

利用 GMDH 方法作风速预报

潘长明 叶卓文

(天津海洋测绘研究所, 300061)

提 要

利用 GMDH 方法对渤海中、西部海面 1991 年 2 月逐日 08 时风速作了 3 天内的预报。预报准确率较高。

关键词: GMDH 建模 预报

引 言

目前,用于单站天气要素预报的客观预报方法,主要是 MOS 预报和 PP 预报,而 MOS 预报和 PP 预报在建模过程中,存在着所需资料样本大,因子信息量小,并且要经常更改模式等不足。本文使用的 GMDH 方法可以较好地克服以上不足,并取得了明显的效果。

GMDH(The group method of data handling, 简称 GMDH, 即数据处理组合法)方法,由原苏联乌克兰科学院的 A. G. Ivakhnenko 等人于 70 年代初,根据控制论中自组织原理首先提出的^[1]。该方法一问世,便受到人们的关注。进入 80 年代中期,我国也有许多学者,特别是工业自动化和自动控制领域里的学者,开始关注 GMDH 方法的研究和应用。随着应用研究的深入, GMDH 方法的应用范围,也由冶炼、化工、医学等领域,逐步扩展到气象、农业、水利等领域。

利用 GMDH 方法作风速短期预报建模,主要有以下特点:

(1) 可以根据较少的样本资料,通过目标函数,确定预报因子与预报对象之间的输入与输出关系。

(2) 通过自组织原理识别和确定预报模型结构。

(3) 在采用动态样本建模的情况下,预

报模式可以连续使用,而不需要经常更改。

(4) 由于采用串行计算,计算量较小。

1 GMDH 方法简介

假设 $y=f(X) \in R^1, X \in R^m, m$ 为自然数。若 $f(X)$ 在 $X=X_0$ 处连续且存在各阶偏导数,则可把 $f(X)$ 在 $X=X_0$ 点展开成 Volterra 无穷级数:

$$Y = a + \sum b_i x_i + \sum \sum c_{ij} x_i x_j + \sum \sum \sum d_{ijk} x_i x_j x_k + \dots \quad (1)$$

式中 $a, b_i, c_{ij}, d_{ijk}, \dots$ 为待定系数, m 为自然数。式(1)又称为 Kolmogorov-Gabor 多项式,简称 K-G 多项式。

不难看出,当 m 取值很大时,估算式(1)中系数所需的数据量和计算量都是惊人的。例如:当 $m=10$, 式(1)右边截取到 3 次方项时,需求解待定系数的个数为 $1+C_{10}^1+C_{10}^2+C_{10}^3=176$, 如果将式(1)右边多项式截取到 5 次方项,需求待定系数的个数则迅速增加到 638 个。可想而知,当 m 取值和多项式截取次数同时增大时,所需求解待定系数的个数是十分惊人的。人们常把该现象称为维数灾难^[2]。

为了解决维数灾难问题, GMDH 方法利用低元、低次多项式的不断组合的方法逼近式(1)。假设有 m 个自变量 $X_1, X_2, \dots, X_m$, 将其中任意两个自变量 X_i, X_j 与变量 Y 的关系通过一个低次多项式联接起来,构成一

个回归方程,比如一个二次多项式:

$$Y = A + BX_i + CX_j + DX_i^2 + EX_j^2 + FX_iX_j \quad (2)$$

其中 A, B, C, D, E, F 为待定参数。如此类推, m 个自变量两两组合,共可得到 $m(m-1)/2$ 个二次回归方程。通过 $m(m-1)/2$ 个回归方程可以产生相应的 Y 的输出,再按照一定的准则,从 $m(m-1)/2$ 组 Y 的输出中,选择若干组作为新的自变量,为下一层的计算提供输入,产生新的 Y 值的输出,从而构成对式(1)的逼近。以此类推,直到求得满意的 Y 值为止。GMDH 方法的算法网络如图 1 所示。图中各层的方程形式皆为二次多项式,但各层的方程数和中间变量的值不同。

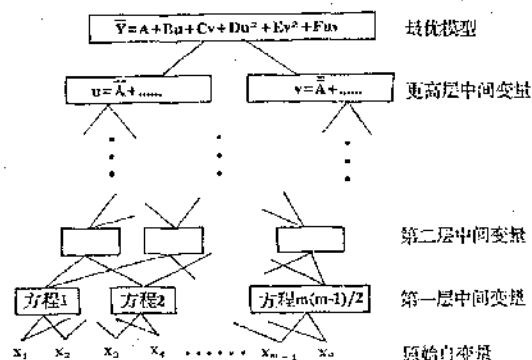


图1 GMDH 方法算法网络

在实际应用中,GMDH 算法可分为以下步骤:

(1)数据资料的预处理

资料预处理的方法不一,一般包括相关分析、样本离散度分析、中心化处理等。在我们的研究中,是采用下列方法对样本资料作预处理的。

$$X'_i = \frac{X_i - \bar{X}}{\bar{X}} \quad (3)$$

其中, $\bar{X} = \sum |X_i|/n$, X_i 表示某一因子, X'_i 为 X_i 经预处理后的值。

(2)把样本数据分成 A、B 两组

将样本数据分成两组是 GMDH 方法的重要环节,划分方法因需要而定。我们经过试

验认为,A 组占样本总量的 70%,B 组占样本总量的 30%左右为宜。参见图 2。

Y	X				
	X_1	X_2	$\dots$	X_m	
y_1	$x_{1,1}$	$x_{1,2}$	$\dots$	$x_{1,m}$	拟合数组 A
y_2	$x_{2,1}$	$x_{2,2}$	$\dots$	$x_{2,m}$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
y_{n-1}	$x_{n-1,1}$	$x_{n-1,2}$	$\dots$	$x_{n-1,m}$	
y_{n-1+1}	$x_{n-1+1,1}$	$x_{n-1+1,2}$	$\dots$	$x_{n-1+1,m}$	检查数组 B
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
y_n	$x_{n,1}$	$x_{n,2}$	$\dots$	$x_{n,m}$	

图2 GMDH 算法样本数组的划分

(3)选定目标函数,对所有变量进行两两组合

目标函数的形式及个数根据建模的特点而定,一般情况下可采用以下形式:

$$Y = f(X_i, X_j) = A + BX_i + CX_j + DX_i^2 + EX_j^2 + FX_iX_j$$

式中 A, B, C, D, E, F 由图 2 中的拟合数组 A 估计,在本文的研究中采用最小二乘求估系数。

(4)选择中间变量

用检查数组 B 数据按照选定的准则对所有拟合方程分别进行评估,筛选较好的模型输出构成中间变量,作为下一层的输入继续计算。本文中我们采用均方误差准则对中间模型进行评估。

$$r_j^2 = \sum (Y_i - Y_{ij})^2 / \sum Y_i^2 \quad (4)$$

其中 Y_{ij} 为 Y_i 的估值, $j=1, 2, \dots, C_m^2$ (m 为自变量个数)。

(5)若该层最小均方误差 $r_{\min}^2$ 比上一层小,则以该层模型的输出作为下一层的输入,转向步骤 3 继续计算,否则转向步骤 6。

(6)用上一层最好的模型作为最终模型(参见图 3),分析模型输出指标,若满足要求则停机,否则调整各层中间模型选择的比例,

重新计算。

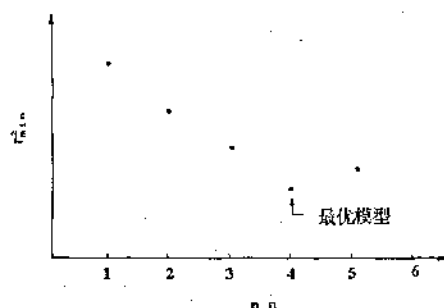


图3 GMDH算法停止准则示意图
 $r_{\min}$ 为某层最小均方误差, m 为计算层数

2 预报试验

2.1 样本资料的构造

我们知道,某一地区或某个单站短期天气要素的变化,与其周围地区短期天气要素的变化有着密切的联系。根据这一特点,在作渤海海面风速预报时,我们以该地区为中心,选择了秦皇岛、熊岳、塘沽、烟台、成山头、羊角沟、锦西、大连、旅顺等周围地区地面和高空常规气象要素作为预报因子,把渤海海面风速作为预报对象,利用GMDH方法建模预报。

在确定样本时间长度的问题上,我们也作了多种试算,最后采用了动态样本建模的方案。动态样本的构造是这样实现的,利用最近30天的要素实测值构成最初样本,对第31天的预报对象作预报,完成预报后,把第31天的实测要素资料即时充实到建模样本中,对原样本进行更新,但原样本的大小保持不变。再利用新样本对第32天的预报对象作预报,以此类推,逐日预报,逐日更新样本。预报结果表明,利用动态样本建模预报,准确率较高。

