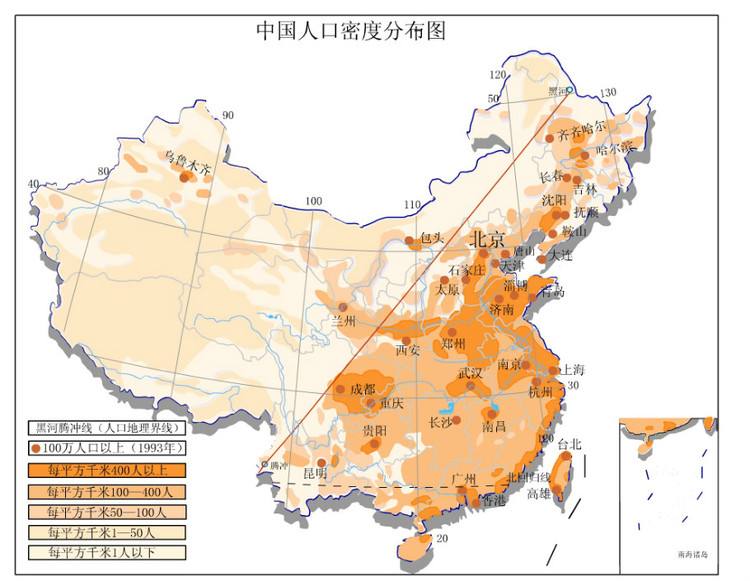
论文《基于GIS的气象要素空间插值方法研究》

该论文主要是对我国及周边地区2114个气象站点1961-1999年的平均温度、年降水量以及年积温数据，比较反距离权重法、样条函数法和普通克里格法的空间插值，并用交叉检验方法对插值精度进行评估。最终对这三种要素的插值结果验证发现，普通克里格法均具有最好的插值精度。

补充：

（1）胡焕庸线

胡焕庸线，即中国地理学家胡焕庸在1935年提出的划分我国人口密度的对比线。这条线从黑龙江省瑷珲（1956年改称爱珲，1983年改称黑河市）到云南省腾冲，大致为倾斜45度基本直线。线东南方36%国土居住着96%人口（根据2000年第五次全国人口普查资料，利用ArcGIS进行的精确计算表明，按胡焕庸线计算而得的东南半壁占全国国土面积43.8%、总人口94.1%。），以平原、水网、丘陵、喀斯特和丹霞地貌为主要地理结构，自古以农耕为经济基础；线西北方人口密度极低，是草原、沙漠和雪域高原的世界，自古游牧民族的天下。因而划出两个迥然不同自然和人文地域。



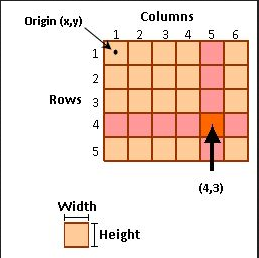
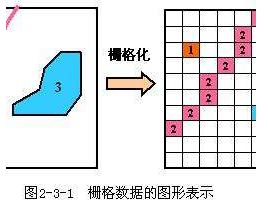
“胡焕庸线”在某种程度上也成为目前城镇化水平的分割线。这条线的东南各省区市，绝大多数城镇化水平高于全国平均水平；而这条线的西北各省区，绝大多数低于全国平均水平。

（2）区域化变量

区域化变量，地质统计学学术用语。当一个变量呈现空间分布时，就称之为区域化。这种变量往往反映某种空间特质，用区域化变量来描述的现象，称之为区域化现象。亦称区域化随机变量。具有两个显著的特征，即随机性和结构性。首先，区域化变量是一个随机函数，具有局部的、随机的、异常的性质；其次区域化变量具有一般的或平均的结构性质，即变量再点x与偏离空间距离为h的点x+h处的数值Z(x)与Z(x+h)具有某种程度的自相关，这种自相关依赖于两点间的距离h及变量特征。

