

文章编号: 1004-4353(2016)01-0081-04

# 离散余弦变换插值处理方案在 海洋数据分析中的应用

陈璇<sup>1</sup>, 郑崇伟<sup>2,3\*</sup>, 刘文鹏<sup>2</sup>, 孙威<sup>3</sup>, 田妍妍<sup>4</sup>, 高凡<sup>5</sup>

( 1. 解放军 75822 部队, 广东 广州 510510; 2. 海军大连舰艇学院 航海系, 辽宁 大连 116018;  
3. 解放军理工大学, 江苏 南京 211101; 4. 91379 部队气象台, 山东 青岛 266002;  
5. 92853 部队, 辽宁 兴城 125106 )

**摘要:** 传统的插值法在处理海洋数据中端点附近的数据时,其处理方法较为复杂,不同的插值方案及端点处理方法对数据处理结果有较大影响。据此,本文对比分析了海洋资料处理中常见的空间插值方案和应用较少的频域插值方案(离散余弦变换,DCT),结果表明:与常见的空间插值方案相比,频域插值方案消耗处理时间较少,程序设计和编辑以及在端点位置的处理上较为便捷,同时受数据量的影响较小,实用性较高。

**关键词:** 空间插值; 频域插值; 离散余弦变换

**中图分类号:** P731

**文献标识码:** A

## DCT interpolation method for marine data analysis

CHEN Xuan<sup>1</sup>, ZHENG Chongwei<sup>2,3\*</sup>, LIU Wenpeng<sup>2</sup>, SUN Wei<sup>3</sup>, TIAN Yanyan<sup>4</sup>, GAO Fan<sup>5</sup>

( 1. No. 75822 of PLA, Guangzhou 510510, China; 2. Department of Nautical Navigation,  
Dalian Naval Academy, Dalian 116018, China; 3. People's Liberation Army University of  
Science and Technology, Nanjing 211101, China; 4. No. 91379 of PLA, Qingdao 266002,  
China; 5. No. 92853 of PLA, Xingcheng 125106, China )

**Abstract:** For near the endpoints of the ocean data, the traditional interpolation method is more complex, and different interpolation schemes and processing methods for near endpoints have a great impact on the processing results. Thus, in this paper, we compared and analyzed the common spatial interpolation method and the frequency domain interpolation (FDI) method that is relatively little used in ocean data. The results show that, compared to the common spatial interpolation method, the FDI method is optimized in time consumption, convenient for program designing, program editing and processing the data near the endpoints; in addition, the FDI method is less impacted by the amount of data, and has advantages of high practicability.

**Keywords:** spatial interpolation; frequency domain interpolation; discrete cosine transform

## 0 引言

海洋资料是研究海洋各种现象、探寻海水运动等规律的重要支撑,同时也是海洋环境特征分析、海洋能资源评估的基础,如:南海岛礁建设评

估<sup>[1-2]</sup>、海上丝绸之路建设<sup>[3]</sup>等。相对于气象观测资料而言,海洋观测资料受环境因素的制约影响较为显著,如海洋观测缺乏固定的测站,无法实时作业等,使得难以及时获取高质量的海洋观测资料;因此,海洋数据插值成为处理海洋资料的重要

手段. 插值法主要有利用导数(斜率)等空间分布特征构造插值函数的方法(空间方案)和利用基函数的方法(变换方案或频域方案)两大类. 比较典型的方案有: 样条插值、Akima 插值、拉格朗日插值等<sup>[4]</sup>. 这类插值法的共同点是: 依据数据空间分布的特点, 构造插值函数. 但这类方案对于二维数据或高维数据的插值处理则较为复杂, 其复杂性在扩展维度时尤为明显. 此外, 在处理端点附近的数据时, 该类方案也存在一定的局限性<sup>[5-6]</sup>.

相较于传统的空间方案, 基于基函数的频域方案则不存在这一问题, 目前该方案在气象、海洋等学科中多用于周期、波谱分析, 而较少用于空间插值处理<sup>[7-9]</sup>. 为解决传统插值方案在端点附近数据的处理上以及维度扩张时所存在的问题, 本文将利用离散余弦变换(DCT)方案对一维(二维)数据进行插值处理, 并将插值结果与传统空间插值方案的处理结果进行对比分析.

## 1 方法介绍

在常见的插值方案中, 本文仅对拉格朗日插值方案和 Akima 插值方案进行介绍.

