

基于 LM-BP 神经网络的 Argo 数据西北太平洋海水温度模型

韩震^{1,2} 赵宁¹

(1. 上海海洋大学 海洋科学学院, 上海 201306; 2. 上海海洋大学 大洋渔业资源可持续开发省部共建教育部重点实验室, 上海 201306)

摘要: 以2007年西北太平洋海域 Argo 海表面温度、经纬度、深度为输入参数, 利用 LM-BP 神经网络, 构建了西北太平洋海水温度模型。将均方根差以及 Pearson 相关性系数作为检验指标, 利用2008年和2009年的 Argo 数据对模型进行了检验。检验结果为: 2008年均方根误差为0.7140℃, Pearson 相关性系数为0.9968; 2009年均方根误差为0.7615℃, Pearson 相关性系数为0.9965。表明所建立的基于 LM-BP 神经网络的 Argo 数据西北太平洋海水温度模型是可行的。

关键词: 海水温度; 西北太平洋; LM-BP 神经网络; Argo 数据

中图分类号: X143 文献标识码: A 文章编号: 1007-6336(2012)04-0555-06

Seawater temperature model from Argo data by LM-BP neural network in Northwest Pacific Ocean

HAN Zhen^{1,2} ZHAO Ning¹

(1. College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China; 2. Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, Shanghai Ocean University, Shanghai 201306, China)

Abstract: Using the LM-BP neural network and choosing the sea surface temperature, longitude, latitude and depth obtained from Argo data in 2007 as input parameters, the seawater temperature model of the Northwest Pacific Ocean was built. Using the root-mean-square error (RMSE) and the Pearson's correlation coefficient (R) as test indices, the model was evaluated by the data in the period 2008 ~ 2009. The results were that the RMSE was 0.7140 °C and R was 0.9968 in 2008. The RMSE was 0.7615 °C and R was 0.9965 in 2009. It shown this seawater temperature model was.

Key words: seawater temperature; Northwest Pacific Ocean; LM-BP neural network; Argo data

海水温度是研究海洋的重要参数之一,它是海洋热力、动力以及海-气相互作用的综合结果^[1]。传统海洋观测方法无法获取大面积同步的海水温度数据,而遥感观测方法虽然克服了传统观测方法的部分缺陷,但其观测也仅能停留在海洋表层,无法获取表层以下的垂向的海水温度分布信息。如果可以建立海洋表层与表层以下的海水温度关系模型,对海洋热结构、水文动力过程、水团等海洋现象的研究将有重要的实际意义。

国外很早就开展了此项研究。Carnes 等^[2] 1990 年在美国海特瑞斯角 (Cape Hatteras) 的墨西哥湾流区 (the Gulf Stream), 利用海表面温度数据以及垂向温度数据,建

立了海水温度垂直反演的基本关系模型。Chu 等^[3] 2000 年在中国南海区域的现场观测水温剖面的基础上,利用表层水温反演了水温的垂直剖面。Guinehut 等^[4] 2004 年利用海表面温度与 Argo 数据模拟了北大西洋水下 200 m 的温度场,结果表明可以采用线性模型简单模拟出一定深度的温度大致分布。González 等^[5] 在 2006 年以北大西洋东部的比斯开湾为研究区域,建立了上层海水的垂直结构模型。

国内开展此项研究较晚,徐玉涓等^[6] 2009 年采用美国海军研究实验室开发的模块化海洋数据同化系统 (modular ocean data assimilation system, MODAS) 系统进

收稿日期: 2011-07-12, 修订日期: 2011-09-08

基金项目: 教育部科学技术研究重点项目 (209047); 国家发改委高技术产业化示范工程项目 (2009214)

作者简介: 韩震 (1969-) 男, 山东省德州市人, 博士, 教授, 主要从事海洋遥感技术研究, E-mail: zhhan@shou.edu.cn

行了东海、日本海以及黑潮流域的试验数据统计分析。修树孟等^[7] 2009年对黄海海水温垂直变化简化成分段线性模型,并利用海表面温度直接反演得到了黄海海水温垂直剖面图。池建军等^[8] 2010年利用MODAS的原理,将海表面温度、高度等信息与Argo数据结合,对西北太平洋海域的水温垂直分布进行了反演。

