。

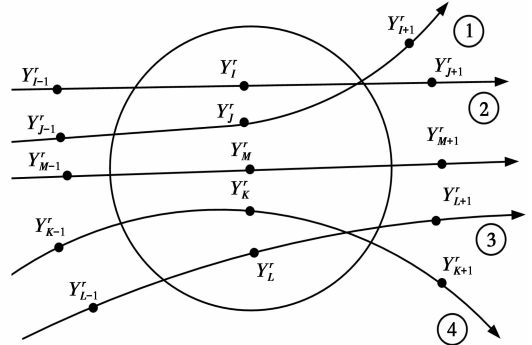


图 1 预测参考点演化轨迹示意图
Fig. 1 The schematic diagram of evolutionary track of reference points

3 城市供水量预测实例

3.1 预测实例

根据文献[15], 原始供水数据中含有季节性、趋势性和噪声, 为消除这些不利影响、提高预测精度, 仅选取 2005—2011 年每年 1 月的日供水时序作为单独的时间序列进行预测, 对于其它月份的预测可照此方法依次处理。由小数据量法计算得该时序的 Lyapunov 指数为 $\lambda = 0.132$, 说明该供水系统是一种混沌系统。采用文献[16]中推荐的方法计算嵌入维度 $m = 10$, 则传统法邻近点个数计为 $K = 11$ 。考虑到互信息法^[17]较其他嵌入延迟算法, 更适用于提取高维混沌系统的非线性特征, 因此选用互信息法计算嵌入延迟得 $\tau = 6$ 。

利用演化追踪法回溯两步历史选取邻近点^[18], 先计算基准点 X_M^r 的 $K (K \leq N - (m - 1)\tau)$ 个初始邻近点欧式距离 $Y_i^r (i = 1, 2, \dots, K)$, 从中寻找与 X_M^r 在演化过程中“一直邻近”的邻近点。再计算 X_M^r 的上一步历史点 X_{M-1}^r 的 K 个初始邻近点欧式距离 $Y_i^r (i = 1, 2, \dots, K)$; 再计算 X_M^r 的前两步演化点 X_{M-2}^r 的 K 个初始邻近点欧式距离 $Y_i^r (i = 1, 2, \dots, K)$; 若 Y_i^r 的上一步演化点欧式距离在 $\{Y_i^r (i = 1, 2, \dots, K)\}$ 中, 且 Y_i^r 的前两步演化点在点集 $\{Y_i^r (i = 1, 2, \dots, K)\}$ 中, 则点 Y_i^r 是基准点 X_M^r 的真邻近点(即 $Y_i^r = Y_i^0 \cap Y_i^1 \cap Y_i^2$), 否则, 为伪邻近点。同理, 依次判断剩余的相点, 此案例计算得邻近点个数 $K = 7$, 比传统的参考邻近点个数少 4 个。

选取 210 个日供水量时序进行相空间重构,考虑到局域模型只能进行短期预测及 Lyapunov 指数,则验证样本数为 $1/\lambda$ 个日供水时序。总体预测趋势如图 2 示,局部预测细节如图 3 示。

