

基于多级相似-差额方法制作乡镇气温预报

连志鸾 李国翠 卞 韬 高连山

(河北省石家庄市气象局, 050081)

提 要: 应用 2005 年以来的 ECMWF 资料、地面常规观测资料和分布在石家庄范围内的 184 个自动站气温资料, 基于多级相似-站际间气温差额预报方法, 制作石家庄 204 个乡镇点 24 小时最高、最低气温预报。为便于该方法的业务运行, 建立了 ECMWF、地面常规观测资料、自动站气温等历史资料数据库, 并实现实时定时追加。该方法在实际业务试验中取得初步的效果, 2007 年 6—9 月乡镇点 24 小时最高、最低气温预报平均绝对误差分别为 1.6°C 、 1.5°C , 绝对误差 $\leq 2^{\circ}\text{C}$ 准确率平均为 75.5%。通过误差成因分析, 对该方法的优缺点做了较为全面、客观的讨论, 并提出了进一步完善的方向。

关键词: 场相似 要素相似 气温差额 乡镇气温预报

引 言

精细化预报是天气预报业务的发展方向, 从 1980、1990 年代以来, 各地大量地开展了空间分辨率到县的精细预报技术研究。对气温预报而言, 山西^[1]、青海^[2]、河北^[3]、贵州^[4]、北京^[5]等基于 T63、T106、T213、MM5 等一种或数种数值预报产品, 结合常规观测资料和县站气温实况, 应用 PPM、MOS、KF、神经网络等方法进行客观分县预报试验, 输出 24 小时、1~3 天、1~5 天内各时段气温预报, 满足了当时的气温预报业务需求。近年来, 社会需求不断对预报精细化程度提出了更高的要求, 特别是乡镇自动站观测网的建设, 获得了各乡镇点第一手的实况观测资料, 这为开展乡镇气温预报业务提供了可能。

天气预报精细化, 一是要有高质量、高精度的数值预报产品, 二是要有与预报的时空尺度相匹配的高密度的监测资料作为基础。目前地市级可接收到的数值预报产品包括国内的 T213、GRAPES、MM5 等, 国外的 ECMWF、日本、德国的部分数值预报产品。“三站四网”建设使自动站资料更规范、更丰富。石家庄从 2001 年开始进行自动站建设, 到 2005 年全市 184 个自动站建设全部完成, 覆盖了全市 90% 的乡镇。到 2007 年已完整业务运行 2 年多时间, 积累了一定数量的资料。

2007 年石家庄气象台尝试制作全市 204 个乡镇点的气象要素预报。充分利用现有条件, 研究本地客观预报方法, 为制作精细准确的乡镇预报提供技术支撑就成了当务之急。乡镇天气预报要素中, 24 小时最高、最低气温预报是目前县级电视天气预报节目对公众

发布的一项基本气象要素,气温又是精细化预报相对容易的一项,以此为突破口,通过反复比较试验,并参考各地制作分县、分市气温预报方法的经验,尝试用一种相似-差值方法制作乡镇气温预报,取得了初步的效果。

1 资料选取及处理

相似预报方法要从大量历史样本中找相似,因此就涉及到数值预报产品、自动站资料、实况天气要素资料 3 个历史资料库的建立和实时定时追加。

(1) 数值预报产品的选取

应用 2005—2006 年石家庄本站逐日最高气温、最低气温分别与 T213 (850hPa 和 925hPa)、ECMWF (850hPa)、日本 (850hPa) 24 小时气温预报求相关,比较其相关系数的大小。相关程度越高者,认为其 24 小时气温趋势预报与实况气温的变化趋势越接近,其预报能力强。结果发现 ECMWF 08 时的 24 小时 850hPa 气温预报与实况最低、最高气温的相关系数分别为 0.930 和 0.913,ECMWF 20 时的 24 小时 850hPa 气温预报与实况最高气温的相关系数为 0.896。日本 08 时 24 小时 850hPa 气温预报与实况最低、最高气温的相关系数分别为 0.906 和 0.932,日本 08 时 36 小时 850hPa 气温预报与实况最高气温的相关系数为 0.913。总体看,ECMWF、日本数值预报高于 T213、MM5 对地面气温的趋势反映能力。另外考虑客观方法的业务系统实现,故采用 ECMWF 数值预报产品。

建库:利用 2005 年以来的 ECMWF 历史资料,选定区域、要素进行处理,建立 ECMWF 高空要素格点场历史数据库。对实时 ECMWF 资料,编制定时处理程序,对所选定区域、要素场资料每天定时处理入库。

