

基于 GIS 的气温和降水推算方法研究

陆忠艳¹ 袁子鹏¹ 蔡福² 吴曼丽¹ 刘文明¹

(1 沈阳中心气象台, 2 中国气象局沈阳大气环境研究所, 沈阳 110016)

摘要 针对开展乡镇天气预报对高精度逐日气象要素输入值的需求, 以辽宁地区为例, 选用克立格法(Kriging)、距离权重反比法(IDW)、带高度梯度订正的距离权重反比法(GIDW)及样条函数法(Spline)4种插值方法, 进行有限气象站点1~12月逐日气象要素空间插值方法研究并对估值进行检验。结果表明: 对温度而言, GIDW方法估值精度较高, 插值结果分布趋势也较为接近实际站点的分布; 对降水而言, IDW估值精度高于其他插值方法, 更适用于日降水量的空间插值。

关键词 气温 降水 空间插值 估值检验

引言

辽宁省东西、南北跨度较大, 各地气候差异明显, 部分地区乡镇间气象要素梯度较大, 各种气象灾害频繁发生, 给农村经济社会发展带来严重影响, 给农民生命财产造成严重损失。开展乡镇天气预报对农村防灾减灾具有重要意义, 由于观测资料缺乏, 制约精细化预报服务的开展。对于短时间序列气象要素进行空间扩展方法研究成为必要的基础性工作。本文以辽宁地区为例, 通过插值提供建立乡镇天气预报客观预报模型的观测资料, 对多种插值方法进行了比较, 旨在从中选取实现短时间序列气象要素空间化的最优方法, 为开展乡镇天气预报提供合理、可靠的基础数据。

1 插值方法介绍

地理信息系统(GIS)中获得的空間数据往往是离散点的形式。离散的点数据常是通过空间采样点进行观测获得, 无法对空间所有点进行观测, 但可以设置一些关键的样本点, 这些样本点的观测值能反映空间分布的全部或部分特征, 然后利用空间内插方法来获取未采样点的值。人们比较熟悉和用得较多的是点内插, 一般的空间内插就是指内插。

点内插根据其基本假设和数学本质可分为几何方法、统计方法、空间统计方法、函数方法、随机模拟方法、物理模型模拟方法和综合方法^[1~3]。内插法都是基于假设进行的, 即空间位置上越靠近的点, 越有可能具有相似的特征值, 离得越远的点, 其特征值相似的可能性越小。

2 插值方法比较

各种插值方法各有优缺点, 没有绝对最优的, 必须根据数据的特点, 在对结果进行严格检验后选择一种相对最优的方法。考虑到可操作性, 本文借助地理信息系统软件中的空间分析模块, 采用距离平方反比法(IDW)、梯度距离权重反比法(GIDW)、克立格法(Kriging)、样条函数法(Spline)对辽宁省2005年的最低温度、最高温度、日降水量数据进行空间内插, 通过估值评估检验选择一种最优插值方法, 建立乡镇天气预报历史资料库。

2.1 距离权重反比法(IDW)

距离权重反比法(IDW: Inverse Distance Weighting)是一种常用而简便的空间插值方法, 它以插值点与样本点间的距离为权重进行加权平均, 离插值点越近的样本点赋予的权重越大。若权重用距离反比, 称为距离反比法; 权重用距离平方反比时

称为距离平方反比法。在实际应用中,通常选择后者,表达式如下:

$$Z = \frac{\sum_{i=1}^n (Z_i W_i)}{\sum_{i=1}^n W_i} \quad (1)$$

式中, n 为用于插值的气象观测站点数, Z 为估计的格点气象要素值, Z_i 为气象要素在第 i 个站点的实测值, W_i 为第 i 个站点的权重系数,通常以格点到观测点的大圆半径的平方反比作为权重系数。

2.2 梯度距离权重反比法(GIDW)

梯度距离权重反比法(GIDW: Gradient Inverse Distance Weighting)是对距离权重反比法的扩充,即在距离权重的基础上,进一步考虑了气象要素随海拔高度的变化,表达式如下:

$$Z = \frac{\sum_{i=1}^n (Z_i W_i)}{\sum_{i=1}^n W_i} + \left[e - \frac{\sum_{i=1}^n (e_i W_i)}{\sum_{i=1}^n W_i} \right] G \quad (2)$$