通过前人的研究可以看出,国外对于水温的垂直结构的研究较为深入,其所建立的模型多是根据观测所得海水温度垂直结构,将海水温度垂直分布进行一定的简化,从而建立经验统计模型。国内的研究则大多为对国外已有模型的验证和应用,而对模型开发的研究非常少见,仅有的研究也是采用分段线性的模型,而对于复杂多变的海水温度结构,线性模型效果并不十分理想。本文以西北太平洋为研究区域,采用基于Levenberg-Marquardt算法的BP神经网络模型,将实测的海水温度数据、经纬度以及深度作为参数,通过模拟获得某一经纬度某一深度的海水温度值,旨在通过本次研究,建立西北太平洋海洋表层与表层以下的海水温度关系模型,为西北太平洋资源的开发利用提供科学依据。

1 数据获取与预处理

本文采用的数据是西北太平洋2007年1月至2009年12月月平均温度数据产品。数据来源于中国Argo实时资料中心(www.argo.org.cn),由国家海洋局海洋环境信息保障技术重点实验室与国家海洋信息中心共同制作。

该数据水平分辨率为 $1^{\circ} \times 1^{\circ}$,深度范围为0~2 000

m。垂向分层划分方式为:0~50 m内,每10 m划分一层;50~150 m内,每25 m划分一层;150~300 m内,每50 m划分一层;300~1 500 m内,每100 m划分一层;1 500~2 000 m内,每250 m划分一层;总计26层。本文选取的研究区域为 $20^{\circ}\text{N} \sim 35^{\circ}\text{N}$, $145^{\circ}\text{E} \sim 180^{\circ}\text{E}$,总计576个站点(图1)。该区域处在西北太平洋范围内,远离大陆架,足以确保海水深度可以达到2 000 m,并且不易受到人类活动因素影响。本研究选择2007年至2009年共计36个月的数据,将每个月的每一点的每一深度层的温度、经度、纬度和该经纬度的海表面温度(0 m处)数据作为模型的一组参数,每年共有179 544组(原数据中 32°N , 173°E 以及 35°N , 172°E 两处仅有0~1 000 m的19层数据,因此每月均缺少了14组)。

