

# 淮安市水源水质与胃肠道疾病发病关系的 探讨分析

高 健, 秦振声, 徐子为, 尚 进\*

(南京医科大学附属淮安第一医院, 淮安 223300)

**摘 要: 目的** 探讨水源水质与胃肠道疾病之间的关系, 以及胃肠道疾病的发病规律。**方法** 选择淮安市 2006 年 1 月至 2013 年 12 月间水源水质指标及胃肠道患病例数, 并进行相关性分析, 然后进行具有输出时延反馈的非线性自回归(NARX)神经网络模型的构建、训练及模拟。**结果** 水源水质的石油类、挥发酚、总砷、总汞、6 价铬、氟化物及粪大肠菌群等指标与发病人数有显著的相关性。经过训练后 NARX 神经网络的可决系数为 0.716, 拟合效果比较理想。**结论** 水源水质与胃肠道疾病之间存在一定的相关性, 基于 NARX 神经网络对水源水质与胃肠道疾病进行拟合是可行的。

**关键词:** 胃肠道疾病; NARX 神经网络; 水源水质; 拟合

## Analysis of water quality relationship with the incidence of gastrointestinal disease in Huai'an

GAO Jian, QIN Zheng-Sheng, XU Zi-Wei, SHANG Jin\*

(Department of neurology Huai'an First People's Hospital, Nanjing Medical University, Huai'an 223300, China)

**ABSTRACT: Objective** To discuss the relationship between water quality and gastrointestinal diseases, and variation of gastrointestinal diseases. **Methods** Data of water quality indicators and gastrointestinal cases in Huai'an from January 2006 to December 2013 were collected. Correlation analysis was made and the NARX model was built, trained and simulated. **Results** Petroleum, phenol, total arsenic, total mercury, hexavalent chromium, fluoride and fecal coliform had a significant correlation with the gastrointestinal diseases. The coefficient of determination after the NARX neural network training was 0.716, and the fitting effect was perfect. **Conclusion** There was a certain correlation between water quality and gastrointestinal diseases. Fitting of water quality and gastrointestinal diseases based on nonlinear autoregressive neural network with external input was feasible.

**KEY WORDS:** gastrointestinal diseases; nonlinear autoregressive neural network with external input; water quality; fitting

基金项目: 江苏省淮安市科技支撑计划课题(has08006)

**Fund:** Supported by the Science and Technology Support Program of Huai'an(has08006)

\*通讯作者: 尚进, 副主任医师, 江苏省南京医科大学附属淮安第一医院神经内科, 223300, Email:sj.hy@163.com

\*Corresponding author: SHANG Jin, Associate Professor, Department of Neurology, Huai'an First Hospital affiliated to Nanjing Medical University Huai'an 223300, China. E-mail:sj.hy@163.com

1 引 言

胃肠道是人体最大的免疫器官,也是人体最大的排毒器官。胃肠道疾病包括消化性溃疡、胃癌和食道癌等疾病<sup>[1]</sup>,目前已成为影响淮安市居民健康的第 2 大非传染性疾病<sup>[2-4]</sup>,尤其食道癌在全国更属高发地区。而饮用水作为人类生存的最基本需求,其水质的好坏将直接决定了人民的健康与安全。影响水源水质的指标如 pH 值等无不与胃肠道疾病有着密切的关系,然而这些指标的影响很难进行量化描述。人工神经网络是由大量简单处理单元广泛连接而构成的一个复杂的、非线性系统,是对大脑神经网络结构和功能的初步模拟,是建立时间序列动态模型进行预测的有力工具。本研究尝试从以往各月饮用水源水质与发病情况的角度出发,采用具有输出时延反馈的非线性自回归神经网络(nonlinear autoregressive neural network with external input, NARX)<sup>[5]</sup>,利用 MATLAB R2010b 建立胃肠道疾病的 NARX 神经网络模型,探讨水源水质与胃肠道疾病之间的关系并对疾病的发病情况进行拟合等。

