

我国主要江河和海域水质状况 比较和趋势分析

赵博娟 崔东华 秦成

内容提要: 我国的江河和海域存在着严重的水质污染问题,致使水环境面临危机。本文利用广义相对风险的概念和 Cox 比例风险模型,量化比较了我国主要江河流域之间、近岸各海域之间以及全海各海域之间水质的相对优劣,并分析了各江河流域和海域水质的变化趋势。结果表明,虽然我国主要江河流域的水质在总体平均意义上是逐年变好,但是我国近岸海域以及全海海域的水质却逐年显著变坏。特别地,按近岸污染面积数据,黄海近岸海域水质逐年显著变好,而渤海、东海和南海近岸海域的水质逐年显著变坏;按全海污染面积数据,渤海、东海、南海和黄海全海海域水质都逐年显著变坏。无论是近岸还是全海海域,东海污染最为严重。

关键词: 水污染;水质;广义相对风险;Cox 比例风险模型

中图分类号: C812 文献标识码: A 文章编号: 1002-4565(2014)01-0078-06

Comparison and Trend analysis of Water Quality for Main Rivers and Sea Area in China

Zhao Bojuan Cui Donghua Qin Cheng

Abstract: Serious water pollution problems exist in the rivers and seas in China, and our water environment is facing unprecedented crises. Using the generalized concept of relative risk and the Cox proportional hazards model, we quantify the water quality differences among the major rivers, the coastal waters and the seas, and study the changing trend over years. The results show that the water quality in the rivers improves in general (under the sense of weighted average), but that in the coastal waters and the seas are deteriorating significantly. In particular, based on the data on polluted areas (in squared kilometers) in the coastal waters, the water quality in the Yellow Sea improves, but that in the Bohai Sea, the East China Sea and the South China Sea deteriorates over the years; based on the data on the polluted areas in the seas, the water quality in the Bohai Sea, the East China Sea, the South China Sea and the Yellow Sea all deteriorate over the years. Among the coastal waters or the seas, the East China Sea is the most severely polluted.

Key words: Water Pollution; Water Quality; Generalized Relative Risk; Cox Proportional Hazards Model

一、引言

自 20 世纪 50 年代以来,国内外学者对水体环境质量进行了深入的研究,各类水质监测评价方法逐步建立与完善。Chen 等(2012)^[1]采用非负约束因子分析和支持向量机的方法分析了中国晋江流域水污染的状况。Zhang 等(2009)^[2]采用了因子分析、主成分分析以及聚类分析等多种多元统计方法对湖南湘江水质状况进行了分析。Zhang 等(2013)^[3]采用 CWQII 法对中国华县的河流水质进

行了评价,并比较了华县段支流的水质。柯丽娜等(2013)^[4]采用可变模糊识别模型对海水环境质量进行了评价。楼文高(2001)^[5]采用人工神经网络模型对海水水质进行了评价研究。史晓新和夏军(1997)^[6]在灰色关联度的基础上,以定义的关联差异度为测度,并结合水质综合指数构造一种新的水环境质量评价灰色模式识别模型。杨磊磊等(2012)^[7]分别使用改进内梅罗指数法和模糊综合评价法进行水质评价,指出内梅罗指数法更适用于掌握水体被污染程度,模糊综合评价法更适用于评

价水体功能及水体类型。潘晓辉等(2008)^[8]采用模糊综合评价模型对2005年黄浦江6个监测断面水质状况进行评价。李名升等(2012)^[9]比较了常用的水环境质量评价方法,并利用长江流域和淮河流域控断面的水质监测数据对单因子评价法等七种常用的方法进行了实证研究。总的来说,针对不同水环境质量问题,所使用的统计方法和统计模型也不同。

本文着重分析我国主要江河流域之间以及不同海域之间水质的相对好坏,量化估计水质随年份的变化趋势。文献中,关于这些数据的使用仅限于引用各类水质占总体的百分比,没有系统分析的文章。由于数据中对水质的评估结果是有序分类变量,本文分析中采用了Zhao(2013)^[10]提出的广义相对风险的概念和Cox比例风险回归模型,下文中将对有关数据和模型进行详细的介绍。

