

Ocean Hydrological Data Visualization Based on Radar Chart Series

Zhaoyang Feng, Caixia Li

Satellite Marine Tracking and Control Department of China, Jiangyin Jiangsu
Email: zyfeng@ustc.edu

Received: Aug. 22nd, 2018; accepted: Aug. 31st, 2018; published: Sep. 7th, 2018

Abstract

Traditional data visualization can only visualize single element of ocean hydrological data. Multi-dimensional data visualization can visualize multi-element in planar coordinate. During to its shortage in displaying multidimensional multivariate data, a new method to add dimension was proposed, which could extend Radar Chart into 3D space (Radar Chart Series). So that multidimensional multivariate data could be displayed as polygonal plane cluster for marine elements showing and analyzing. Results reveal that Radar Chart Series can effectively analyze characters and provide a powerful tool to reveal the rules of abstract data and the complex oceanic phenomena.

Keywords

Ocean Hydrological Data, Multidimensional Multivariate, Radar Chart, Data Visualization

基于雷达图序列的海洋多维数据可视化方法

冯朝阳, 李彩霞

中国卫星海上测控部, 江苏 江阴
Email: zyfeng@ustc.edu

收稿日期: 2018年8月22日; 录用日期: 2018年8月31日; 发布日期: 2018年9月7日

摘要

传统的海洋水文数据可视化方法通常只对单一要素进行显示, 而多维数据可视化方法通常只在二维平面坐标系上显示多要素数据。本文针对海洋高维多元数据提出了一种通过添加时空维度坐标轴, 将雷达图扩展到三维空间的方法, 绘制由二维雷达图和一个时空维构成的雷达图序列Radar Chart Series, 融合

了剖面图和雷达图各自的信息, 既在每个雷达图表示多个水文要素的数据信息, 又在时空维度上反映各要素沿某一时空维度的变化趋势, 从而在全局上展现整个数据集分布特征以及各个研究对象(要素)之间可能存在的关系。实验结果表明, 雷达图序列方法在很大程度上拓宽了一个图所能承载的数据信息, 对数据中的隐藏信息进行了更深层次的挖掘。

关键词

海洋水文数据, 高维多元数据, 雷达图, 数据可视化

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

海洋的信息化应用程度是衡量一个国家海洋实力的重要标志。利用可视化技术来仿真海洋水文要素在空间和时间的分布与运动, 一直是国内外学者和相关机构关注和着力研究的重大课题[1]。

目前国内外几种典型图件的绘制都是侧重表现单要素的空间变化。比如剖面图, 描述某一测站海洋水文要素(温度、盐度、密度或声速)随深度的变化, 通常一幅图中最多展现两个要素; 断面图是垂向二维分布图, 描述特定断面深度剖面上海洋水文要素的分布情况; 平面图是水平二维分布图, 描述典型层面海洋水文要素的分布情况, 与断面图一起可以获得关于海洋水文要素的立体分布情况[2]。这些基于等值线图、散点图、折线图等传统可视化方法绘制的图件通常只能表现某一单一海洋水文要素的变化规律。而专业分析关注的特征不仅仅依赖一个要素本身, 通常需要将影响海洋环境变化的多种复杂海洋水文要素同时表达出来, 在空间坐标系上直观地比较一个数据维与其它数据维之间的联系, 分析数据集的特性如趋势、聚类 and 关联等等。因此, 传统的可视化方法在表现具有多维数据特征的海洋水文要素时面临着许多挑战。