2 对象与方法

2.1 研究对象

水源水质数据来源于淮安市环境保护局网站发布的 2006 年 1 月至 2013 年 12 月的环境质量的水质月报,选取淮安古黄河的淮安市取水口各月的 pH 值、溶解氧(mg/L)、生化需氧量(mg/L)、高锰酸盐指数(mg/L)、氨氮(mg/L)、石油类(mg/L)、硝酸盐氮(mg/L)、挥发酚(mg/L)、总氰化物(mg/L)、总砷(mg/L)、总汞(mg/L)、6 价铬(mg/L)、总铅(mg/L)、总磷(mg/L)、总镉(mg/L)、氯化物(mg/L)、氟化物(mg/L)和粪大肠菌群(mg/L)等 18 个指标数据<sup>[6]</sup>,另以 ICD-10 疾病编码为依据,选取淮安市疾病预防控制中心所监测的胃肠道的月患病人数作为研究对象。

2.2 研究方法

首先利用 SPSS 14.0 软件对淮安市水源水质的 18 个指标分别与胃肠道患病人数进行相关性分析,然后利用 MATLAB R2010b 软件的神经网络工具箱进行 NARX 神经网络模型的构建、训练及模拟,最终建立基于 NARX 神经网络的时间序列模型,并利用所建立的模型进行拟合等,以  $P<0.05$  为差异有统计

学意义。

3 结 果

3.1 水源水质与胃肠道患病人数的相关性分析

除溶解氧指标外其余指标的取值一般均在标准范围内,此外淮安市水源水质的各指标与胃肠道患病人数的相关系数  $r$  与  $P$  值显示各指标与胃肠道患病人数的相关系数均较小。而石油类、挥发酚、总砷、总汞、6 价铬、氟化物及粪大肠菌群等指标与发病人数有显著的相关性,且部分指标与发病人数呈现出负相关性的特点,水源水质的标准范围及相关系数等详见表 1。

表 1 水源水质及其与胃肠道患病人数的相关系数表  
Table 1 Water quality of source the water and its relationship with the number of gastrointestinal illness

指标	标准范围	均值±标准差( $\bar{x} \pm s$ )	$r$	$P$
pH 值	6~9	7.94±0.34	0.11	>0.05
溶解氧	4	7.61±2.82	-0.05	>0.05
生化需氧量	5	3.89±2.72	-0.09	>0.05
高锰酸盐指数	6	3.54±0.60	0.04	>0.05
氨氮	1.0	0.38±0.11	0.23	>0.05
石油类	0.05	0.04±0.01	0.38	<0.05
硝酸盐氮	10	0.88±0.39	-0.06	>0.05
挥发酚	0.005	0.0015±0.0008	-0.21	>0.05
总氰化物	0.2	0.0032±0.0010	0.24	>0.05
总砷	0.05	0.0048±0.0021	-0.38	<0.05
总汞	0.0001	0.0010±0.0017	-0.38	<0.05
六价铬	0.05	0.0029±0.0017	0.37	<0.05
总铅	0.05	0.0169±0.0387	0.21	>0.05
总磷	0.2	0.11±0.05	-0.07	>0.05
总镉	0.005	0.0002±0.0005	-0.16	>0.05
氯化物	250	43.77±9.00	-0.02	>0.05
氟化物	1.0	0.56±0.13	-0.36	<0.05
粪大肠菌群	10000	1986.57±1125.58	0.31	<0.05

3.2 NARX 神经网络的构建、训练及模拟

NARX 神经网络共包括输入层,隐含层和输出层对数据。首先进行预处理,以各指标的值除以该指标的最大值,其次从原始数据中随机抽取 70% 的样本作为训练集,15% 的样本作为校验集及 15% 的样本

作为测试集。网络输入层节点的个数由输入向量的个数决定,选取水质的 18 个指标作为输入,对于隐含层目前仍无统一的标准来确定个数,本研究采用公式  $n_1 = \log_2 n$  来确定,其中  $n_1$  为隐含层的单位数,  $n$  为输入层的单位数<sup>[7]</sup>,由此得到隐含层个数为 5,由于输出层仅为患者人数,故输出层节点的个数为 1。隐含层和输出层的激活函数分别取双曲正切函数及线性函数,网络训练采用 Levenberg-Marquardt 优化算法。对于输入层和输出层的自回归阶数目前也无统一的标准来确定,本研究尝试从 2 阶开始实验,经反复实验确定当阶数  $d$  取 4 时具有相对较小的误差和较快的收敛速度。