二、数据与模型

(一) 数据来源

本文所使用数据均来源于国家统计局网站上公布的环境统计数据2003-2011^①,其中包括五套对我国江河流域和海域水质评价的数据,见表1。

表1 我国江河流域和海域水质评价的数据

数据	内容
数据(1)	主要水系干流水质状况评价结果(按监测断面统计)2003-2011年数据
数据(2)	流域分区河流水质状况评价结果(按评价河长统计)2003-2010年数据
数据(3)	全国近岸海域海水水质评价结果(按海水类别分)2003-2008年数据
数据(4)	全国近岸海域海水水质评价结果(按海域污染程度分)2003-2010年数据
数据(5)	全海域海水水质评价结果2003-2011年数据

注:由于缺失或统计指标发生变化,有些年份的数据未纳入本文的分析。

如表1所示,数据(1)和数据(2)的评价标准是《地表水环境质量标准(GB3838-2002)》,即按I类~劣V类,共六个类别进行评价,其中数据(1)是按照监测断面采用单因子评价法,即根据评价时段内该断面参评的指标中类别最高的一项来确定水质类别。数据的格式是按I~劣V,各类水质占全部评估断面的百分比形式。而数据(2)则是按所检测的主要江河和流域的河长(单位:公里)按照I~劣V,各类水质的河长所占评价的总河长的百分比形式。

数据(3)、数据(4)和数据(5)的分类标准按照

《海水水质标准 GB3097-1997》,即将水质分为四类:第I类、第II类、第III类和第IV类。实践中将海水水质分为五类:①清洁海域(符合第I类海水水质),即适用于海洋渔业水域、海上自然保护区和珍稀濒危海洋生物保护;②较清洁海域(符合第II类海水水质),即适用于水产养殖区、海水浴场、人体直接接触海水的海上运动或娱乐区,以及与人类食用直接有关的工业用水区;③轻度污染海域(符合第III类海水水质),即适用于一般工业用水区和滨海风景旅游区;④中度污染海域(符合第IV类海水水质),即适用于海洋港口水域和海洋开发作业区;⑤严重污染海域(劣于第IV类海水水质),见国家海洋局2009年中国海洋环境质量公报^[11]和环境统计数据2006中主要统计指标解释。

数据(3)是根据不同监测点观测的水质结果,按第I类~劣于第IV类,各类水质占总监测点个数的百分比形式;数据(4)是按海水污染程度,较清洁海域~严重污染海域,即第II类~劣于第IV类标准,各类污染级别海域面积(单位:平方公里);数据(5)与数据(4)的标准相同,给出各污染级别的海域面积。

按观测点的污染数据(1)和数据(3)相当于抽样数据,结果受抽样点影响较大,但数据观测全面。按江河测量长度和按海域面积的污染数据(2)、数据(4)和数据(5)可作为数据(1)和数据(3)的重要补充。

虽然这五套水质评价结果数据都是有序分类的,但要量化比较不同江河流域或海域之间的水质状况,或量化研究某一江河或海域水质状况随年份的变化却不太容易。需要寻找合适的统计模型。

(二) Cox比例风险回归模型及应用拓展

1. Cox比例风险回归模型。

Cox比例风险回归模型(Cox proportional hazards model,简称Cox回归模型)是一种生存分析方法,由Cox在1972^[12]年提出。此模型可以用来分析多种因素或风险因子对生存时间的影响,可以处理有截尾的生存时间。Cox比例风险回归模型的一般形式为:

$$h(t; X) = h_0(t) \exp(X^T \beta) \quad (1)$$

其中, $X^T = (x_1, x_2, \dots, x_p)^T$,是对某个个体的 p 维观测值, $h(t; X)$ 是自变量观测值为 X 的个体在

① <http://www.stats.gov.cn/tjsj/qtsj/hjtjzl>.

生存时间 $T = t$ 时的风险函数 $h_0(t) = h(t, X = 0)$ 是在 $X = 0$ 情况下的基准危险率函数, $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ 是未知参数。

在估计回归未知系数 β 时, 需要构造偏似然函数^[13]。

$$L(\beta | (X, T)) = \prod_{k=1}^K \frac{\exp(X_k \beta)}{\sum_{j \in R(t_{(k)})} \exp(X_j \beta)} \quad (2)$$

这里 $t_{(1)} < t_{(2)} < \dots < t_{(k)}$ 是 k 个数值不同的失败时间 (failure times), X_j 是第 j 个个体的观测值, $R(t_{(k)})$ 是在时间 $t_{(k)}$ 时的风险集。