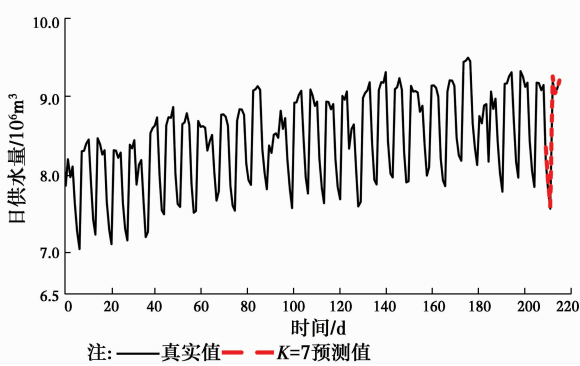


图 2 日供水量总体预测趋势

Fig. 2 General forecast trend of daily water supply

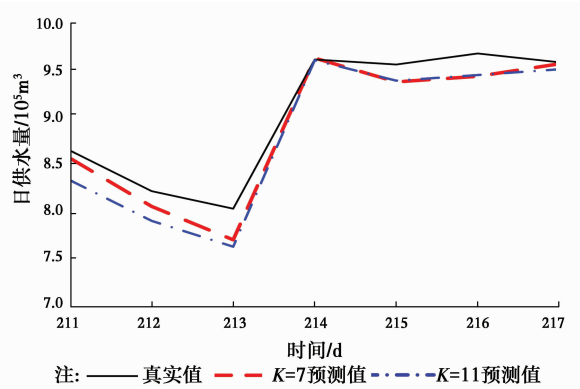


图 3 日供水量局部预测细节

Fig. 3 Local forecast details of daily water supply

由图 2、3 可见,“K=7 预测值”与真实值总体走势相同,其能较好地把握供水系统的总体运动趋势;演化追踪法与传统法二者预测值虽走势相同,但演化追踪法对系统的走势细节把握更准确。图 4 可见,演化追踪法的相对误差总体小于传统法相对误差,说明演化追踪法能提高预测精度。

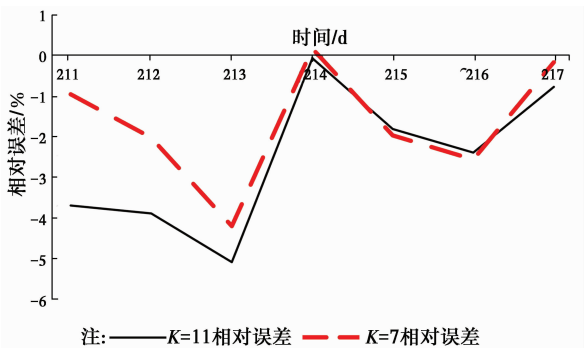


图 4 相对误差

Fig. 4 Relative error

3.2 模型精度检验

这里选取常用的平均绝对百分比误差作为模型精度的检验标准,计算式为

$$MAPE = \frac{1}{n} \sum_{i=1}^n |P_i| \quad (10)$$

式中: n 为样本数据的个数; P_i 为相对百分比误差。

表 1 预测精度比较

Table 1 Comparison of prediction accuracy

邻近点数	分步预测相对误差 $P_i/\%$			
K=11	-3.639	-3.832	-5.008	-0.087
K=7	-0.946	-1.983	-4.097	0.146
邻近点数	分步预测相对误差 $P_i/\%$			MAPE/%
K=11	-1.794	-2.365	-0.778	2.501
K=7	-1.930	-2.505	-0.176	1.683

相同预测步数情况下, K 值大的预测精度反而低,说明初始邻近相点中存在伪邻近相点或邻近程度较弱的相点,对预测起负面作用,需要进行辨识并剔除。运用演化追踪法对邻近点进行优选,预测精度有显著提高,其平均绝对误差由 2.501% 减少到了 1.683%。

4 结论

1) 在利用混沌局域法模型预测具有混沌特性的城市日供水量时,传统方法通常凭经验选取邻近点,可能会引入伪邻近点,预测精度不理想。

2) 一般地, K 值越大越容易掺杂伪邻近点,运用演化追踪法挖掘邻近点的历史演化信息,可有效辨别真、伪邻近点。实例中日供水量预测的 MAPE 由 2.501% 降低到 1.683%,说明该方法提高了预测精度。

3) 演化追踪法能辨别、剔除伪邻近点,获取有效参考样本,克服了凭主观经验或多次试验来确定邻近点个数问题,也为其他模型优选参考样本提供了新思路。

参考文献:

[1] LILY H P, BETHANY P. Effects of urban spatial structure, sociodemographics, and climate on residential water consumption in hillsboro, oregon [J]. Journal of the American Water Resources Association, 2010, 46(3):461-472.