近年来研究人员已经提出很多有效的多维数据可视化方法, 如散点图矩阵[3]、平行坐标[4] (Parallel Coordinates)、星形坐标法(雷达图)等, 这些方法是基于二维坐标系的可视化方法, 在二维可视空间中展现多维抽象信息的多属性数据特征[5], 在统计、金融、文本等信息可视化领域中得到广泛应用。但海洋水文数据与这些非时空数据不同, 属于高维多元数据, 即高维(Multidimensional)是海洋水文数据具有多个独立属性, 如时空维度; 多元(Multivariate)是海洋水文数据具有多个相关属性, 如温、盐、密等海洋水文要素与经度、纬度、深度等空间维度和时间维度相关。这就要求可视化技术要满足以下四个条件: 1) 能反映数据的时空特性, 即空间维度(经度、纬度、深度)和时间维度; 2) 能同时进行定性或定量的多个维度属性数据的比较, 反映数据的相关度; 3) 能提供整体概况和局部信息; 4) 能快速检索到元数据。

因此, 传统的多维数据可视化方法需要针对时空多维数据进行改进, 从而准确快速地发现数据集中时空维度隐藏的特征信息、关系信息、模式信息及趋势信息等。孙扬等对多变元数据可视化方法[6]进行了研究, 姚中华探索了一种时序多维数据的可视化方法[7], 文献[8]基于几何的平行坐标系法和圆形平行坐标系法实现了地质数据属性信息的分布特征可视化。本文提出一种扩展时空维度的雷达图序列可视化方法(简称为 Radar Chart Series), 在二维雷达图的基础上添加一个时空维度, 将坐标系扩展至三维空间, 从而支持在时空维度观察海洋水文数据多要素随某一时空维度的变化, 在有限的空间里表达更加丰富的信息。

2. 海洋时空多维数据可视化方法

2.1. 时空多维数据模型

将海洋水文数据集中相互完全独立要素(如经度、纬度、深度、时间)称作维度(dimension), 相关要素(如温度、盐度、密度和声速等信息)称为变元(variate) [5]。

定义 1: 海洋水文数据集中的一条记录可以看作定义在 4 维空间 D 中的 k 元函数 $F(X)$ 。其中, $F = (f_1, f_2, \dots, f_k)$ 由 k 个变元组成, 表征温度、盐度、密度、声速等 k 个物理量; $X = (xlon, xlat, xdep, xtime)$ 为 4 维空间 D 中一点, $xlon$ 表征经度, $xlat$ 表征纬度, $xdep$ 表征深度, $xtime$ 表征采样时刻。

以数据立方(Data Cube)的视角来看, 每一个最细粒度的数据立方(或称数据单元)记录 k 个物理量, 由 $xlon, xlat, xdep, xtime$ 4 个维度坐标定位(图 1)。

2.2. 多元数据雷达图绘制

定义 2: 当 $X = (xlon, xlat, xdep, xtime) = \text{常量 } C, F(C)$ 只描述某个数据单元中 k 个变元表征的物理量。在二维空间中用等角度的 k 条半径表示 k 个变元的坐标轴, $F_i = (f_{1i}, f_{2i}, \dots, f_{ki})$ 的 k 个属性值对应到 k 条坐标轴上, 再将 k 条坐标轴上的点用连续线段连接起来表示 F_i 。

在平面直角坐标系中, 以 O 为圆心绘制圆形, 将圆周切成 k 等分, 连接圆心与圆周上各个点, 形成 k 条半径, 将 k 条半径分别定义为温度、盐度、密度、电导率、声速等各个变元的坐标轴; 每个坐标轴上点的位置由物理量值与该物理量最大值的比例决定; 折线连接所有坐标轴上的点, 形成雷达图(Radar Chart)。从雷达图(图 2)中可以看到所有坐标轴连接到共同的原点(圆心), 其布局沿圆周等角度分布, 线段的长短代表温、盐、密等数值的大小, 图形区域的整体形状和大小反应了某时刻某采样点海洋水文数据的特征。