式中, Z 为待估格点的气象要素值, Z_i 为气象要素在第 i 个站点的实测值, e 和 e_i 分别为待估格点与气象观测站点的海拔高度, G 为气象要素随海拔高度变化的梯度。

2.3 克里格法(Kriging)

克里格法近年来在地质、气象等研究领域得到广泛应用。它的分析工具是半变异函数(semivariogram),对空间分布具有随机性与结构性的变量研究具有独特的优点^[4]。对任一空间变量点处的估计值 Z_x^* ,可以通过对该点影响范围内的 n 个有效观测值 $Z(X_i)$ 的线性组合得到,即

$$Z_x^* = \sum_{i=1}^n \lambda_i Z(X_i) \quad (3)$$

式中, λ_i 是赋予气象观测值 $Z(X_i)$ 的权重系数,表示各站点气象要素值 $Z(X_i)$ 对估计值 Z_x^* 的贡献。在克里格插值中,权重不仅建立在已知点和预测点位置间的距离的基础上,而且还要依据已知点的位置和已知点的值的整体空间分布和排列。应用权重的空间排列,空间自相关必须量化。因此,运用普通克里格插值,权重 λ_i 取决于已知点的拟合模型、距预测位置的距离和预测点周围和已知点间的空间关系。为达到线性无偏估计,使估计方差最小,权重系数由普通或简单克里格(ordinary/simple Kriging)方程组求得。同时权重系数取决于变量的空间结构性,变异估计就是拟合一个数学模型或空间模型,

像已知的结构分析。在已测点结构的建模中,首先得出经验半变异函数的曲线图。

$$h = 0.5 \bar{X} [(X_i - X_j)^2] \quad (4)$$

式中, h 为距离, \bar{X} 为距离平均值, X_i 和 X_j 为点对在 i 位置和 j 位置的值。

用于计算被距离 h 分隔的每一点对相对应的位置。公式用于计算一点对的差值的平方,可计算一点对中的一点的位置和其他所有已测点位置的相应关系。这一步骤延伸了每一个已测点。

而任何变量的空间结构由半变异函数 $\gamma(h)$ 表示:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (5)$$

式中, h 为距离矢量, $N(h)$ 为相距 h 的数据对的数目。

因权重系数取决于变量的空间结构性,而任何变量的空间结构可由半变异函数 $\gamma(h)$ 表示,所以在克里格插值方法中权重系数 λ_i 是由半变异函数 $\gamma(h)$ 决定的。

2.4 样条函数法(Spline)

样条插值法是用一种数学函数,对一些限定的点值,通过控制估计方差,利用一些特征节点,用多项式拟合的方法来产生平滑的插值曲线,用公式表示为:

$$z = \sum_{i=1}^n A_i d_i^2 \log d_i + a + bx + cy \quad (6)$$

其中, z 为待估点的栅格值, d_i 为插值点到第 i 个气象站点的距离, $a + bx + cy$ 为气温的局部趋势函数, x, y 为插值点的地理坐标, $\sum_{i=1}^n A_i d_i^2 \log d_i$ 为一个基础函数,通过它可以获得最小化表面的曲率, A_i, a, b 和 c 为方程系数, n 为用于插值的气象站点数。本文选用规则样条插值法。

规则样条插值将生成一个平滑、渐变的表面,对于规则样条插值来说,权重定义了曲率最小化表达式中表面的3阶导数的权重。权重越高,表面越平滑,这一参数的值必须等于或大于零。本文选用地理信息系统软件中样条函数插值时的默认值,权重值为1,插值运算中所采用的样本点数目为12个。 a, b 和 c 由已知点和插值点要素值以及插值点的地理坐标进行拟合得到。