### 1.1 拉格朗日插值法

拉格朗日插值法<sup>[4]</sup>,<sup>64</sup> 是对于  $n$  个序列点, 其相应坐标为  $(x_i, y_i)$ ,  $1 \leq i \leq n$ , 其中  $x_i$  表示目标位置和坐标位置, 如时间、空间坐标,  $y_i$  表示相对应要素的值. 对应  $n-1$  次拉格朗日插值函数  $L_n$ , 其表达式为  $L_n = \sum_{k=1}^n \left( \prod_{j=1, j \neq k}^n \frac{x - x_j}{x_k - x_j} \right) \cdot y_k$ , 该式对于样本数目限制较小.

### 1.2 Akima 插值算法

Akima 插值算法<sup>[4]</sup>,<sup>68-70</sup> 是利用相邻的 6 个点之间的空间关系对数据进行插值处理, 具体操作如下:

相邻的 6 个实测点的序号用  $i$  ( $i=1, 2, \dots, 6$ ) 表示, 相应坐标为  $(x_i, y_i)$ , 其中插值点  $(x, y)$  位于第 3、4 个点之间,  $x$  及  $x_i$  表示目标位置和坐标位置, 如时间、空间坐标,  $y$  及  $y_i$  表示相对应要素的值, 即  $x_3 < x < x_4$ .  $y$  可用式(1)计算:

$$y = p_0 + p_1(x - x_3) + p_2(x - x_3)^2 + p_3(x - x_3)^3, \quad (1)$$

式(1)中的系数计算采用式(2):

$$\begin{cases} p_0 = y_3; \\ p_1 = t_3; \\ p_2 = \frac{\frac{3(y_4 - y_3)}{(x_4 - x_3)} - 2t_3 - t_4}{x_4 - x_3}; \\ p_3 = \frac{t_3 + t_4 - \frac{2(y_4 - y_3)}{x_4 - x_3}}{(x_4 - x_3)^2}. \end{cases} \quad (2)$$

在式(2)中,  $t_3, t_4$  的计算方式如式(3):

$$t_i = \frac{|m_{i+1} - m_i| m_{i-1} + |m_{i-1} - m_{i-2}| m_i}{|m_{i+1} - m_i| + |m_{i-1} - m_{i-2}|}. \quad (3)$$

式(3)中,  $i=3, 4$ ,  $m_i$  为斜率, 其表达式为

$$m_i = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}. \quad (4)$$

利用式(4)对于首末两端处做插值时, 采用外推法构造出相应点, 并进行相关插值. 在对二维数据插值处理时, 一般先处理数据的一个维度, 再基于处理后的数据对另一个维度进行重复处理. 从实现手段上来看, 其计算总量与数据总维数呈指数关系.

### 1.3 离散余弦变换(DCT)

DCT 是余弦变换的离散化形式, 其逆变换就是原数组或函数本身, 该类方案基于频域基函数变换的原理, 在数字图像、信号处理中有广泛的应用<sup>[10-12]</sup>.

1) 一维情形. 一维数据  $\{f_k \mid k=0, 1, 2, \dots, N-1\}$  ( $N$  为数据总长度) 经 DCT 变换后的数据为  $\{F_m \mid m=0, 1, 2, \dots, N-1\}$ , 其中

$$F_m = \alpha_m \sum_{k=0}^{N-1} f_k \cos \left[ \frac{(2k+1)m\pi}{2N} \right]. \quad (5)$$

式(5)中  $\alpha_m = \begin{cases} \sqrt{\frac{1}{N}}, & m=0; \\ \sqrt{\frac{2}{N}}, & m \neq 0. \end{cases}$  与式(5)相对应的

逆变换为:

$$f_n = \sum_{k=0}^{N-1} \alpha_k F_k \cos \left[ \frac{(2n+1)k\pi}{2N} \right], \quad (6)$$

其中  $0 \leq n \leq N-1$ . 理论上, 式(6)可以对数列中的任意位置进行插值, 同时相较于 Akima 插值方案, DCT 变换对数据样本容量的要求也较小.