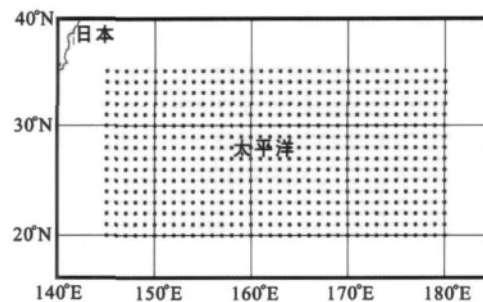


图1 研究区域站点分布

Fig. 1 Spatial distribution of stations in region of interest

图2为经过处理的2007~2009年每年的温度数据以及海表面温度数据,总计538 632组。从图中可以看出数据基本涵盖了各个温度段。

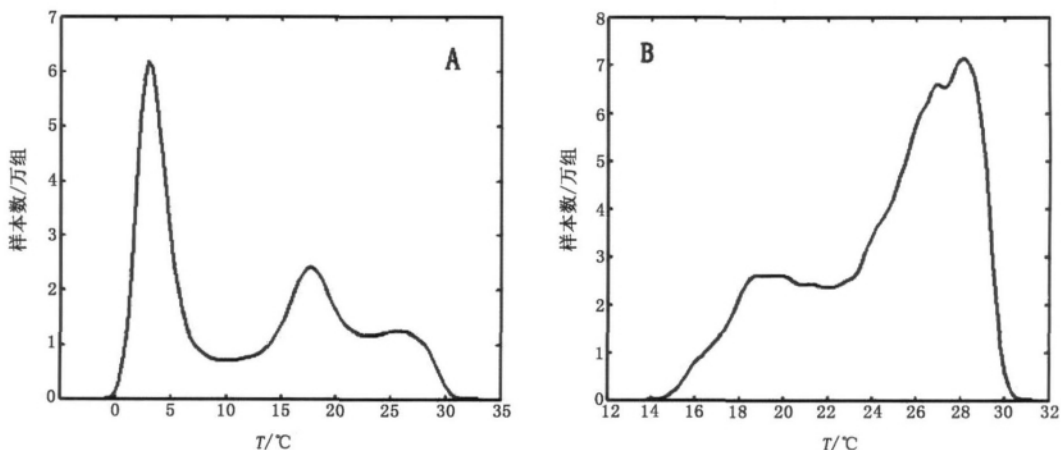


图2 2007~2009年样本温度分布:各组样本的温度(A);各组样本所处经纬度的海表面温度(B)

Fig. 2 Frequency of sample temperature in period 2007-2009 (A) temperature of each sample (B)

sea surface temperature of each sample

2 BP神经网络模型的建立及隐含层神经元数目选择

人工神经网络(artificial neural network, ANN)是20世纪80年代迅速发展起来的一门非线性科学,其中反向

传播网络(back-propagation neural network),即BP神经网络,是一种具有逼近非线性映射关系的神经网络^[9-11],我们利用Matlab 2010b软件进行了BP网络设计。网络的训练函数为自适应Levenberg-Marquardt算法,即LM算法。该算法可以使中等大小的神经网络学习时间更短,

并且收敛速度很快^[12]。LM-BP 神经网络模型输入参数为海表面温度、深度、经度、纬度, 输出参数为温度。其基本原理可以表示为式(1):

$$F(SST, Depth, Lon, Lat) \rightarrow T \quad (1)$$

式中: F 表示输入参数与输出参数的映射关系, SST 表示样本所处经纬度的海表面温度, $Depth$ 表示样本所在深度, Lon 与 Lat 分别表示样本的经度与纬度, T 表示样本温度值。图 3 为本研究所采用的具有 4 个输入参数与 1 个输出参数的含有 n 个神经元的单隐含层神经网络示意图。为了使模型可以快速收敛, 隐含层的非线性传递函数为 S 型正切函数(tan-sig)。

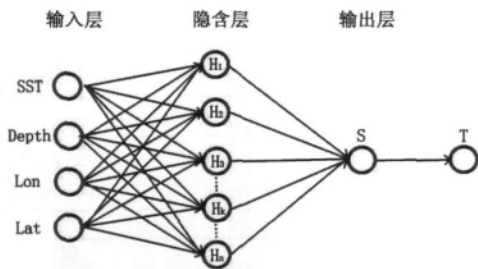


图 3 LM-BP 神经网络结构

Fig. 3 Structure of LM-BP neural network

图中: 隐含层函数为 tan-sig; 输出层函数为线性函数; T 为

权重; b 为偏差; k 为 1 至 n 中的某一个数; 圆形代表参数神经元。

在神经网络模型中神经元的个数选择是影响模型输出的重要因素。尽管神经元个数的增加可以使得模型能更快速的拟合, 但是同时也会带来模型的不稳定, 导致使用非建模数据模拟时产生拟合度变差。因此, 为了找到最优的神经元个数, 我们采取枚举法进行最优值选取。我们在 2007 年数据集中随机选取了某一个月的数据建立 LM-BP 神经网络, 并用 2008 年与 2009 年同一个月的数据作为检验。检验标准为模拟结果与实际结果的均方根差 (root-mean-square error, RMSE) 和 Pearson 相关性系数 R (以下简称相关性系数)。其计算公式为式(2)与式(3)。RMSE 越小, R 越接近 1, 则代表模型越接近实际映射。

$$RMSE = \sqrt{\frac{1}{N} \sum_{p=1}^N (T_1 - T_2)^2} \quad (2)$$

$$R = \frac{\sum_{p=1}^N (T_1 - \bar{T}_1)(T_2 - \bar{T}_2)}{\sqrt{\sum_{p=1}^N (T_1 - \bar{T}_1)^2 \sum_{p=1}^N (T_2 - \bar{T}_2)^2}} \quad (3)$$