3 插值站点数的确定

空间数据往往是根据用户要求所获取的采样点观测值,这些点的分布一般是不规则、不连续的,在用户感兴趣或模型复杂区域可能采样点多,反之则少。采样获得的数据一般是研究因素在某点的具体数值,是空间的矢量点数据。研究区域某空间因子采样的个数是有限的,不可能布满整个研究区域。当用户对未采样点的数值需要准确地了解时,希望能根据已知采样点的信息对附近未知点的属性进行预测和估计,内插的目的是根据已知点的属性合理推断和预测附近未知点的属性值,所有的内插都是建立在空间上距离越近的点相似性越强的假设基础上的。空间站点的选取首先要计算每个台站与其余各站距离,因为在内插时选取的空间点数是有一定限制的,且空间分布均匀的站点内插出的结果会更接近实际观测值。因此按大距离优先的原则筛选站点,这样做是在保证气象站点入选的前提下,去掉距离小的站点,以实现所选气象站点在辽宁地区均匀分布。站点的均匀分布选择利用两点间距离公式法,假定一个经度代表 97 km,一个纬度代表 110 km。用下式计算站点间的距离: $L_i = [(l_{oi} - l_{oi+1}) \times 97]^2 + [(l_{ai} - l_{ai+1}) \times 110]^2$, 其中 L_i 表示任意两个站点间的距离平方, l_{oi} 、 l_{oi+1} 和 l_{ai} 、 l_{ai+1} 分别代表任意两个不相重合的台站的经度和纬度。这样将每两个台站的距离进行计算,对每一个当前站点与其它所有站的距离从大到小排序,去掉排在最后的站点,被去掉的站点将不再参加下一次计算和排序。对 80 个辽宁省常规观测站和加密自动站资料按上述方法计算,每次去除 10 个站点,以 80 站、70 站、60 站、50 站、40 站、30 站、20 站这样的站点密度进

行研究。

影响空间内插精度的因素主要有空间分辨率和站点空间密度,选取不同的空间分辨率进行插值,旨在找出空间分辨率的不同对插值结果所造成的影响,进一步证明用常规观测站和自动站的资料通过内插技术建立乡镇天气预报历史资料库的可行性。文中选取的空间分辨率分别为 100 m, 300 m, 600 m, 1 km, 2 km, 6 km, 12 km。每个空间分辨率下对应一个插值结果,通过对各月份数据以不同站点密度进行误差比较,发现在任何空间分辨率下,误差值随站点密度的增加都呈减小趋势。图 1 给出了用距离平方反比法和样条函数法插值后计算的 2005 年降水的平均绝对误差随站点密度变化曲线。从图 1 可以看出,当站点密度达到 50 左右时,插值后计算的平均绝对误差值对于距离平方反比法而言,是随站点的增加稳定降低;而对于样条函数法,是随站点增加误差基本稳定。虽然站点密度越大,精度越高,但考虑到资料问题,本文最终选择 46 站插值^[5]。注意到误差随站点的增加,偶有起伏,可能是因为在计算误差时没有去除边界外的点。

文中边界外点的选取是选择临近省份距离相近的资料,以保证所要插值区域的每一点都有内插结果,避免所要插值区域的边缘点无资料。因为内插的结果都是基于空间距离越近的点相似性越高的假设基础上的,在选择边界点的时候选择临近省份距离本省距离最近站点的资料效果更好。

空间分辨率是指在进行空间插值时选取的网格大小。从图 1 可以看出,无论那种插值方法,空间分辨率对插值结果影响不大,因而将辽宁省 46 个站的资料插值到 952 个乡镇具有代表性。