2) 二维情形. 二维情形有重复利用一维处理和利用二维变换 2 种做法<sup>[13-14]</sup>, 本文在此介绍二

维变换做法. 设一、二维的数据矩阵记为  $[f]_{m \times n}$ , 其中第  $i$  行、第  $j$  列元素为  $f_{ij}$ ,  $1 \leq i \leq m, 1 \leq j \leq n$ . 该数据矩阵对应于 DCT 变换后的矩阵记为  $[F]_{m \times n}$ , 其中第  $u$  行、第  $v$  列元素为  $F_{uv}$ :

$$F_{uv} = \frac{2}{\sqrt{mn}} c_u c_v \cdot \sum_{i=1}^m \sum_{j=1}^n f_{ij} \cos \left[ \frac{(2i-1)(u-1)\pi}{2m} \right] \cdot \cos \left[ \frac{(2j-1)(v-1)\pi}{2n} \right]. \tag{7}$$

式(7)中  $1 \leq u \leq m, 1 \leq v \leq n, c_u = \begin{cases} \frac{1}{\sqrt{2}}, & u=1; \\ 1, & u \neq 1, \end{cases}$   $c_v$  与  $c_u$  相似. 与式(7)相对应的 DCT 逆变换为:

$$f_{ij} = \frac{2}{\sqrt{mn}} \sum_{u=1}^m \sum_{v=1}^n c_u c_v F_{uv} \cdot \cos \left[ \frac{(2i-1)(u-1)\pi}{2m} \right] \cdot \cos \left[ \frac{(2j-1)(v-1)\pi}{2n} \right]. \tag{8}$$

如前所述, 理论上式(8)也可以进行任意位置的插值. 从实现方案上看, 此类计算只涉及卷积运算, 其计算复杂度和计算量较空间方案要小.

## 2 插值效果对比

本文利用 Matlab 数学软件实现拉格朗日方案、Akima 方案和 DCT 方案, 并基于这 3 种方案对相同数据进行处理和对比分析.

### 2.1 一维效果对比

所用样本数据见表 1. 各个方案的插值点均选在原始数据中相邻数据点之间的中点位置, 如表 1 中  $x=1.5$  的位置. 3 种插值方案的处理效果如图 1 所示, 插值结果的相关性分析见表 2. 从图 1 可以看出: DCT 方案和 Akima 方案对数据均有很好的描述, 两者的结果基本吻合; 拉格朗日方案则存在较大波动, 对数据趋势突变及端点位置附近的描述效果较差.

表 1 原始数据											
$x$	0	1	2	3	4	5	6	7	8	9	10
$y$	10	10	10	10	10	10	10.5	15	50	60	85

注: 数据来源于文献[4]<sup>70</sup>; 表中  $x$  表示相应序号,  $y$  为对应的数组的取值.

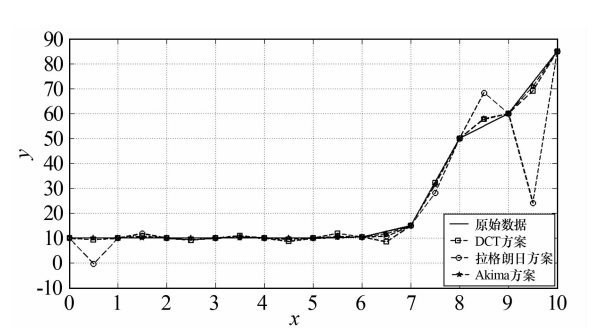


图 1 3 种插值方式的对比图

表 2 插值后数据相似度

插值方法	拉格朗日插值	Akima 插值	DCT 插值
拉格朗日插值	1.000 0	0.899 4	0.904 5
Akima 插值	0.899 4	1.000 0	0.999 1
DCT 插值	0.904 5	0.999 1	1.000 0

### 2.2 二维效果对比

所用的样本数据来源于 CORA 再分析资料<sup>[15]</sup>. 该资料是中国近海及邻近海域(海区范围为  $99^{\circ}\text{E} \sim 150^{\circ}\text{E}, 10^{\circ}\text{S} \sim 52^{\circ}\text{N}$ )海面高度场及三维温、盐、流场的再分析月平均数据, 其时间长度为 23 年, 数据文件命名方式为: CORA\_trial\_YYYYMM.nc, 其中 YYYY 为年份, MM 为月份. 本文所选用的数据资料为 CORA\_trial\_200807.nc, 处理的要素为海表温度场.

为避免岸界影响, 本文选取无岸界的矩形区域(海区范围为  $130^{\circ}\text{E} \sim 144^{\circ}\text{E}, 10^{\circ}\text{N} \sim 30^{\circ}\text{N}$ )内的数据作为原始数据, 并将该区域内的数据偶数列取出, 保留其剩余部分作为预处理数据, 并利用 Akima 方案和 DCT 方案还原原数据偶数列的温度值, 然后对比分析插值数据和原始数据的差异, 对比方案采用比较计算处理后的标准差、误差极值和处理时间. 数据插值处理效果见图 2 和表 3, 图 2 中等值线的间隔为  $0.25^{\circ}\text{C}$ .