其中: N 为样本数; T_1 为实测温度值; T_2 为模拟温度值; \bar{T}_1 为实测温度平均值; \bar{T}_2 为模拟温度平均值。模拟结果如图 4 所示。

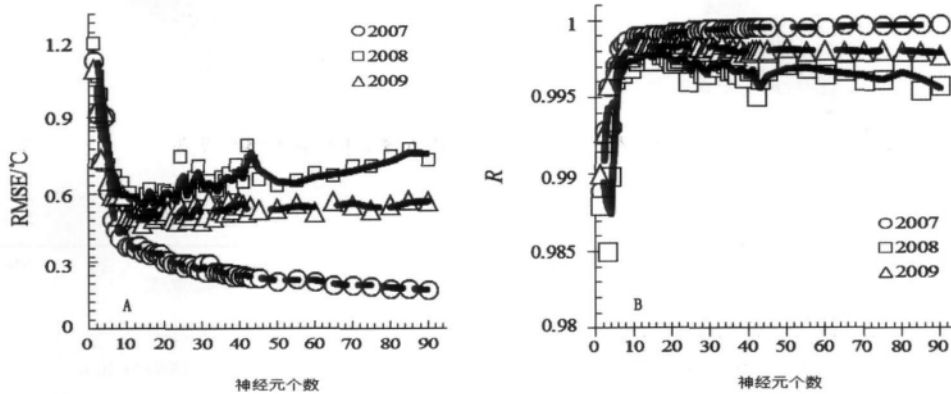


图 4 不同隐含层神经元个数的神经网络模型比较: (A) 均方根误差 RMSE; (B) 相关性系数 R

Fig. 4 Comparison of networks with different number of neurons in hidden layer (A) RMSE (B) R

从图 4 可以看出, 随着神经元数目的增加, 作为建模参数的 2007 年数据的 RMSE 在逐渐降低, 其相关性系数 R 也不断接近 1。但是作为检验参数的 2008 年与 2009 年数据的 RMSE 则表现为先降低, 随后小幅波动且缓慢上升的趋势。这两年数据的相关性系数 R 表现为先接近 1, 随后小幅波动且缓慢下降的趋势。经过比较, 我们选择的神经元个数为 14。

3 LM-BP 神经网络模拟海水温度

通过神经元选择试验, 我们最终确定建立一个具有 14 个隐含层神经元的 3 层 LM-BP 神经网络模型, 输入神经元为 4 个(海表面温度 SST、深度 Depth、经度 Lon、纬度 Lat), 输出神经元为 1 个(温度 T)。我们利用 2007 年的

数据作为建模参数, 建立了 2007 年 12 个月的海水温度模型, 并使用 2008 年与 2009 年数据进行模拟与检验, 结果见表 1。

4 模拟结果分析

4.1 总体样本比较

从表 1 可以看出 LM-BP 神经网络模型模拟的结果较好, 2008 年和 2009 年相关性系数 R 的均值分别为 0.996 8 和 0.996 5, 表明模拟结果与实际温度有显著的相关性。2008 年和 2009 年 RMSE 均值分别为 0.714 0 °C 与 0.761 5 °C, 较 2007 年 RMSE 均值 0.452 4 °C 偏大, 这主要是模拟数据和建模数据的差异性造成的。我们选择 1 月、4 月、7 月、10 月代表四个季节, 分别进行了回归分析。每个月的样本