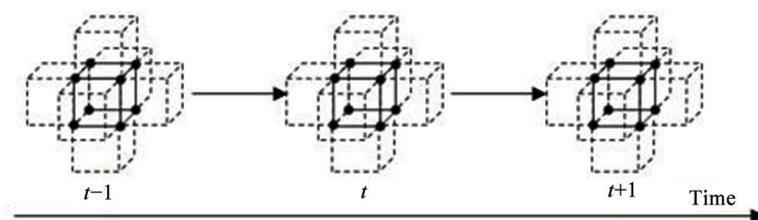


Figure 1. Data Cube model

图 1. 数据立方体模型

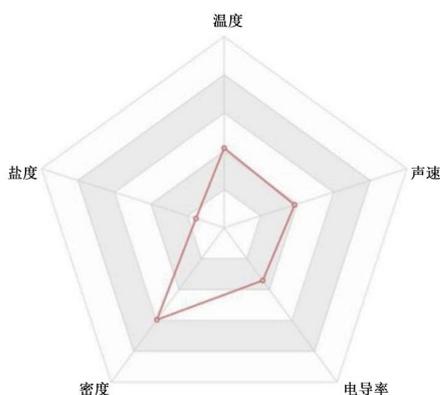


Figure 2. Radar Chart of ocean hydrological data

图 2. 某时刻某采样点水文数据雷达图

2.3. 雷达图序列可视化方法

定义3: 当 $X = (xlon, xlat, xdep, xtime)$ 中有 3 个维度为常量 C , 只有一个维度为变量, $F(X)$ 描述 k 个变元表征的温度、盐度、密度、声速等物理量沿该维度(经度或纬度或深度或时间维度)的变化。

雷达图基于二维坐标提供了一种多维数据表示方法, 本文针对海洋高维多元数据对雷达图进行改进, 通过添加时空维度坐标轴, 将雷达图扩展到三维空间, 绘制由二维坐标、一个时空维构成的雷达图序列 Radar Chart Series, 即由一组雷达图沿一个时空维度构成平面簇。从时空维度(只研究 1 个维度变化的情况)的角度, 同一个时空维度采样的数据组成海洋要素多元数据, 利用雷达图绘制在同一平面上; 不同时空维度的采样数据表现为多个平面, 形成如图 3 所示的平面簇。

添加时空维度后, 将二维雷达图拓展到三维空间, 在增大了数据表现容量的同时, 提供了观察分析数据的新视角。绘制时采用两种表现形式: 1) 离散映射, 海洋水文时空多维数据表现为平面簇中的一系列空间折线形成的多面体轮廓, 视线方向好比轴线将这些平面像糖葫芦一样串起, 也可以理解为将多个雷达图形沿某一时空维度拉伸形成拉花(节日装饰用的手工艺品); 2) 嵌套映射, 对于沿时空维度(比如深度)变化差异大的数据集, 可沿视线方向将不同平面叠压在一起形成嵌套雷达图序列, 用颜色对不同雷达图进行区分, 如图 4。