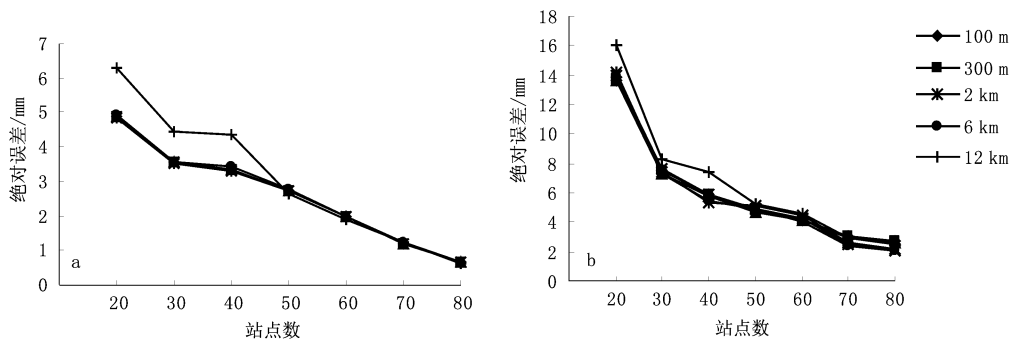


图 1 不同空间分辨率的降水量平均绝对误差随站点密度变化:(a)距离平方反比法,(b)样条函数法

4 插值站点的确定

对气象要素进行空间插值,一个必要的前提就是要使所选台站在所研究区域上均匀分布。辽宁省

有常规气象站 56 个,本文选取空间分布比较均匀的 46 个站作为插值建模站,剩余的 10 个站作为检验站(图 2)。

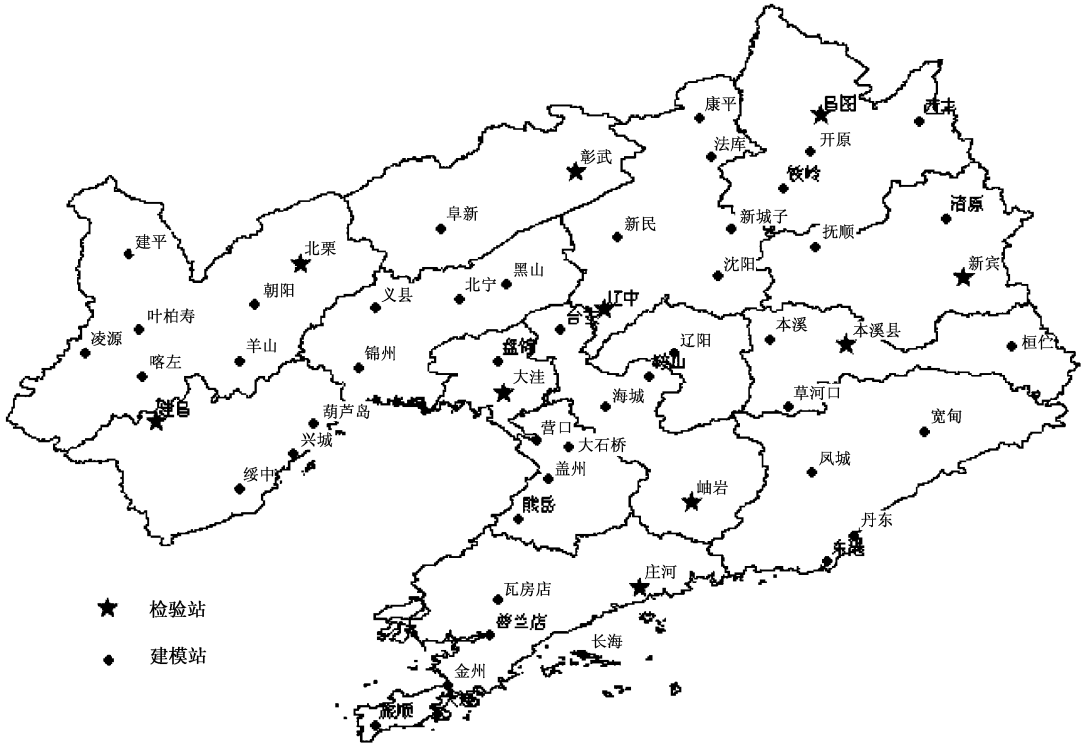


图2 建模站和检验站空间分布

5 检验评估

采用平均绝对误差 (MABE) 和均方根误差 (RMSE) 作为评估 3 种插值方法效果的标准。前者反映样本数据估值的总体误差或精度水平,后者反映利用样本数据的估值灵敏度和极值。

$$\text{平均绝对误差 } E_{\text{MABE}} = \sum_{i=1}^n |k_i - \hat{k}_i| / n$$

$$\text{均方差 } E_{\text{RMSE}} = \sqrt{\sum_{i=1}^n (k_i - \hat{k}_i)^2 / n}$$

将建模 46 个站分别用 IDW、Kriging、Spline 3 种方法进行空间插值,用 10 个检验站进行检验,气温选取每月的奇数日;降水选取每月有降水的日数。3 种插值方法气温和降水的绝对误差见图 3。