[2] YASAR A, BILGILI M, SIMSEK E. Water demand forecasting based on stepwise multiple nonlinear regression analysis [J]. Arabian Journal for Science and

- Engineering, 2012, 37(8): 2333-2341.
- [3] 孙月峰, 闫雅飞, 张表志. 基于 T-S 模型的模糊神经网络城市需水量预测方法研究[J]. 安全与环境学报, 2013, 13(2): 136-139.
SUN Y F, YAN Y F, ZHANG B Z. Fuzzy neural network method based on the T-S model for the urban water demand forecasting and its application [J]. Journal of Safety and Environment, 2013, 13(2): 136-139. (in Chinese)
- [4] MOMBENI H A, REZAEI S, NADARAJAH S, et al. Estimation of water demand on SARIMA models [J]. Environmental Modeling and Assessment, 2013, 18(5): 559-565.
- [5] CHANG K, GAO J, CHEN M L, et al. Urban daily water demand short-term forecasts based on the chaos theory [C]// International Conference on Natural Computation, IEEE Computer Society, 2008:696-699.
- [6] JAYAWARDENA A W. Neighbourhood selection for local modelling and prediction of hydrological time series [J]. Journal of Hydrology, 2002, 258:40-57.
- [7] 张琴, 汪雄海, 朱庆建. 基于联合时序的混沌时用水量短期预测调度[J]. 灌溉机械工程学报, 2011, 29(4): 353-360.
ZHANG Q, WANG X H. Short-term prediction of chaotic hourly water consumption based on united time series[J]. Journal of Drainage and Irrigation Machinery Engineering, 2011, 29(4):353-360. (in Chinese)
- [8] 高金良, 常魁, WEN Y W, 等. 城市用水量预测的混沌理论研究[J]. 哈尔滨商业大学学报(自然科学版), 2009, 25(1):88-91.
GAO J L, CHANG K, WEN Y W, et al. Study on urban daily water demand short-term forecasts based on chaos theory [J]. Journal of Harbin University of Commerce (Natural Sciences Edition), 2009, 25(1):88-91. (in Chinese)
- [9] BAI Y, WANG P, LI C, et al. A multi-scale relevance vector regression approach for daily urban water demand forecasting [J]. Journal of Hydrology, 2014, 517: 236-245.
- [10] TAKENS F. Determining strange attractors in turbulence [J]. Lecture Notes in Mathematics, 1981, 898: 361-381.
- [11] 韩敏. 混沌时间序列预测理论与方法[M]. 北京: 中国水利水电出版社, 2007:28-39.
- HAN M. Prediction theory and method of chaotic time series [M]. Beijing: China Water&Power Press, 2007: 28-39. (in Chinese)
- [12] DU K, PH D, LONG T Y. Inversion model of water distribution systems or nodal demand calibration [J]. Journal of Water Resources Planning and Management, 2015, 4(15): 1-12.
- [13] 刘年东, 杜坤, 周明, 等. 局域法邻近点选取对降雨量预测精度影响研究[J]. 给水排水, 2016, 42: 285-288.
LIU N D, DU K, ZHOU M, et al. Study on the influence of select the local-region method nearest neighbours on rainfall forecasting accuracy [J]. Water & Wastewater Engineering, 2016, 42: 285-288. (in Chinese)
- [14] 唐巍, 谷子. 基于相关邻近点与峰谷荷修正的短期负荷时间序列预测[J]. 电力系统自动化, 2006, 30(14): 25-29.
TANG W, GU Z. Short-term load time series forecasting based on correlative neighboring points and peak-valley correction [J]. Automation of Electric Power Systems, 2006, 30(14):25-29. (in Chinese)
- [15] 张善文, 雷英杰, 冯有前. MATLAB 在时间序列分析中的应用[M]. 西安: 西安电子科技大学出版社, 2007: 20-27.
ZHANG S W, LEI Y J, FENG Y Q. The application of MATLAB in time series analysis [M]. Xi'an: Xidian University Press, 2007:20-27. (in Chinese)
- [16] MIRMOMENI M, LUCAS C. Model based method for determining the minimum embedding dimension from solar activity chaotic time series [J]. Materials and Energy Research Center, 2008, 21(1):31-41.
- [17] JIN S H, PETER L, MARK H. Linear and nonlinear information flow based on time-delayed mutual information method and its application to corticomuscular interaction [J]. Clinical Neurophysiology, 2010, 121: 392-401.
- [18] 高俊杰. 混沌时间序列预测研究及应用[D]. 上海: 上海交通大学, 2013: 53-55.
GAO J J. Study and application of chaotic time series prediction [D]. Shanghai: Shanghai Jiao Tong University, 2013: 53-55. (in Chinese)