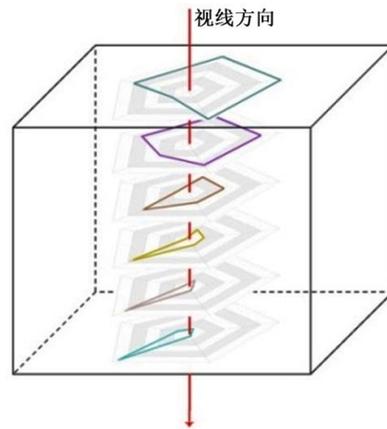


Figure 3. Radar Chart Series in discrete mapping
图 3. 离散映射雷达图序列

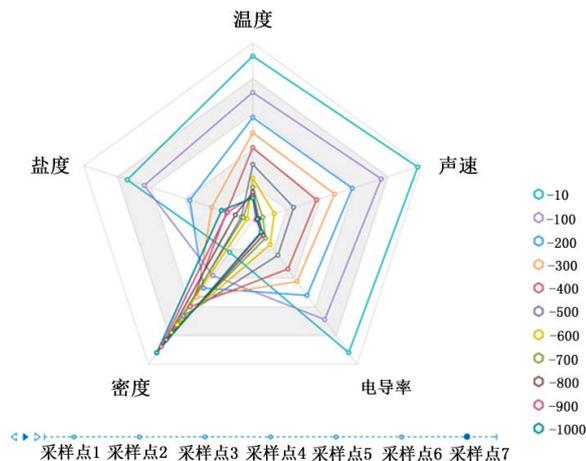


Figure 4. Radar Chart Series in nesting mapping
图 4. 嵌套映射雷达图序列

3. 实验结果分析

3.1. 数据描述

本文选取 2014 年太平洋上 7 个测站点(纬度接近, 经度不同)温度、盐度、密度数据进行实例分析。数据集提供了经、纬、深度坐标以及海洋水文相关的温度、盐度、密度、电导率、声速 5 个要素, 每个要素的取值范围和精度都是不同的: 其中“温度”的取值范围是 0~30, “盐度”取值范围是 34~36, “密度”取值范围是 1020~1033, “电导率”取值范围是 30~59, “声速”取值范围是 1480~1543。

由于实验数据 5 个要素的取值范围的差异, 会使绝对值相对大的变量淹没绝对值相对小的变量, 所以对各要素进行规范化预处理。借鉴文献[9]的思想, 首先对实验数据进行无效数据剔除, 然后使用自适应归一方法将要素数值映射到 0~1 之间。

3.2. 雷达图序列绘制

本文用数据的平均值来度量数据的中心变化趋势, 以 100 米为间隔来抽象描述, 将大量数据以简洁的形式显示出来, 以此来揭示数据的变化规律, 同时可以减少折线交叠带来的影响。如图 5 所示, 采用嵌套映射的雷达图序列采用不同的颜色来表示深度维度信息变化, 简单明了的显示出不同采样点温度、盐度、电导率、声速等水文要素随深度维度变化而变化的总体趋势。

采用离散映射的雷达图序列, 如图 6 所示。通过绘制同一要素随维度变化的轮廓图, 纵向比较折线趋势、幅度, 展现样本数据动态变化的态势。

3.3. 结果分析

综合雷达图序列显示结果, 比对温盐散点图(见图 7), 分析如下:

1) 从各站点 10 m 层到 1000 米层的水文要素分布可见, 随着深度的增加各要素呈明显梯度分布, 7 个采样点雷达图形状变化趋势一致, 有明显的时空变化。