### 3.3 NARX 神经网络的训练结果

在以上的条件下,仅经过 13 次的迭代训练以后,网络的系统误差降至 0.007129。经过训练后 NARX 神经网络的拟合值  $y'$  与实际值  $y$  的关系为  $y' = 0.84y + 0.06$ ,可决系数  $R^2 = 0.716$ ,图 1 为输出层的时间序列拟合图,除个别极端点外训练集、校验集及测试集的拟合效果均比较理想。

## 4 讨 论

对于本研究如发病人数与水源水质的时间序列数据,一般常采用传统的时间序列方法,其假定变量间的关系为线性的,而实际问题一般均为复杂的非线性系统,使用传统的时间序列模型进行拟合、预测难免误差较大<sup>[8-11]</sup>。NARX 神经网络能学习训练集输

入、输出数据之间的函数关系,适用于解决复杂的非线性问题,从而对未来的输出数据作出准确预测,本研究很好地证明了这一点。然而,目前 NARX 神经网络在流行病学领域的应用还不够成熟,如输入变量的选择,如何合理地确定隐含层个数及自回归阶数等还需进一步的研究。

本研究尝试不仅利用胃肠道疾病的历史数据,而且引入了水源水质的相关指标建立 NARX 神经网络模型,结果证明该方法是可行的。另外, NARX 神经网络模型能根据模型拟合的效果,自主调节自回归的阶数,具有很强的适用性。本文构建的 NARX 神经网络模型表明胃肠道发病人数主要与过去 4 个月的水源水质有关,这也为淮安市胃肠道疾病的防治提供了一定的参考意义。

水是居民最基本的生活需要品之一,水质的好坏直接影响居民的健康<sup>[12]</sup>。然而,目前居民饮水类型的主体已发生了较大的变化,从原来的分散式供水逐步转变为集中式的供水<sup>[13-15]</sup>,我市古黄河的取水口的水源水质监测指标显示除溶解氧指标高出标准范围外,其水质质量均在标准范围内,比较符合正常饮用水的标准。相关性分析表明各指标与胃肠道患病人数的相关系数均较小,且部分指标与发病人数也存在显著的相关性。

此外,胃肠道疾病作为临床上常见的、多发的疾病之一<sup>[16]</sup>,对其进行研究有着十分积极的意义。胃肠道疾病的发病情况通常与当地的饮食习惯、食品质量及水源水质等因素有关,避开这些因素仅利用胃肠