2.2 预报结果

我们分别以渤海中部(8号平台)和西部(A平台)海面风速为预报对象,利用1991年1月逐日08时资料为样本,对该年2月逐日

08时的风速分别作了24小时、48小时、72小时的预报。附表为渤海中部海面风速预报结果(西部预报结果略)。

附表 渤海中部海面1991年2月逐日08时风速预报(单位: $m \cdot s^{-1}$)

日期	实测值	24小时		48小时		72小时	
		预报值	预报误差	预报值	预报误差	预报值	预报误差
1	8.0	7.8	+0.2	6.0	+0.2	6.9	+1.1
2	4.0	3.4	+0.6	4.8	-0.8	11.0	-7.0
3	6.0	8.5	-2.5	6.1	-0.1	7.2	-1.2
4	9.0	7.0	+2.0	7.4	+1.6	9.4	-0.4
5	11.0	9.2	+1.8	7.3	+3.7	9.6	+1.4
6	11.0	10.8	+0.2	7.7	+3.3	9.8	+1.2
7	9.0	7.6	+1.4	8.2	+0.8	15.5	-6.5
8	4.0	1.9	+2.1	6.2	-2.2	8.9	-4.9
9	4.0	4.3	-0.3	4.7	-0.7	4.3	-0.3
10	8.0	5.2	+2.8	6.6	+1.4	6.4	+1.6
11	10.0	4.8	+5.2	7.6	+2.4	11.5	-1.5
12	5.0	4.8	+0.2	5.5	-0.5	7.9	-2.9
13	4.0	5.2	-1.2	4.4	-0.4	5.3	-1.3
14	4.0	4.8	-0.8	3.1	+0.9	4.5	-0.5
15	11.0	15.5	-4.5	14.4	-3.4	10.4	+0.6
16	6.0	8.8	-2.8	8.9	-2.9	7.7	-1.7
17	6.0	7.0	-1.0	6.7	-0.7	5.9	+0.1
18	6.0	5.6	+0.4	5.9	+0.1	5.9	+0.1
19	12.0	11.9	+0.1	9.1	+2.9	8.6	+3.4
20	14.0	14.3	-0.3	16.8	-2.8	10.1	+3.9
21	4.0	7.5	-3.5	6.9	-2.9	6.0	-2.0
22	11.0	14.9	-3.9	16.9	-5.9	11.1	-0.1
23	4.0	4.2	-0.2	5.2	-1.2	3.1	+0.9
24	3.0	7.6	-4.6	3.5	-0.5	6.6	-3.6
25	6.0	6.8	-0.8	7.5	-1.5	9.3	-3.3
26	9.0	7.7	+1.3	9.2	-0.2	7.5	+1.5
27	6.0	6.4	-0.4	7.1	-1.1	6.7	-0.7
28	5.0	5.3	-0.3	5.0	0.0	4.3	+0.7
平均	7.14	7.45	-0.31	7.45	-0.31	7.94	-0.80

(下转第23页)

(上接第 35 页)

3 几点结论

3.1 用 GMDH 方法制作要素预报,突出的优点是选择因子客观性强,信息量大,克服了统计预报中选择因子人为因素多、信息量少的不足。

3.2 动态样本建模的好处不仅能反映短期天气要素预报的特性,而且还能把短期要素变化的信息及时地吸收到模式中来,这体现了用天气学理论制作天气预报的思想。

3.3 动态样本建模的另一个显著优点是,不需要随着季节的变化而更改模式,这是目前统计预报方法难以做到的。

参考文献

- 1 孟含明,杨百厚,李宝泽. 数据处理组合方法(GMDH)及其应用. 控制与决策,1987,第1期,45—51.
- 2 A. G. Ivakhnenko. Heuristic Self-organization in Problem of Engineering Cybernetics. Automatica, 1986, Vol. 6, 207—219.

GMDH for Forecasting Wind-Velocity

Pan Changming Ye zhuowen

(Tianjin Institute of Hydrographic Surveying and Charting)

Abstract

The group method of data handling (GMDH) is used to forecast wind-velocity at eight o'clock each day within three days for the middle and western areas of Bohai sea in Feb. 1991. It's accurate rate is higher.

Key Word: the group method of data hendling modelling forecast