从图 3 出,温度的平均绝对误差小于 2°C ,对降水的模拟误差相对大一些,大多数都在 10 mm 以

下,但个别站的误差也较大,在 20 mm 左右,查看当天的雨情实况,属于局地强降水,这种天气对于只考虑距离权重而进行插值显然是不够的,由于天气系统的复杂性,导致产生局地强降水的因素很复杂,因而误差会较大。对于比较均匀的大范围降水,3 种方法模拟的效果都相对比较好。

均方根误差表示误差的离散程度(图 4),图 4 表明 IDW 的插值结果相对来说也是最好的。从温度误差的时间分布来看,夏季太阳高度角大,辐射空间分布差异小,辐射是影响气温空间分布的主要因子,气温受地形影响小,因而用插值得到的气温分布更接近实况,误差较小。冬季太阳高度角小,温度空间分布受地形影响大,不考虑地形影响仅考虑空间距离相近得到的气温误差必然也会增大。

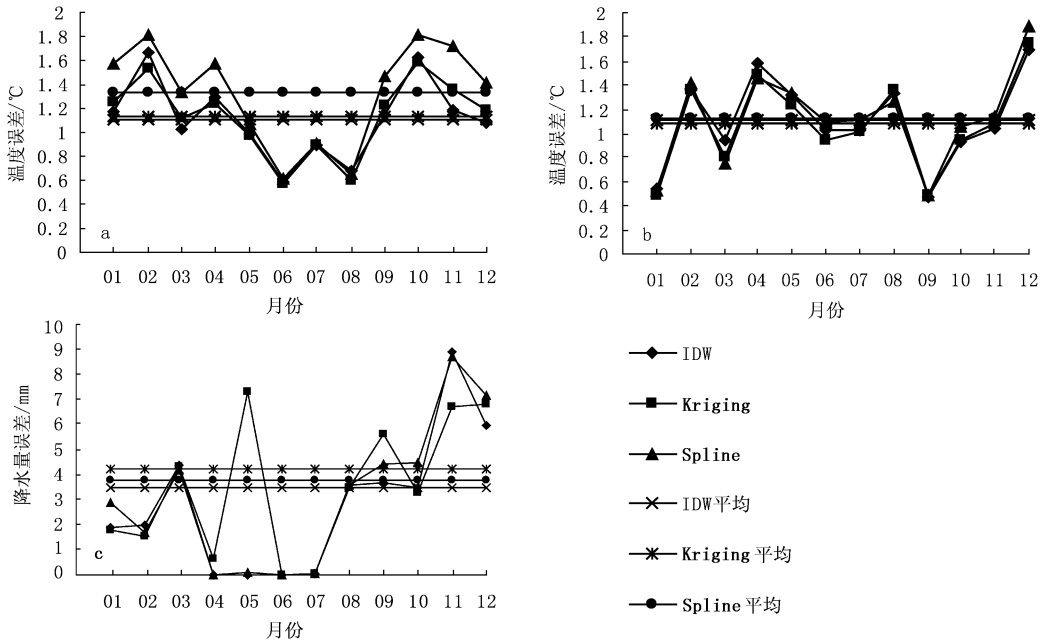


图 3 各种插值方法的气温和降水量绝对误差分布:(a)平均最低气温,(b)平均最高气温,(c)平均降水量

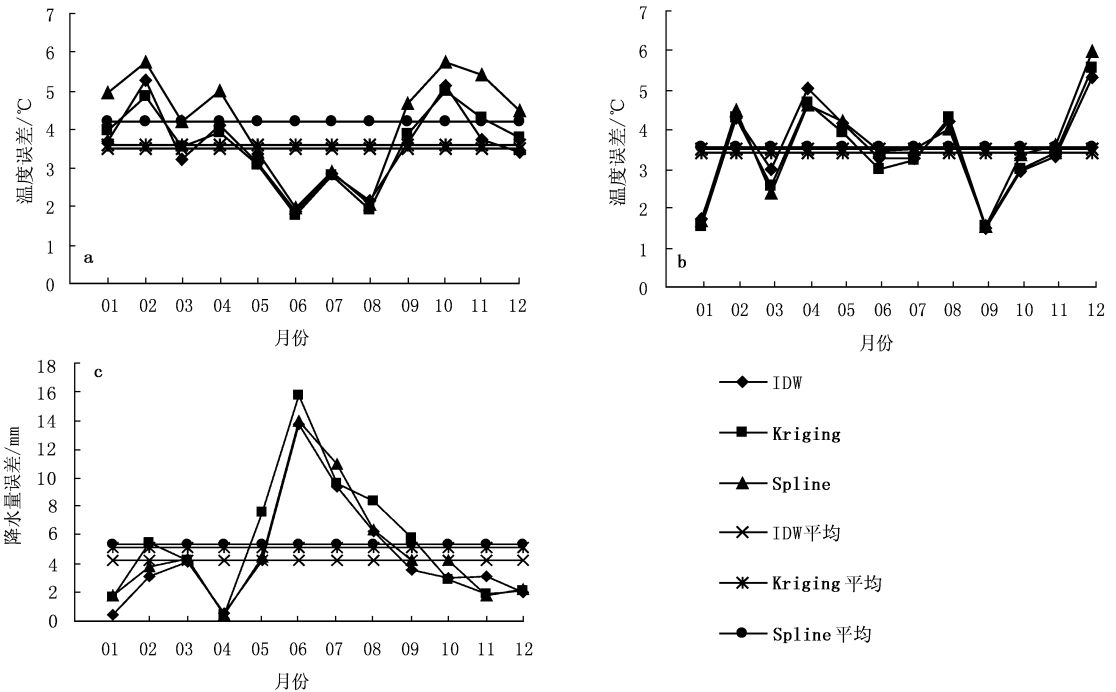


图 4 各种插值方法的气温和降水量均方根误差分布:(a)平均最低气温,(b)平均最高气温,(c)平均降水量

6 海拔高度对气温的订正

梯度距离权重反比法是距离权重反比法的扩展,其中海拔高度对气温的影响用下式表示:

$$\Delta T = \left[e - \frac{\sum_{i=1}^n (e_i W_i)}{\sum_{i=1}^n W_i} \right] G \quad (7)$$

对内插后的气温用式(7)进行海拔高度订正:

(1)用 56 个站每日的最低温度、最高温度资料和海拔高度进行回归,得到气温随海拔高度的递减率(G)。

(2)对气温递减率采用距离权重反比法进行空间内插,得到气温递减率的空间分布。

(3)权重系数(W_i)的计算方法同式(1)。

(4)将 W_i , G 和 DEM 高程值代入式(7), 得到海拔高度对气温的影响值。

近年来气候异常, 暖冬和局地高温打破了气温随纬度和距海远近的时空分布特征, 气象要素时空分布随局地自然条件的改变不断发生着变化, 导致日温度随海拔高度的变化并不简单地遵从通常的温度垂直递减率 $0.6\text{ }^{\circ}\text{C}/\text{hm}$, 而是有一定的波动范围。在进行带有高度订正的空间插值时, 应当使用动态的温度高度梯度。