在利用该模型进行统计分析之前, 需要验证模型 (1) 是否符合比例风险的假定。一般地, 可以利用 Kaplan-Meier Curves 和 $\log(-\log(\text{survival}))$ 图画察看比例风险的假定是否基本符合。

2. Cox 比例风险回归模型的应用拓展。

Cox 比例风险回归模型中, 偏似然函数式 (2) 仅与生存时间的长短顺序有关, 对生存时间的分布没有特殊要求。因此, 该模型应用很广泛。为了进一步拓展 Cox 回归模型的应用, Zhao (2013) 提出了广义相对风险的概念, 将此模型推广到分析一般连续因变量和某些有序分类因变量数据。推广中的几个核心结论如下:

(1) 一般的连续型随机变量可以看作生存分析中时间事件数据 (time to event data) 的特例。生存分析中的因变量一般用两个变量描述, 连续型变量生存时间和 0-1 事件指示变量。一般地用 “1” 代表事件 (如死亡) 发生, “0” 代表数据丢失或被截尾。对于一般的连续型变量如收入, 没有 “绑定” 的事件发生。收入数据或者被观测或者缺失, 不像生存时间那样, 个体信息在研究终止之前消失; 截尾数据仅在人为安排下才能发生。因此, 一般连续型随机变量可以看作生存时间数据的特例。

(2) 给出了广义的风险函数的概念。如果 T 是生存时间, $h(t)$ 是 t 时刻死亡的风险。

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = -\frac{S'(t)}{S(t)}$$

对于一般的连续型随机变量 T , 比如收入, 仍然有意义, 可以看作收入为 t 时, 收入不再增高的风险函数。相应地, 对于生存时间 T $h(t)$ 也可以解释为不长寿的风险函数。

(3) 给出了广义相对风险的概念, 简称相对风险, 即 Cox 回归模型中死亡的相对风险可以解释为

不长寿的相对风险。如果 T 是收入, 可以解释为收入不高的相对风险。对于一般的连续型随机因变量, 都可以有类似的相对风险的定义。

(4) 对于满足成比例风险式 (1) 的模型, Cox 回归模型中的参数含义就可按照 (3) 解释。对于不太满足比例风险的情况, 通过研究 Cox 回归模型中的参数和 Mantel-Haenszel 估计的关系, 从 Cox 回归模型中得到的相对风险可以解释为一些相对风险的加权平均。换句话说, 所得到的相对风险可以理解为在总体平均意义下的数值。

(5) 尽管生存时间在理论上是连续变量, 但只要观测单位 (如天数、小时等) 生存时间就被离散化了, 会出现相同的生存时间, 要用 Breslow 方法^[14] 或 Efron 方法^[15] 求解模型中的未知参数。许多有序分类变量也可以看成连续型随机变量的观测。如 “水质好坏程度” 可以看成连续型变量, 由于难以观测, 用有序污染分类标准 I ~ V 度量。如果将 I ~ V 类分别用数值 1 ~ 5 代替, Cox 回归模型中得到的结果可以解释为水质不更坏的风险。

本文主要是利用 Cox 回归模型及广义相对风险的概念, 分析和解释我们所关心的五套江河流域和海域的水质状况。

三、主要江河和海域污染状况比较和趋势分析

(一) 基于监测断面的七大江河水质评估

对 2003 - 2011 年度的主要水系干流 (按监测断面统计) 表中的数据, 我们用 Cox 回归模型分析了每条江河水质 (按 I 类 ~ 劣 V 类六个类别) 随年份变化情况, 也分析了所有江河水质数据放在一起时, 各江河水质污染的差异和总体平均随年份的变化情况, 结果见表 2。

从表 2 的结果可以看出, 运用 Efron 方法和 Breslow 方法得出的结论是一致的。随着年份的增加, 每一年各条江河的水质比上一年不更坏的相对风险都大于 1, 其中黄河、松花江和淮河水质逐年显著变好, 但长江、珠江、海河和辽河水质随年份变化不显著。

将所有江河数据放在一起, 以长江作为比较基线, 发现七大江河中, 水质由好到坏的顺序为: 长江、珠江、黄河、松花江、淮河、辽河、海河, 其中珠江和长江的水质差别不显著, 其他的江河水质显著不如长

表 2 水质不更坏的相对风险(Relative Risk , 按监测断面的污染情况)