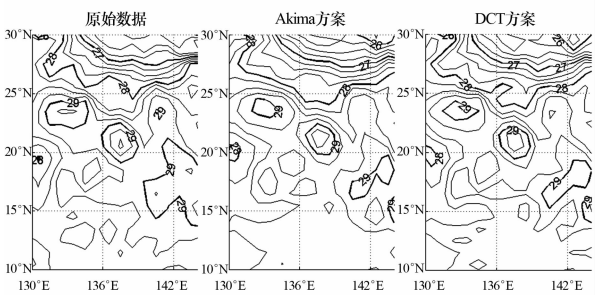


图 2 二维插值对比图

通过比较 DCT 方案和 Akima 方案的插值处理结果发现,2 种插值方案还原的信息与原始场具有较高的相似度,但相较于 Akima 方案,DCT 方案的标准差和误差极值的绝对值都相对较小.此外,DCT 方案所消耗的时间也只有 Akima 方案的一半左右.

表 3 2 种方案的插值效果对比

对比参数	DCT 方案	Akima 插值
标准差/℃	0.026 8	0.044 6
最大值/℃	0.106 5	0.316 2
最小值/℃	-0.561 4	-0.650 9
时间消耗/s	0.077 9	0.136 3

3 结论

本文对空间插值方案和频域插值方案进行对比分析表明:1)相较于空间方案的局部性,DCT 方案的全局性可以优化程序设计、降低时间消耗;2)DCT 处理后其相应插值点的值与原数据的差异较 Akima 方案要小,相应误差分布也趋于零;3)理论上 DCT 可不受数据量的限制,而 Akima 方案要求插值样本的数据量至少为 6 个,对于二维插值,其数据量则要大于 36 个;此外,相较于 Akima 方案,DCT 不存在端点处理的复杂问题.由以上结果表明,DCT 方案在大规模的数据插值处理上具有广泛的应用前景.随着近年来对于大气分形及分数维度研究的不断深入,基于以正余弦函数为代表的魏尔斯特拉斯函数的研究<sup>[16]</sup>也逐步展开,这也使得 DCT 方案在描述大气分形特征等相关领域具有较好的应用价值.

参考文献:

[1] 郑崇伟,李崇银.中国南海岛礁建设:重点岛礁的风

候、波候特征分析[J].中国海洋大学学报(自然科学版),2015,45(9):1-6.

[2] 郑崇伟,李崇银.中国南海岛礁建设:风力发电、海浪发电[J].中国海洋大学学报(自然科学版),2015,45(9):7-14.

[3] 郑崇伟,潘静,孙威,等.经略 21 世纪海上丝路之海洋环境特征系列研究[J].海洋开发与管理,2015,32(7):4-9.

[4] 陈上及,马继瑞.海洋数据处理分析方法及其应用[M].北京:航空工业出版社,1991:64-70.

[5] 王勇标.克里金算法的改进[D].荆州:长江大学,2012:4-15.

[6] 邵才瑞.井筒数据的 Akima 插值重采样方法[J].测井技术,2004,28(2):112-114.

[7] 薛宇峰.近 50 年来湛江、韶关气候变化的小波分析[J].广东气象,2006(1):29-32.

[8] 万明波,程智,王文.青藏铁路沿线气温和降水的小波分析[J].干旱气象,2006,24(4):35-39.

[9] 王盘兴,吴兴毓,华文漪,等.热带气候风场的傅立叶分析方案及试验[J].大气科学学报,2008,31(3):300-307.

[10] 孙燮华.数字图像处理:原理与算法[M].北京:机械工业出版社,2010:80-81.

[11] 万永革.数字信号处理的 MATLAB 实现[M].北京:科学出版社,2012:112-124.

[12] Rafael C, Gonzalez, Richard E Woods. Digital Image Processing[M]. New Jersey: Prentice Hall, 2007:150-156.

[13] 吴红文,李久贤,夏良正.一种新二维离散余弦变换快速算法[J].东南大学学报(自然科学版),1996(2):7-15.

[14] 黄杏英,黄华灿.一种二维离散余弦变换系数快速算法[J].华侨大学学报(自然科学版),2007,28(2):131-134.

[15] 韩桂军,李威,张学峰,等.中国近海及邻近海域海洋再分析产品[M/OL]. [2015-10-12]. <http://www.cora.net.cn/>.

[16] 刘式达,刘式适.物理学中的分形[M].北京:高等教育出版社,2014:42-66.