数均为 14 962 个(图 5 ~ 图 8)。从图 5 ~ 图 8 可以看出,总体拟合结果较好。其中 2008 年与 2009 年的 1 月模拟效果

相对最好,2008 年与 2009 年的 10 月模拟效果次之,而 2008 年 4 月与 2009 年 7 月的模拟结果相对差些。

表 1 模拟结果

Tab. 1 Results of simulation

月份	2007 仿真结果		2008 模拟数据		2009 模拟数据	
	RMSE/°C	R	RMSE/°C	R	RMSE/°C	R
1	0.369 3	0.998 9	0.581 5	0.997 6	0.512 7	0.998 4
2	0.365 7	0.998 9	0.616 7	0.997 2	0.571 1	0.997 5
3	0.373 9	0.998 9	0.658 2	0.996 1	0.561 2	0.997 5
4	0.375 2	0.999 0	0.893 7	0.995 2	0.525 5	0.998 0
5	0.424 5	0.998 8	0.784 7	0.995 9	0.795 7	0.996 3
6	0.531 9	0.998 3	0.698 4	0.997 2	0.940 4	0.994 3
7	0.558 0	0.998 2	0.820 5	0.996 2	0.892 2	0.995 9
8	0.464 0	0.998 8	0.932 4	0.996 2	0.872 8	0.996 0
9	0.481 1	0.998 7	0.793 3	0.996 6	0.838 0	0.996 7
10	0.469 5	0.998 7	0.639 6	0.997 7	0.835 5	0.996 7
11	0.507 6	0.998 3	0.604 5	0.997 7	0.929 5	0.995 0
12	0.508 1	0.998 2	0.544 8	0.997 9	0.863 3	0.995 1
均值	0.452 4	0.998 6	0.714 0	0.996 8	0.761 5	0.996 5

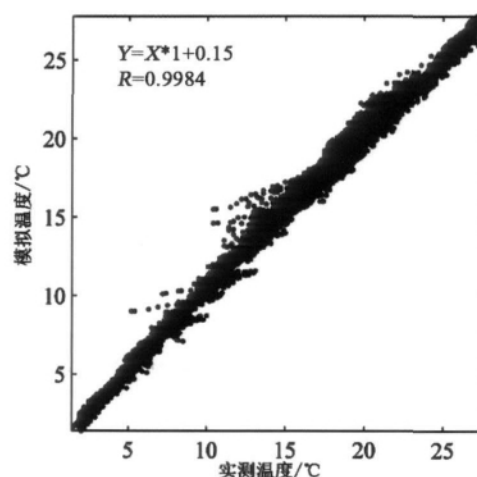
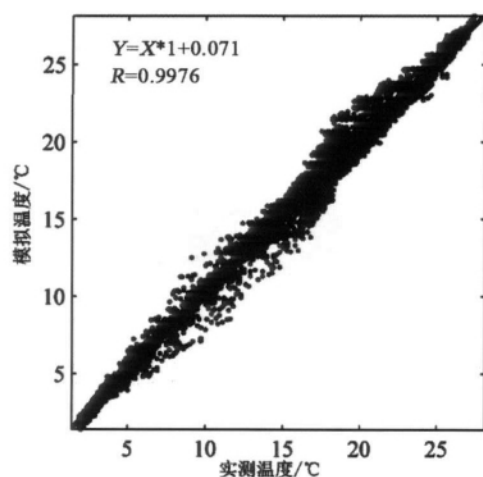


图 5 2008 年(左)与 2009 年(右)1 月份模拟数据与实测数据比较

Fig. 5 Comparison of simulation and measured data in Jan. , 2008(left)/2009(right)

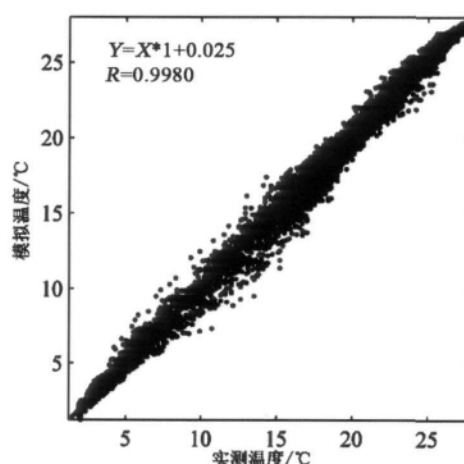
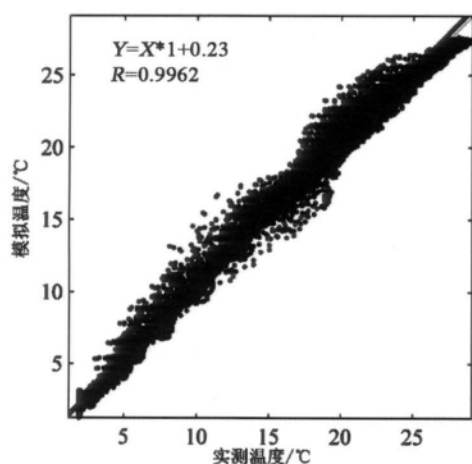