（3）栅格结构

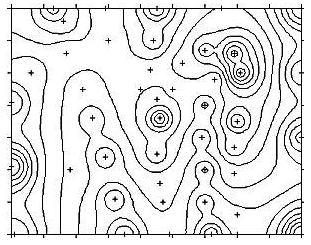
栅格结构是最简单直观的空间数据结构，又称网络结构或像元结构，是指将地球表面划分为大小均匀、紧密相邻的网络阵列，每个网络作为一个像元或像素，有行号、列号定义，并包含一个代码表示该像素的属性类型或量值，或仅仅包含指向其属性记录的指针。栅格结构是以规则的阵列来表示空间地物或现象分布的数据组织方式，阵列的每个数据表示地物或现象分布的非几何数据特征。

（4）空间数据库

空间数据库指的是地理信息系统在计算机物理存储介质上存储的与应用相关的地理空间数据的总和，一般是以一系列特定结构的文件的形式组织在存储介质之上的。由于传统的关系数据库在空间数据的表示、存储、管理、检索上存在许多缺陷，从而形成了空间数据库这一数据库研究领域。而传统数据库系统只针对简单对象，无法有效的支持复杂对象（如图形、图像）。

（5）反距离加权法一般在数据点比较集中，且均匀分布的时候使用，是一种经常用来定位异常值的方法。但是如果数据点太少或者太稀疏，或分布不均匀，那这种方法就不太好用。如果你看到一幅等高线地图上过于频繁地出现下面的这种『牛眼』形状，那么这幅图很可能就是用反距离权重法画的。由于该方法在格网区域内会产生围绕观测点的“牛眼”，给电法与磁法数据解释带来不便，因此实际应用较少。



（6）积温

一年内日平均气温≥10℃持续期间日平均气温的总和，即活动温度总和，简称积温。是研究温度与生物有机体发育速度之间关系的一种指标，从强度和作用时间两个方面表示温度对生物有机体生长发育的影响。一般以度·日(d·℃)为单位。

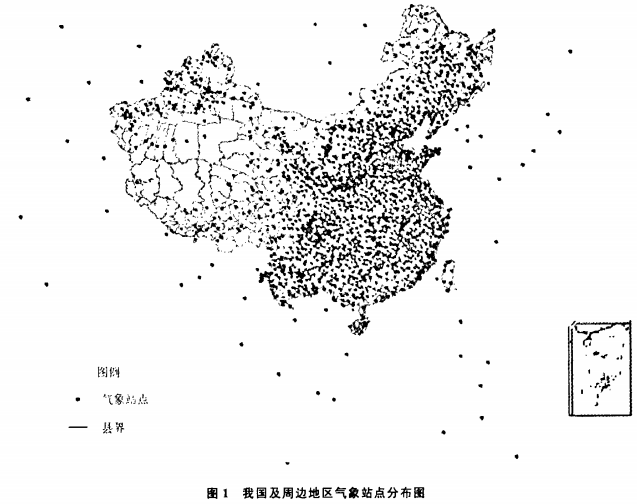
积温空间分布与地形、植被等下垫面特性有着密切关系，根据聚类的思想，将研究区域进行聚类分区，使分区内积温分布特征一致性趋强。然后利用统计方法寻找各分区年积温随海拔变化的规律，以获得较高精度的估算结果。之后在积温插值中引入地理位置和地形因子，很好地反映了积温空间分布的整体趋势，而且也表现出了积温分布的局部细微特征，并提高了积温插值的整体精度。

# 1. 空间插值

空间插值的实质是通过已知样点的数据来估算未知点的数据。常用于气象要素的空间插值方法按照是否考虑海拔高度等相关因素，可以分为以下两类。第一类是不考虑这些因素的方法：有泰森多边形法、反距离权重法、样条函数法、趋势面分析和普通克里格法等；第二类是考虑相关因素的插值方法：多元回归技术、协克里格插值法、Lapse比率和梯度距离反比法等。

# 2. 数据来源

气温和降水数据来自我国及周边2300个气象站点，去掉异常值和缺值，实际参与空间插值的气象站点为2114个，1961-1990年各气象站点30年的均值。



# 3. 研究方法

## 3.1 反距离权重法（IDW）

反距离权重法是对采样点进行线性的加权来决定输出的栅格值，加权与距离成反比，输入点离输出栅格越远，对输出栅格的影响越小。

其中，Z为待估计的气温栅格值，为第个气象站点的数据，为用于气象数据插值的站点数目，为插值点到第个气象站点的距离为距离的幂。

## 3.2 样条函数法（Spline）

样条函数法是使用一种数学函数，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。这种方法适用于逐渐变化的表面，如温度、高程、地下水位高度或污染程度等。样条函数法又分为张力样条函数法阿和规则样条函数法，由于规则样条函数法容易产生极值现象，所以选用的是张力样条函数法进行插值，并在插值时对参数进行了优化。