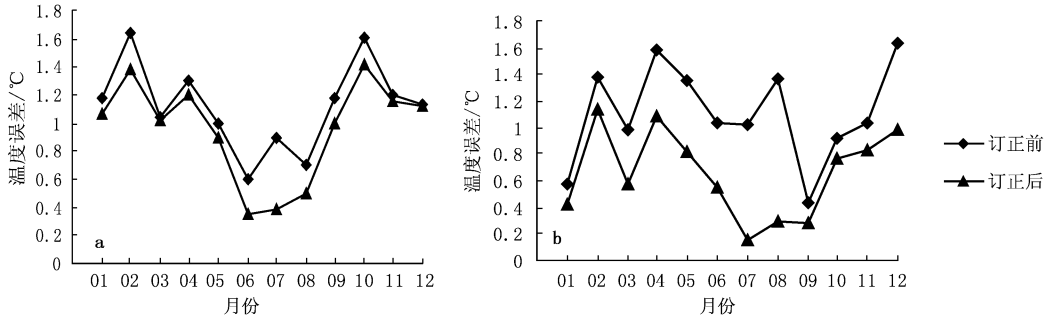


图5 海拔高度订正前后的日最低温度(a)和日最高温度(b)绝对误差

7 自动站资料对资料库的检验

辽宁省南北跨5个纬度, 东西跨6个经度, 由于范围比较小, 因而气温受经纬度影响较小, 地形和海拔高度是影响气温空间分布的主要因子。为了使选取的自动站能代表不同的地形, 在选取时根据其地理位置、海拔高度选取5个自动气象站作为检验站(沈阳市崇山西路(L1070)代表平原, 抚顺县后腰农场(L3507)代表内陆, 宽甸青山沟(L4533)和铁岭鸡冠山(L7604)代表山区, 葫芦岛南票兰甲(L9505)代表沿海)。

选取2005年8~12月自动站日最高温度、日最低温度、日降水量(8~10月)资料, 与同一天的插值结果进行比较, 其误差分布见图6。

从气温的误差分布看: 日最高温度误差一般都小于 $2\text{ }^{\circ}\text{C}$, 只有个别站的个别月份误差在 $4\text{ }^{\circ}\text{C}$ 左右, 日最低温度误差在 $4\text{ }^{\circ}\text{C}$ 以下。从误差时间分布看: 冬季误差大于其它季节, 这是由于冬季太阳高度角较低, 地形是影响气温空间分布的主要因子, 因而误

差较大。在太阳高度角较高的季节, 太阳辐射是影响气温空间分布的主要因子, 气温受地形的影响很小, 因而误差较小。从误差空间分布看: 山区大于沿海, 这是由于在沿海地区海水的调节作用, 使气温的日较差不会很大, 气温分布比较均匀, 均匀分布的数据插值效果会好些。从降水误差分布看: 由于降水是不连续分布要素, 每次降水过程成因复杂, 既有天气系统的影响也有地形、海拔高度等局地地形的影响, 因而误差分布并没有规律可循, 误差一般都小于 5 mm , 比较可信, 可为乡镇天气预报提供具有参考价值的基础数据。

利用海拔高度订正后计算的日最高温度、日最低温度的均方差和平均绝对误差见图5。由图5可见, 经过海拔高度订正后的最低温度、最高温度的绝对误差和均方差都有所减小, 尤其在太阳高度角较高的夏季, 海拔高度是影响气温空间分布差异的主要因子, 对气温进行海拔高度订正必然会使误差减小。而在冬季, 太阳高度角较低, 地形(坡向、坡度等)是影响气温空间分布的主要影响因子, 因而对冬季气温进行海拔高度订正后的误差减小不明显。

差较大。在太阳高度角较高的季节, 太阳辐射是影响气温空间分布的主要因子, 气温受地形的影响很小, 因而误差较小。从误差空间分布看: 山区大于沿海, 这是由于在沿海地区海水的调节作用, 使气温的日较差不会很大, 气温分布比较均匀, 均匀分布的数据插值效果会好些。从降水误差分布看: 由于降水是不连续分布要素, 每次降水过程成因复杂, 既有天气系统的影响也有地形、海拔高度等局地地形的影响, 因而误差分布并没有规律可循, 误差一般都小于 5 mm , 比较可信, 可为乡镇天气预报提供具有参考价值的基础数据。

8 结论

通过插值方法实现辽宁省气温和降水的空间分布推算是可行的。对温度而言梯度距离权重反比法(GIDW)估值精度较高, 插值结果分布趋势也较为接近实际站点的分布。对降水而言, 距离权重反比法(IDW)估值精度高于其他插值方法, 更适用于日降水量的空间插值。