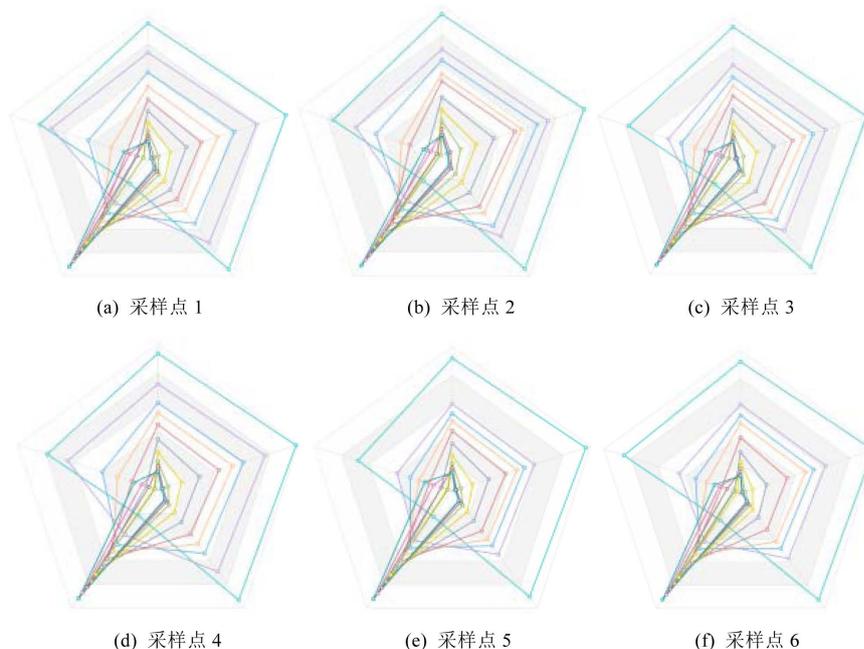
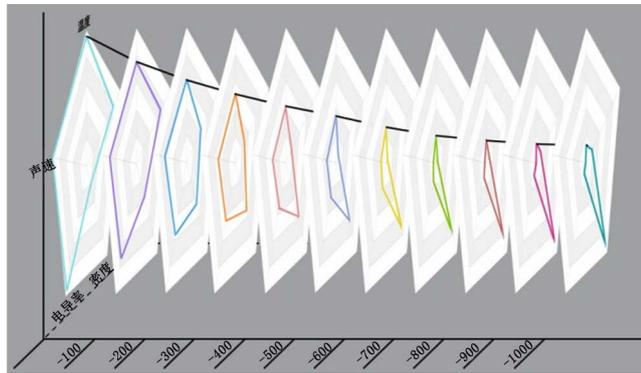
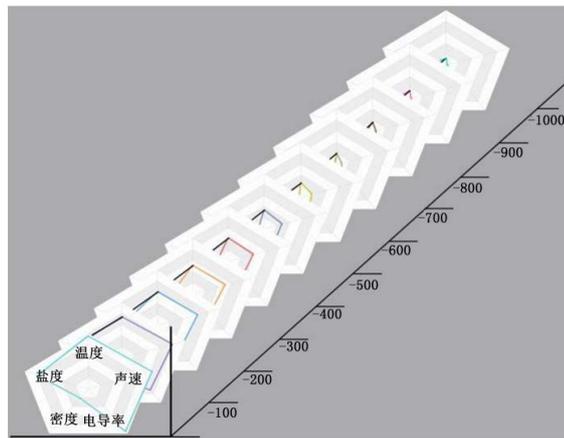


Figure 5. Radar Chart Series in nesting mapping
图 5. 不同采样点雷达图序列(嵌套映射)



(a) 侧视图



(b) 前视图

Figure 6. Radar Chart Series in discrete mapping
图 6. 某一采样点雷达图序列(离散映射)