其中，Z为待估值的栅格值，插值点到第个气象站点的距离，为气温的局部趋势函数，为插值点的地理坐标，为一个基础函数，通过它可以获得最小化表面的曲率，为方程系数，为用于插值的气象站点的数目。

## 3.3 普通克里格法（OK）

普通克里格法是利用区域化变量的原始数据和变异函数的结构特点，对未采样点的区域化变量的取值进行线性无偏最优估计的一种方法。

其中，Z为待估值的气象数据栅格值，为赋予站点气象数据的一组权重系数，为用于气象数据插值的站点数目，为站点的气象数据值。

为满足无篇性和最优性两个条件，通过建立如下克里格方程组来确定权重系数：

其中，为气象站点之间的协方差函数，为气象站点与插值点之间的协方差函数，为朗格朗日乘数。

# 4. 插值气象站点年均温度为负值时的处理方法

对于一些气象站点，年平均温度有可能出现负值。而插值的方法是将每个地区所有临近站点的值都参与进插值过程。所以对年平均气温度为负值的气象站点，有必要通过一定的数学处理，使之统一成正值，然后进行插值，确保插值结果的精度与可信度。分析我国及周边地区气候站点原始数据，发现年平均温度最低为-7.4℃，所以对所有参与气象插值的气象站点年平均温度加上8以后，再进行插值。插值出的结果再减去8。这样处理的目的：消除负值站点对插值整体精度的影响。【问题：负值有何影响？】

其中，为参与插值的气象站点数目，为处理后的气象站点年均温度值，为气象站点原始平均温度值，为最终生成的栅格数据库每一点的年均温度值，为经过插值后栅格数据库每一点的年均温度值。

# 5. 不同气象站点数目的选择

对反距离加权法、样条函数法和普通克里格法这3种插值方法，各选择3、5、10、15、20、25、30个站点数据，分别对年均温、年积温和年降水量进行空间插值比较分析，从中选出最适合的站点数目。【问题：这选的站点数目少了吧？】

# 6. 检验标准

采用交叉验证法来验证插值的效果。即首先假设某一站点的气象要素值未知，都用周围站点的值来估算，然后计算所有站点实际观测值与估算值的误差，以此来评判插值方法的优劣。

这里采用平均误差和均方差误差来作为评估几种插值方法的插值效果的标准。选择平均误差和均方差误差相对最小的插值方法来建立全国气象数据的栅格空间数据库。平均误差反映总体估计误差的大小，均方差误差反映利用样点数据的估值灵敏度和极值效应。均方差误差表达式为：

其中，为均方差误差，为第个站点的实际观测值，为估计值，为用于检测的站点数目。

# 7. 结果分析

将我国及周边地区气象站点年均温度统一为正值后再进行插值，与使用原始数据插值的结果进行对比发现：在不同参与插值气象站点数目下，均可以在一定程度上提高插值结果的精度。【问题：本文中没有再进行提高精度的操作，进一步提高精度要使用哪些方法？】



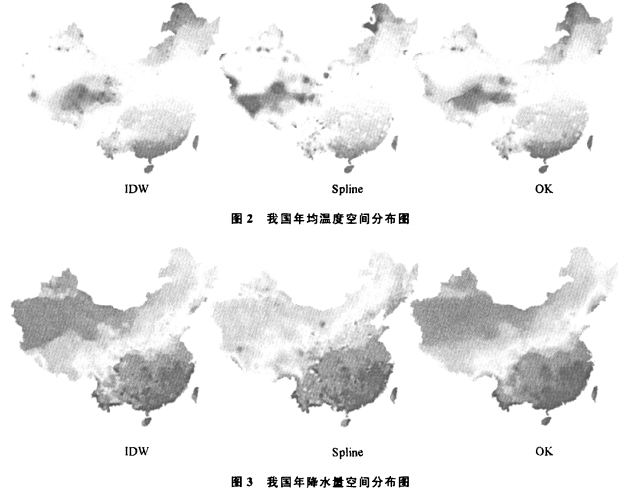
结果：不同插值方法对于适宜气象站点数目的要求是不同的，插值的站点数目不同，插值结果精度不同，空间插值时，适宜数量的站点插值结果最优，并非是越多越好。