(2) 自动站资料处理

定时从自动站资料 SQL-Server 数据库

中提取资料,生成逐时的最高、最低和定时气温文件,每天定时生成逐日的最高、最低气温。两种方法实现质量控制:一是提取系统故障信息;二是比较 1~3 小时内最高、最低气温,差值为零者,认为采集器没工作或系统故障,该时段记录舍去。故障出现在日最高、最低气温形成时段,则当日的最高、最低气温也舍去。

建库:根据上述自动站资料处理方法,建立 2005 年以来自动站资料逐日最高、最低气温及逐时气温历史数据库。

(3) 实况天气要素库:建立 2005 年以来石家庄逐日的实况天气要素库。其中包括:最高、最低气温,08 时、14 时相对湿度,夜间、白天段天空状况。天空状况包括晴、多云、阴、雨(小雨、中雨、大雨、暴雨)、雪(小、中、大雪)、雾、大风等对气温变化有明显影响的天气。

2 预报方法

(1) 基本思路

预报员认为,在相似的大气环流形势背景下所产生的天气现象也往往是相似的。基于这样的基本考虑,模仿预报员制作单点气温预报的基本思路,采用多级相似的方法,点面结合进行相似度量。首先通过形势场相似,找出与未来 24 小时形势特征相似的若干场相似日。然后,以场相似日为样本,再通过要素相似度量,找出与未来 24 小时气象要素也相似的若干相似日。计算相似日各乡镇日最高、最低气温与石家庄本站最高、最低气温的平均差值(差额)。根据未来 24 小时石家庄市气温预报分别与差额相加,即可得出未来 24 小时各乡镇气温预报。图 1 为该方法的业务系统框架。

(2) 相似度量

场相似度量应用“形、值”结合的相似系

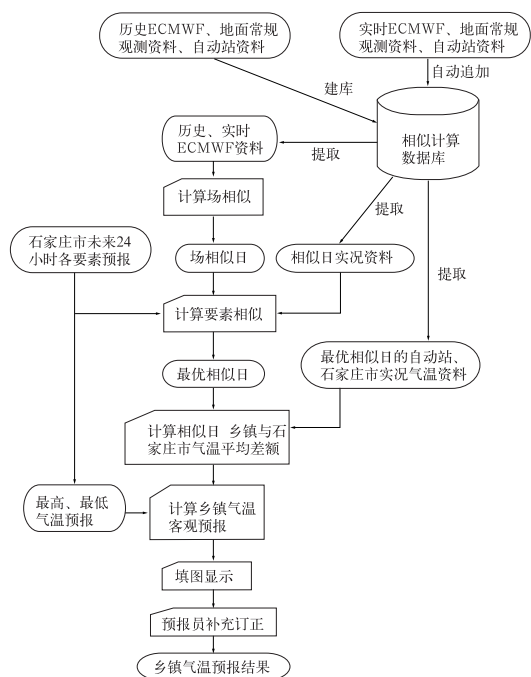


图1 基于相似-差额方法制作乡镇气温预报的业务系统实现流程图

数。其中,“形”相似判据应用“平均因子差值离散度”(式3),“值”相似判据应用“海明距离”(式2),二者线性相加综合为相似系数 $R_{i,j}$ 。 $R_{i,j}$ 既表征两个场的波形相似程度,也表征了它们的空间格点值的接近程度。该方法计算简单,也能较好地反映两个样本的相似性。场相似测度公式如下^[6-8]:

$$R_{i,j} = \frac{1}{2}(S_{i,j} + D_{i,j}) \quad (1)$$

$$D_{i,j} = \frac{1}{M} \sum_{k=1}^M |X_{ik} - X_{jk}| \quad (2)$$

$$S_{i,j} = \frac{1}{M} \sum_{k=1}^M |(X_{ik} - E_{ik}) - (X_{jk} - E_{jk})| \quad (3)$$

$$E_{ik} = \frac{1}{N} \sum_{l=1}^N X_{il}, E_{jk} = \frac{1}{N} \sum_{l=1}^N X_{jl} \quad (4)$$

要素相似的测度公式采用式(2)“海明距离”判据,此时式(2)中 k 代表各要素, M 为要素个数。

(3) 样本选择及相似计算

计算场相似时,对每一个预报日,选择该预报日前后各25天的历史样本参与相似计算。这样处理可减少运算时的历史总样本数,提高运算效率,并能有效地消除季节性误差。不同要素场选择的扫描区域有差别,以考察“场形”相似为主的500hPa高度、地面气压、风等,其相似场扫描区域定为30~35°N,105~120°E,选区域较大;以考察要素值相似为主的850hPa温度、湿度,其相似场扫描区域为35~40°N,110~115°E,选区域较小。应用式(1~4)计算场相似系数。