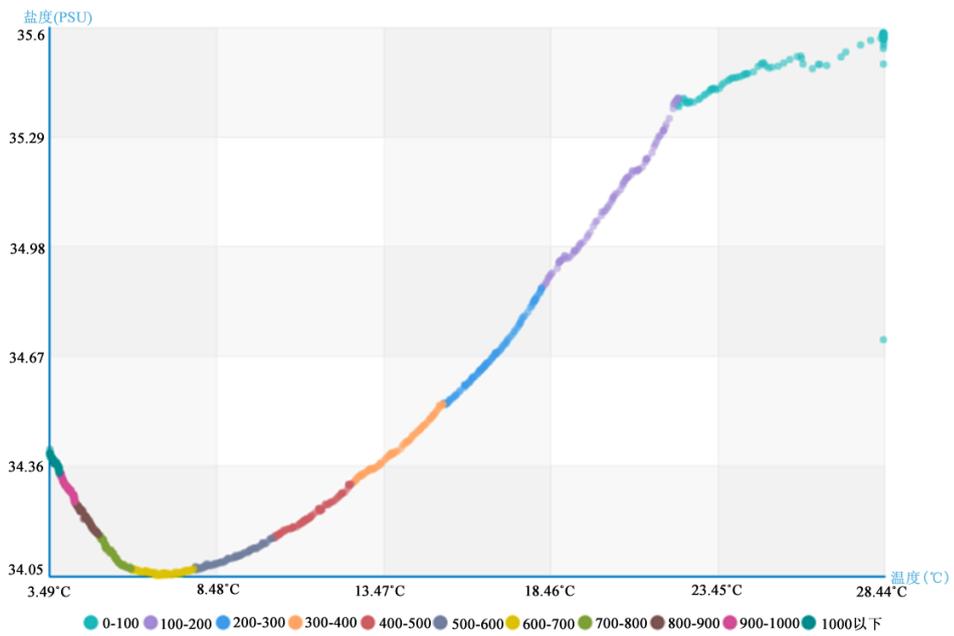


Figure 7. The distribution of thermohaline scatter plot
图 7. 某一采样点温盐散点图

2) 温度、电导率、声速 3 个要素分布趋势为: 随深度增加呈现数值减少的线性变化趋势, 最高值出现在表层; 相反, 密度随深度增加呈现数值增加的线性变化趋势。

3) 盐度随深度增加呈现非线性变化, 在小于 600 米时呈现不均匀减少趋势, 在大于 600 米时呈现不均匀增加趋势。

4) 利用多边形的相似性对水文要素进行定性聚类分析, 表明电导率、声速呈明显的正相关。

4. 结论

传统的海洋水文数据可视化方法只能对单一要素进行显示, 最多叠加两个要素(如温盐), 而多维数据可视化方法通常只在二维平面坐标系上解决了多要素的数据显示。鉴于此, 本文提出了一种基于雷达图序列的可视化方法, 在雷达图的基础上加入时空维度坐标, 既保留了原来雷达图可视化技术的优点, 又弥补了其不能在时空维度进行数据挖掘的缺陷。经过仿真实验验证该方法较传统方法有以下两点进步:

1) 离散映射的雷达图序列融合了剖面图和雷达图各自的信息, 既在每个雷达图表示多个水文要素的数据信息, 又在时空维度上反映了各要素沿某一时空维度的变化趋势, 能够在全局上展现整个数据集分布特征; 2) 嵌套映射的雷达图序列中坐标轴间线段的斜率还可以表示各个研究对象(要素)之间可能存在的关系。实验表明, 用雷达图序列方法来研究海洋高维多元数据在很大程度上拓宽了一个图所能承载的数据信息, 对数据中的隐藏信息进行了更深层次的挖掘, 为海洋数据进一步分析、决策起到重要作用。

当前, 随着观测技术的提高、手段的丰富, 粒度越来越细的与海洋有关的气象、水文数据已经成为大数据的一个重要分支, 如何有效分析、处理是业内的热点方向。本文提出的方法是对高维多元数据进行可视化的一种尝试, 下一步的研究方向是研发出海洋水文数据可视化分析、处理的原型系统。

参考文献

- [1] 李文庆. VVOcean 海洋环境仿真与海洋数据动态可视化系统的研究与实现[D]: [博士学位论文]. 青岛: 中国海洋大学, 2011: 3-5.
- [2] 林丽娜. 南极普里兹湾及邻近海域水文特征分析[D]: [硕士学位论文]. 北京: 国家海洋局第一研究所, 2013: 14-18.
- [3] 陈为, 沈则潜, 陶煜波. 数据可视化[M]. 北京: 电子工业出版社, 2013: 65-168.
- [4] Inselberg, A. (1985) The Plane with Parallel Coordinates. *The Visual Computer*, **1**, 69-91. <https://doi.org/10.1007/BF01898350>
- [5] Santos, S.D. and Brodli, K. (2004) Gaining Understanding of Multivariate and Multidimensional Data through Visualization. *Computers & Graphics*, **28**, 311-325.
- [6] 孙扬, 封孝生, 唐九阳, 等. 多维可视化技术综述[J]. 计算机科学, 2008, 35(11): 1-6.
- [7] 姚中华, 张龙飞, 宋汉辰. Parallel Cluster: 一种时序多维数据的可视化方法[J]. 系统仿真学报, 2013, 25(9): 2135-2139.
- [8] 李建忠, 崔静, 刘彦兵, 等. 基于几何坐标法的多维数据可视化技术在地质数据处理中的应用[J]. 地学前缘(中国地质大学(北京)), 2012, 19(4): 159-164.
- [9] 王鹏. 基于 HLA 的空间环境要素建模与仿真技术研究[D]. 郑州: 郑州信息工程大学, 2006.

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2161-8801，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：csa@hanspub.org