# 8. 误差分析

看表2可知，对于年均温来说，ME的排序为反距离权重法 > 样条函数法 > 普通克里格法；RMSE的排序为反距离权重法 > 样条函数法 > 普通克里格法。对于年降水来说，ME的排序为反距离权重法 > 样条函数法 > 普通克里格法；RMSE的排序为反距离权重法 > 样条函数法 > 普通克里格法。对于年积温来说，ME的排序为样条函数法 > 反距离权重法 >普通克里格法；RMSE的排序为样条函数法 > 反距离权重法 >普通克里格法。可见普通克里格法均具有最小的ME和RSME。对年均温度和年降水量来说，样条函数法的插值精度优于反距离权重法；而对于年积温，反距离权重法的插值精度优于样条函数法。

# 9. 空间插值结果

将插值结果转换为1km×1km的我国平均温度、年降水量和年积温的栅格空间数据库：【问题：如何转换？空间分布图有点模糊，所以有些结论看的并不是很清晰】





看图2可知，三种插值方法均能较好地反映出年均温度的纬度地带性变化。在西部山区和河谷地区，普通克里格法和样条函数法较反距离权重法更好地反映出温度随地形变化的趋势。普通克里格法和样条函数法插值出来的图较平滑。在反映温度的变化趋势上，样条函数法的能力接近于普通克里格法，而反距离权重法生成的图上存在很多比较小的“牛眼”。

从图3看到，我国降水空间分布的趋势大致呈从东南向西北逐渐减少的趋势，东-西、南-北的空间差异非常显著。样条函数法和反距离权重法生成的图非常相似，相比较来说，普通克里格法生成的年降水量分布图更加接近于真实。

从图4看到，我国年积温和降水量的分布趋势非常一致，说明水热条件的分布大致是相同的。反距离权重法生成的图上具有许多非常明显的极值分布，样条函数法生成的图与反距离权重法比较相似。在西北地区，三种方法的插值结果都较差，没有反映出年积温随地形起伏的变化。

# 10. 结论

（1）水热条件的东南和西北分界非常明显，大致与胡焕庸线一致，即东南水热条件较为充分，而西北水热条件较差。

（2）从本质上来说，空间内插方法是研究区域化变量空间分布的基本方法，各种方法都具有其特定假设、适用范围、算法和优缺点，目前常用的办法是对多种方法进行比较研究，然后根据插值精度选择其中精度最高的一种。对于不同的空间变量，在不同的地域和不同的时空尺度内所谓的最优内插法是相对的。

（3）影响插值结果的因素很多，参数的选择、气象站点的分布和数目等都会影响到最后的插值结果。通过对原始数据进行一定的处理，可以在一定程度上提高最终插值结果的精度。

# 11. 最后

（1）数据图以及结果图较为模糊，导致无法清晰地分析结果。

（2）西部气象站点数量少，无法得到西部地区精确的水热条件情况。

（3）气象站点为何选择2300个我国及周边，既然最后涉及到胡焕庸线的比较，不应该只选取我国各地区的气象数据吗？

（4）除了对原始数据全部处理为正数，并没有做其他改进处理来提高精度。还有哪些方法可以采取呢？

（5）站点数据的选择，最少是3最多是30，这个数据量是不是比较小吧？

（6）如何将插值结果转换为栅格空间数据库？

（7）通过反距离权重法、样条函数法和普通克里格法对不同站点数据插值的对比，得出相对最优的插值方法，保证了一定程度上的结果精确性。

（8）论文中各步骤逻辑清楚，容易看懂作者的整体思路及实验流程。

空间数据库全面版

https://wapwenku.baidu.com/view/2c42708f33687e21af45a9d3?rand=0.4967514800373465