将计算所得场相似系数排序,取相似程度高的前45天,将其所对应的日期作为历史样本。如果未来24小时全市天空状况分布相对比较均匀,则以石家庄市为代表站,取预报日的天空状况、温度、风等与该历史样本计算要素相似系数。计算用式(2),其中 $M=9$ 。

在进行上述场相似、要素相似计算时,首先对所选样本资料进行标准化处理^[9],这样可避免量纲影响造成的相似误差。

(4) 计算差额

将要素相似系数排序,优选出相似程度高的前15天,将其对应的日期作为未来24小时的气温相似日,分别计算相似日各乡镇气温与本市气温的平均差值(式5, i 代表各乡镇),即差额。该差额反映了不同天气和天气形势下,由于各乡镇地理位置、环境因素而造成的气温差异。

$$\Delta T_i = \frac{1}{10} \sum_{k=1}^{10} (T_i - T_{石家庄}) \quad (5)$$

(5) 气温预报结果

根据主客观方法获得的未来24小时本市最高、最低气温预报结果,与上述所得差额相加,得到各乡镇点的未来24小时最高、最低气温预报。

对尚未建自动站的20个乡镇和个别因资料不全算不出客观预报结果的乡镇点的预报,用相邻乡镇点预报结果代替。

(6) 预报员订正

由于采取以点(石家庄市)代面计算要素相似,因此天空状况、降水、雾等分布不均造成的气温分布差异未完全考虑在内,需要做如下预报订正:

首先将降水、雾区预报转化到乡镇站点上,形成站点预报数据。然后根据降水出现时段、大雾持续时间等对最高、最低气温的影响程度以及与石家庄市的差异程度,设定不同的影响值(比如:降雨以中雨以上降水出现在上午,大雾以持续到中午对最高气温的影响最大)。各级影响值是根据统计结果,或由预报员经验设置。根据站点预报信息确定站点影响值,用影响值订正预报结果,为最后的预报结果。

3 预报效果检验

应用 2007 年 1—9 月的资料,通过两种预报试验对该方法进行检验。

(1) 拟合检验

把预报次日的实况最高、最低气温作为预报值,进行拟合检验。结果(表 1)表明,春季 3—5 月预报误差大于其它月份,这与春季

表 1 2007 年 1—9 月 24 小时乡镇气温预报
拟合检验误差统计

月	平均绝对 误差/℃		绝对误差≤1℃ 准确率/%		绝对误差≤2℃ 准确率/%	
	T_g	T_d	T_g	T_d	T_g	T_d
1	0.9	0.9	65.7	66.2	91.3	88.8
2	0.7	1.0	75.0	64.3	96.2	88.4
3	0.9	1.0	65.5	61.5	90.1	87.3
4	1.1	1.0	64.1	62.9	87.8	87.6
5	1.0	1.1	62.0	60.6	87.5	86.5
6	0.8	0.8	68.6	72.4	94.1	93.5
7	0.8	0.7	75.5	72.9	93.0	94.9
8	0.6	0.6	81.2	82.7	97.2	98.3
9	0.6	0.8	82.8	71.0	97.8	93.3
平均	0.8	0.9	71.2	68.3	92.8	91.0
6—9 月 平均	0.7	0.7	77.0	74.8	95.5	95.0

气温变幅大有关,符合气温预报误差的常情。1—9 月最高、最低气温平均绝对误差分别为 0.8℃和 0.9℃,绝对误差≤1℃的准确率分别为 71.2%和 68.3%,绝对误差≤2℃准确率分别为 92.8%和 91%。

(2) 实际检验

2007 年 6—9 月实际预报检验(表 2),最高、最低气温平均绝对误差分别为 1.6℃和 1.5℃,绝对误差≤1℃的准确率分别为 43%和 48.5%,绝对误差≤2℃准确率分别为 73.8%和 77.1%。

表 2 2007 年 6—9 月 24 小时乡镇气温预报
实际检验误差统计

月	平均绝对 误差/℃		绝对误差≤1℃ 准确率/%		绝对误差≤2℃ 准确率/%	
	T_g	T_d	T_g	T_d	T_g	T_d
6	1.6	1.4	40.0	48.4	72.7	78.0
7	2.2	1.7	42.4	54.1	69.6	79.9
8	1.6	1.6	40.6	42.5	71.9	68.3
9	1.2	1.3	48.8	49.0	81.0	82.2
平均	1.6	1.5	43.0	48.5	73.8	77.1