图 6 2008 年(左)与 2009 年(右)4 月份模拟数据与实测数据比较

Fig. 6 Comparison of simulation and measured data in Apr. 2008(left)/2009(right)

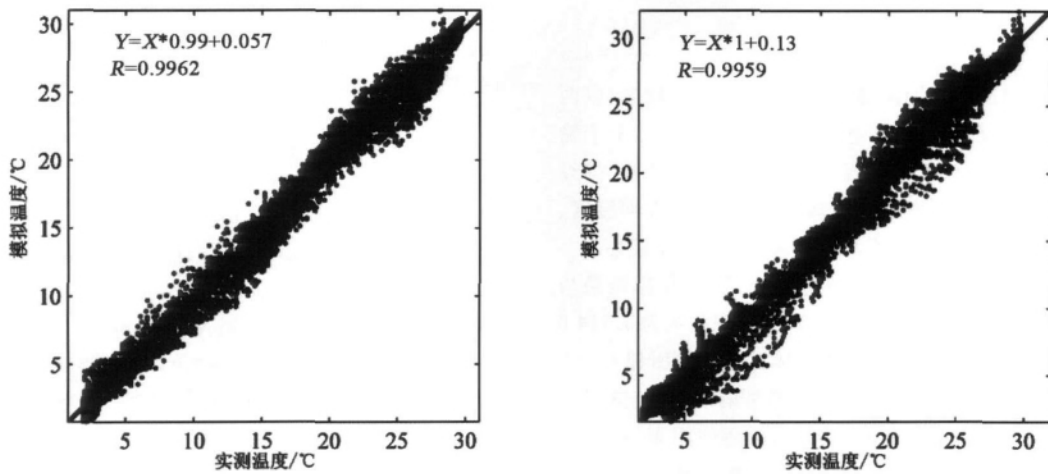


图 7 2008 年(左)与 2009 年(右)7 月份模拟数据与实测数据比较
 Fig. 7 Comparison of simulation and measured data in Jul. 2008(left)/2009(right)

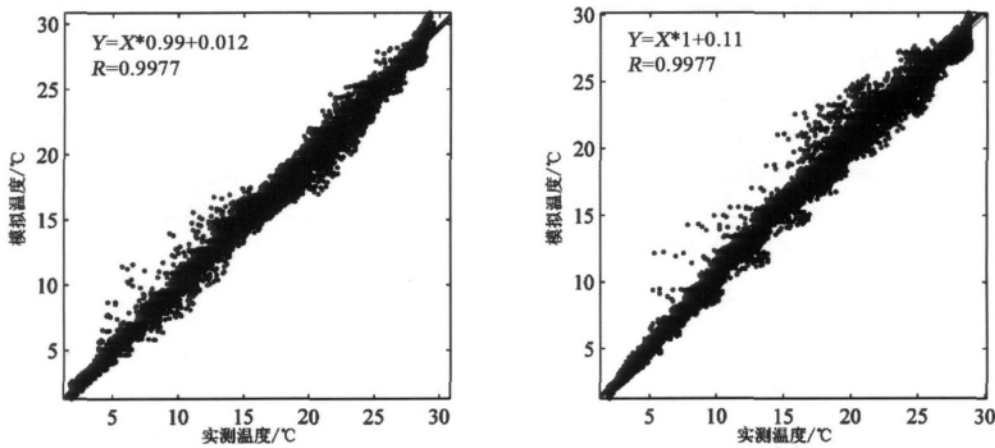


图 8 2008 年(左)与 2009 年(右)10 月份模拟数据与实测数据比较
 Fig. 8 Comparison of simulation and measured data in Oct. 2008(left)/2009(right)