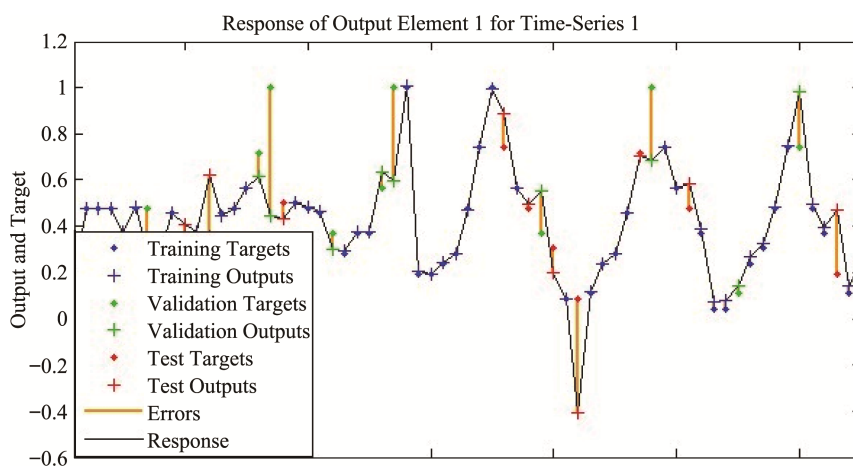


图 1 NARX 神经网络的训练输出集的拟合效果图

Fig. 1 NARX output of neural network training set fitting effect

道疾病的历史数据进行预测则很难避免其他相关影响因素有较大变动时所造成的影响。若能进一步地引入饮食及生活习惯等因素则模型应更为客观、科学。

### 参考文献

- [1] David AW. Oxford textbook of medicine[M]. Oxford: Oxford University Press, 2005: 511-515.
- [2] 潘恩春, 王仪, 姚海波, 等. 淮安市慢性非传染疾病流行现状调查[J]. 现代预防医学, 2006, 33(11): 2083-2084.  
Pan EC, Wang Y, Yao HB, *et al*. The prevalence investigation of chronic non-communicable disease in Huaian[J]. J Prev Med, 2006, 33(11): 2083-2084.
- [3] 孔灵芝. 慢性非传染性疾病流行现状、发展趋势及防治策略[J]. 中国慢性病预防与控制, 2002, 10: 1-2.  
Kong LZ. Prevalence, trends and control strategies in chronic non-communicable diseases[J]. China Chronic Disease Prevention Control, 2002, 10: 1-2.
- [4] 刘荣海, 唐才昌, 蒋忠, 等. 盐城市慢性非传染性疾病综合防治社区诊断报告[J]. 中国初级卫生保健, 2003, 17: 31-32.  
Liu RH, Tang CC, Jiang Z, *et al*. Community diagnostic report of comprehensive prevention and control of chronic non-communicable diseases in Yancheng[J]. Chin Primary Health Care, 2003, 17: 31-32.
- [5] 飞思科技产品研发中心. 神经网络理论与 Matlab 7 实现[M]. 北京: 电子工业出版社, 2005: 120-121.  
Fly Synopsys R & D center. Neural network theory and Matlab 7 to achieve[M]. Beijing: Electronic Industry Press, 2005: 120-121.
- [6] 淮 安 市 环 境 保 护 局 [EB/OL]. <http://hbj.huaian.gov.cn/web/hbj/4924/54396/6346/6346.shtml>.  
Huaian state environmental protection administration of china [EB/OL]. <http://hbj.huaian.gov.cn/web/hbj/4924/54396/6346/6346.shtml>
- [7] 沈花玉, 王兆霞, 高成耀, 等. BP 神经网络隐含层单元数的确定[J]. 天津理工大学学报, 2008, 24(5): 13-15.  
Shen HY, Wang ZX, Gao CY, *et al*. BP neural network to determine the number of units in hidden layer[J]. J Tianjin Univ Technol, 2008, 24(5): 13-15.
- [8] 张淑英, 余丽君, 王爱萍, 等. 慢性胃肠道疾病病人生活质量的调查[J]. 现代护理, 2008, 14(1): 29-31.  
Zhang SY, Yu LJ, Wang AP, *et al*. Investigate the quality of life of chronic gastrointestinal disease patients[J]. Mod Nursing, 2008, 14(1): 29-31.
- [9] ShallowS, Samuel M, McNees A, *et al*. Preliminary foodnet data on the incidence of food borne illnesses selected sites[J]. MMWR, 2000, 49(10): 201-205.
- [10] Aymerieh T, Garriga M, Ylla J, *et al*. Application of enterococci as bio preservatives against Listeria innocua in meat Products[J]. J Food Protect, 2000, 62(6): 721-726.
- [11] Smith JL. Foodborne illness in the elderly[J]. J Food Protect, 1998, 61(9): 1229-1239.
- [12] Fidler D. Globalization international law and emerging infectious diseases[J]. Emerg Infect Dis, 1996, 2(2): 77-84.
- [13] Kaferstein FK, Motarjemi Y, Bettcher DW. Foodborne disease control: A transnational challenge[J]. Emerg Infect Dis, 1997, 3(4): 503-510.
- [14] Hall RL. Foodborne illness: Implications for the future[J]. Emerg Infect Dis, 3(4): 555-559.
- [15] 吴家兵, 叶临湘, 尤尔科, 等. 时间序列模型在传染病发病率预测中的应用[J]. 中国卫生统计, 2006, 23: 276.  
Wu JB, Ye LX, You EK, *et al*. The application of time series model to predict the incidence of infectious diseases[J]. Chin Health Statistics, 2006, 23: 276.
- [16] 凌文华, 朱惠莲. 控制食源性疾病是一项全球关注的公共卫生问题[J]. 疾病控制杂志, 1999, 4(3): 304-306.  
Ling WH, Zhu HL. Control of foodborne disease is a public health problem of global concern[J]. Dis Control, 1999, 4(3): 304-306.

(责任编辑: 邓伟)

### 作者简介



高健, 统计师, 江苏省南京医科大学  
附属淮安第一医院统计室。  
E-mail: gg1986jj@163.com



尚进, 副主任医师, 江苏省南京医科  
大学附属淮安第一医院神经内科。  
E-mail: sj.hy@163.com