对 2007 年 6—9 月两种预报试验结果误差(表 1、2)的比较看出,实际预报检验较拟合检验平均绝对误差增长了约 1 倍,绝对误差≤1℃的准确率平均下降了 30.1%,绝对误差≤2℃的准确率平均下降了 19.8%。实际本市最高、最低气温预报的绝对平均误差分别为 1.4℃、1.2℃,而拟合预报中本市气温“预报”误差则为零。乡镇点气温预报误差随石家庄单点预报误差明显增长。

(3) 预报效果分析

将 2007 年 6—9 月逐日的实际预报绝对误差从大到小排序,筛选出预报效果最差、最优的若干日,逐日分析比较产生误差的原因,结果可归纳如下:①气温升降幅度大时,本市气温预报误差易大,导致总体预报效果差。②当有弱焚风时,形势场往往反映不明显,造成焚风天气漏报,从而导致预报效果较差。③云量、降水、雾等天气现象的空间分布不均

匀,导致部分乡镇的预报误差大,特别是后半夜和上午的天气现象分别对最低、最高气温影响最大。

计算逐日站际间预报误差变率,并通过排序筛选出误差变率大的预报日,试图对预报误差空间分布不均的产生原因进行详细分析,结果表明,焚风天气漏报、天气现象的空间分布不均仍为主因。

除上述三种情况外,大多数预报比较理想。特别是当天空状况、气温变化都比较平稳,无焚风出现时,预报效果最好。

4 结论和讨论

通过多级相似(场相似、要素相似)制作乡镇气温预报。其优点是:①形、值结合进行场相似测度。②以点(石家庄市)代面计算要素相似,使计算更为简便,也符合预报员日常预报思路,且抓住了造成气温趋势变化的主要因素,预报结果能够满足日常业务应用。实际天气预报中,石家庄本市天气及气温预报是预报员的重点考虑对象,也有一些技术方法作支撑,预报准确率相对较高,使此方法具有一定可行性。③由于本方法模仿预报员思路,容易被预报员掌握。实际业务系统设计时,在每一步都输出中间结果,便于预报员在使用中及时发现问题,有助于对该方法各处理环节的不断改善。

缺点及进一步完善的方向:①该方法首先要知道石家庄本市未来24小时天气和最高、最低气温预报,因此各乡镇点气温预报准

确率依赖于本市气温预报。误差分析表明,石家庄本市预报误差大,则会造成乡镇点气温预报误差的普遍增大;在一日中,站际间误差分布一般比较均匀。②本方法对由于天气现象分布不均所造成的气温空间分布差异,没有完全客观地考虑进去,需要通过人机交互进行订正。要通过客观方法弥补这一缺陷,尚需做大量深入、细致的工作。③本方法涉及的处理环节较多,如:历史样本的数量、场相似计算所选区域及其大小、要素相似计算所选要素及其数据处理等等,都会不同程度地影响最终预报结果,因此该方法尚需在实际业务应用中不断检验并加以完善。

参考文献

- [1] 苗爱梅,胡永祥,郭玉玺,等.以数值产品为基础的省级分县温度预报方法[J].气象,1998,24(9):41-45.
- [2] 李有宏,韦淑侠,王青川,等.青海省最高最低气温预报方法研究[J].气象,2003,29(7):34-37.
- [3] 张国华,张江涛,赵玉广,等.河北省11个地级市极端气温的分析和预报[J].气象科技,2003,31(5):273-279.
- [4] 陈百炼.降水温度分县客观预报方法研究[J].气象,2003,29(7):48-51.
- [5] 胡江林,张德山,王志斌,等.北京地区未来1~3天昼夜气温预报模型[J].气象,2005,31(1):67-68.
- [6] 钟元.多元判据综合评估中期天气客观相似预报模式[J].气象,2003,29(4):3-8.
- [7] 黄嘉佑著.气象统计分析与预报方法[M].北京:气象出版社,2004:181-183.
- [8] 阎惠芳,李社宗,黄跃青,等.常用相似性判据的检验和综合相似系数的使用[J].气象科技,2003,31(4):211-215.
- [9] 李湘阁著.农业气象统计[M].西安:陕西科学技术出版社,1996:99-100.