4.2 单个样本比较

通过随机抽样的方式,我们在 2008 年与 2009 年全年数据中各抽取了 1 个站点数据,对该站点的模拟结果与原始数据进行了比较(图 9)。所抽取的站点分别为 2008

年 12 月站点(31°N, 167°E)和 2009 年 7 月站点(30°N, 147°E),从图 9 可以看出,尽管模拟结果与原始数据在部分深度仍有一定差距,但总体效果较好。

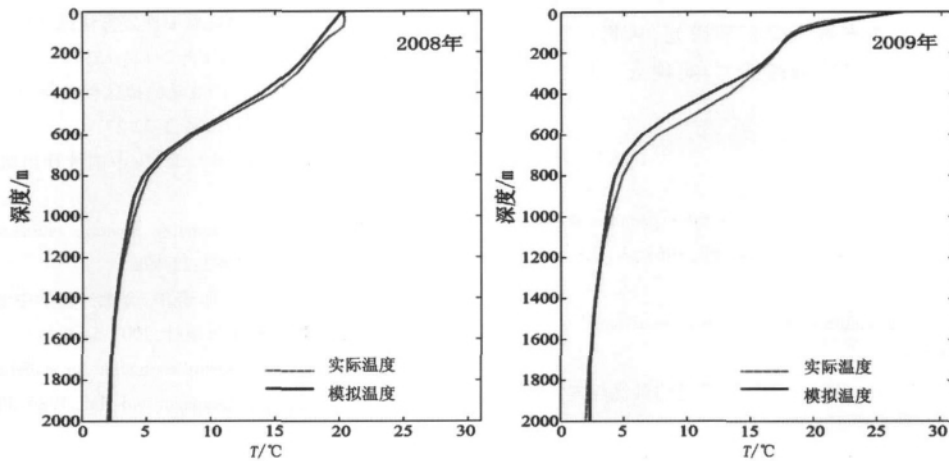


图 9 单个站点的样本实测数据与模拟结果比较
 Fig. 9 Comparison of simulation and measured data at single station

5 结论

本次研究通过采用全球 Argo 网格化产品数据,以西北太平洋为研究区,建立了包含 4 个输入神经元、1 个输出神经元以及具有 14 个神经元的单隐含层 LM-BP 神经网络模型。在模拟研究中,通过输入站点的海表面温度、经纬度和深度数据,得到了该站点某一深度的温度值。

根据均方差与 Pearson 相关性系数检验,模拟结果与原始数据拟合度较高,总体均方差均值 2008 年为 0.714 0 °C,2009 年为 0.761 5 °C。此误差接近 Chu 等所建立的 0~600 m 深度范围的水温垂直结构模型的误差(均方根误差均值 0.72 °C),而与国内学者徐玉涓等使用 MODAS 系统模拟不同区域海面以下 100 m 处深度误差(均方根误差均值为 1~1.8 °C)和修树孟等的模拟 0~50 m 深度范围的误差($RMSE = 0.90$ °C)相比较更低。

本研究所模拟深度范围(0~2 000 m)均超过之前所提的研究范围,而且预测时间跨度也更大,这说明本文所建立的基于 LM-BP 神经网络的 Argo 数据西北太平洋海水温度模型是可行的。由于影响海水温度的因素有很多,海水温度场的变化机制比较复杂,模拟所得的结果与实际数据仍有一定差距。因此加强影响海水温度因素的研究和 LM-BP 神经网络模型的优化是下一步工作的重点。

参考文献:

- [1] 鲍献文,万修全,高郭平等.渤海、黄海、东海 AVHRR 海表温度场的季节变化特征[J].海洋学报,2002,24(5):125-133.
- [2] CARNES M R, MITCHELL J L. Synthetic temperature profiles derived from Geosat Altimetry: Comparison with air-dropped expendable bathythermograph profiles[J]. Journal Geophysical Research, 1990, 95(c10): 17979-17992.
- [3] CHU P C, FAN C, LIU W T. Determination of vertical thermal structure from sea surface temperature[J]. Journal of Atmospheric and Oceanic Technology, 2000, 17(7): 971-979.
- [4] GUINEHUT S, TRAON P Y L, LARNICOL G, et al. Combining Argo and remote-sensing data to estimate the ocean three-dimensional temperature fields—a first approach based on simulated observations[J]. Journal of Marine Systems, 2004(46): 85-98.
- [5] GONZ LEZ-POLA C, FERN NDEZ-D AZ J M, LAV N A. Vertical structure of the upper ocean from profiles fitted to physically consistent functional forms[J]. Deep-Sea Research I, 2007, 54: 1985-2004.
- [6] 徐玉涓,刘春笑,吴振华. MODAS 试验数据统计分析[J]. 海洋测绘, 2009, 29(6): 52-54.
- [7] 修树孟,张钦,逢爱梅. 卫星遥感 SST 反演海水温度垂直剖面的方法研究[J]. 遥感应用, 2009(5): 73-76.
- [8] 池建军,孙祥年,骆永军. 卫星遥感 SSH 数据在水下温度垂直分布反演中的应用研究[J]. 海洋预报, 2010, 27(1): 59-61.
- [9] 袁曾任. 人工神经网络及其应用[M]. 北京: 清华大学出版社, 1999: 66-68.
- [10] HECHT-NIELSEN R. Theory of the back-propagation neural network[C]. IJCNN, International Joint Conference on Neural Networks, 1989, 1: 593-605.
- [11] DAGORN L, PETIT M, STRETTA J M. Simulation of large-scale tropical tuna movements in relation with daily remote sensing data: The artificial life approach[J]. Biosystems, 1997, 44: 167-180.
- [12] HAGAN M T, MENHAJ M B. Training feed-forward networks with the marquardt algorithm[J]. IEEE Transactions on Neural Networks, 1994, 5(6): 989-993.

(上接第 554 页)

动漫宣传片等方式向公众普及、推广应对海洋灾害的相关知识,使得广大公众能够对海洋灾害有一个科学的、全面的认识,增强其防灾减灾的科学意识。使其主动地采取适合自身条件的海洋灾害风险管理措施,从真正意义上提高全社会海洋灾害风险管理能力,实现人与自然的和谐、统一。

参考文献:

- [1] LIND R C. Flood control alternatives and the economics of flood protection[J]. Water Resources Research, 1967, 3(2): 345-357.
- [2] LUAN C. Insurance premium calculations with anticipated utility theory[J]. ASTIN Bullentin, 2001, 7: 27-39.
- [3] 杨朝军,肖彦明,徐为山. 基于均值方差模型的最优巨灾保险计划[J]. 上海交通大学学报, 2006, 4: 622-640.
- [4] 欧阳资生. 极值理论: 巨灾理论的统计理论基础[J]. 统计与决策, 2006, 7: 13-14.
- [5] 谭英平. 非参数密度估计在个体损失分布中的应用[J]. 统计研究, 2003, 8: 40-44.
- [6] 梁来存. 我国粮食保险纯费率厘定方法的比较与选择[J]. 数量经济技术经济研究, 2011, 2: 124-133.
- [7] 钟甫宁,邢鹏. 粮食单产波动的地区性差异及对策研究[J]. 中国农业资源与区划, 2004, 3: 22-27.
- [8] 吴喜之. 非参数统计[M]. 北京: 中国统计出版社, 1999: 13-200.
- [9] PRAKASA RAO. Nonparametric function estimation[M]. London: Academic Press, 1983: 22-498.
- [10] 李竹渝,鲁万波,龚金国. 经济、金融计量学中的非参数估计技术[M]. 北京: 科学出版社, 2007: 5-69.
- [11] SILVERMAN B W. Density estimation for statistics and data analysis[M]. London: Chapman and Hall, 1986: 30-367.