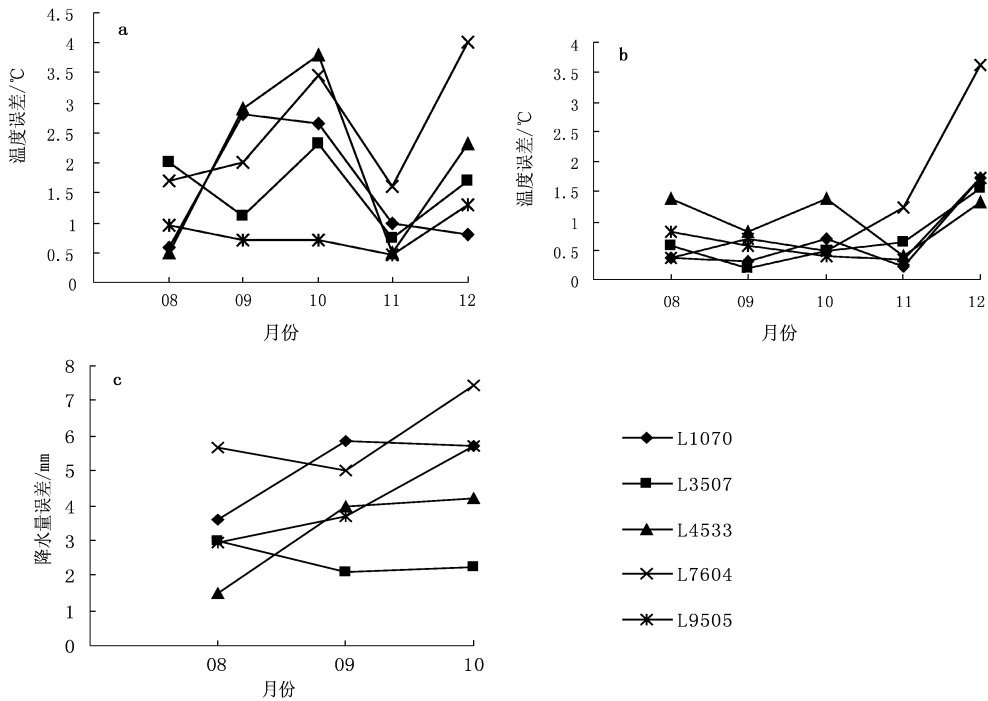


图 6 2005 年 8~12 月日最低温度(a)、日最高温度(b)及日降水量(c)插值结果与自动站实测值误差检验结果

参考文献

[1] 王跃山. 客观分析和四维同化——站在新世纪的回望(II) 客观分析的主要方法[J]. 气象科技, 2000, 28(1): 1-9.
 [2] 廖顺宝, 李泽辉. 气温数据栅格化中的几个具体问题[J]. 气象科技, 2004, 32(5): 352-356.

[3] 崔劲松, 张涛, 杨雨春. GPRS/GIS 技术在自动雨量监测网中的应用[J]. 气象科技, 2006, 34(2): 205-209.
 [4] 林忠辉, 莫兴国, 李宏轩, 等. 中国陆地区域气象要素的空间插值[J]. 地理学报, 2002, 57(1): 47-56.
 [5] 庄立伟, 王石. 东北地区逐日气象要素的空间插值方法应用研究[J]. 应用气象学报, 2003, 14(5): 605-615.

Method for Calculing Temperature and Precipitation Based on GIS Data in Liaoning Province

Lu Zhongyan¹ Yuan Zipeng¹ Cai Fu² Wu Manli¹ Liu Wenming¹

(1 Shenyang Central Meteorological Office, Shenyang 110016;

2 Institute of Shenyang Atmospheric Environment, CMA, Shenyang 110016)

Abstract: Aiming at the needs for high-resolution daily weather data, taking Liaoning Province as an example, the spatial interpolation methods of daily weather data from January to December are studied using the ordinary Kriging, Inverse Distance Weighting (IDW), the IDW with weighting of Gradient Inverse Distance Weighting (GIDW), and spline function methods. The statistic analysis for interpolated values and estimated values are made. The results indicate that for temperature, the precision of the estimated values is the highest, and the space-distributing trend of the interpolated result is the closest to the actual data by using GIDW method. For precipitation, the precision of the estimated values with the IDW method is higher than those with the other methods, which is more fit for the interpolation of daily precipitation.

Key words: temperature, precipitation, spatial interpolation, estimation test