每条江河水质随年份变化			
名称	自变量	相对风险(Efron)	相对风险(Breslow)
长江	年	1.021(0.981 ,1.062)	1.015(0.976 ,1.056)
黄河	年	1.139(1.078 ,1.204)	1.107(1.048 ,1.169)
珠江	年	1.049(0.986 ,1.116)	1.036(0.975 ,1.102)
松花江	年	1.094(1.022 ,1.171)	1.072(1.003 ,1.146)
淮河	年	1.109(1.046 ,1.176)	1.093(1.032 ,1.157)
海河	年	1.023(0.946 ,1.107)	1.022(0.945 ,1.105)
辽河	年	1.061(0.993 ,1.133)	1.054(0.987 ,1.125)
所有江河水质的相互比较及随年份变化			
自变量		相对风险(Efron)	相对风险(Breslow)
黄河 vs. 长江		0.484(0.417 ,0.562)	0.544(0.468 ,0.633)
珠江 vs. 长江		0.975(0.825 ,1.153)	0.963(0.815 ,1.138)
松花江 vs. 长江		0.409(0.346 ,0.482)	0.460(0.390 ,0.543)
淮河 vs. 长江		0.350(0.304 ,0.402)	0.411(0.358 ,0.473)
海河 vs. 长江		0.237(0.200 ,0.280)	0.300(0.254 ,0.355)
辽河 vs. 长江		0.288(0.243 ,0.342)	0.355(0.300 ,0.422)
年		1.068(1.045 ,1.091)	1.055(1.032 ,1.078)

江好。调整了江河之间的差异之后,随年份增加,每一年七大江河的水质比上一年不更坏的相对风险显著大于 1,即按总体平均来说,河流的水质逐年显著变好。另外,通过更改比较基线(结果没在表中列出),发现黄河与松花江、松花江与淮河、海河与辽河水水质差异均不显著。

需要指出的是,按监测断面的污染情况数据受样本量大小和采样点位置影响很大,因此在据上述结果给出结论之前,还要分析按河长污染评估数据。

(二) 基于河流长度的主要江河流域水质评估

假设这套数据中每条江河或江河区的水质数据都是按每公里测量或观测得到的,将所测得的公里数作为权重,得到表 3 的结果。

在表 3 中,两种方法结论基本一致:松花江、黄河区、长江区、珠江区这些江水区的水质随年份显著变坏;松花江区、海河、珠江的水质随年份变化不显著,剩余江河或江河区水质都逐年显著变好。

与按照监测断面的结论相比较,只有松花江水质随年份变好还是变坏的结论不一致。从具体水质监测数据上看,按(I , II , III , IV , V ,劣 V) 的顺序,从监测点得到的 2003、2004、2009、2010 和 2011 年数据分别为: (0 , 0 , 3 , 10 , 1 , 1) 、 (0 , 0 , 4 , 11 , 0 , 0) 、 (0 , 5 , 12 , 20 , 1 , 4) 、 (0 , 0 , 8 , 3 , 0 , 0) 和 (0 , 0 , 8 , 3 , 0 , 0) ,水质似有变好的趋势;而按长度得到的 2003、2004、2009 和 2010 年数据分别为 (107 , 911 , 3777 ,

表 3 水质不更坏的相对风险(Relative Risk , 按所评价的河长)

每条江河或江河区水质随年份变化			
名称	自变量	相对风险(Efron)	相对风险(Breslow)
松花江区	年	1.002(0.999 ,1.004)	1.002(1.000 ,1.005)
松花江	年	0.988(0.985 ,0.991)	0.991(0.988 ,0.994)
辽河区	年	1.040(1.036 ,1.045)	1.036(1.032 ,1.041)
辽河	年	1.047(1.040 ,1.053)	1.044(1.037 ,1.050)
海河区	年	1.005(1.002 ,1.008)	1.005(1.002 ,1.008)
海河	年	0.997(0.994 ,1.000)	0.997(0.994 ,1.001)
黄河区	年	0.989(0.986 ,0.991)	0.989(0.987 ,0.992)
淮河区	年	1.058(1.055 ,1.060)	1.051(1.048 ,1.053)
淮河	年	1.055(1.052 ,1.058)	1.048(1.045 ,1.050)
长江区	年	0.981(0.979 ,0.982)	0.984(0.983 ,0.986)
太湖	年	1.041(1.036 ,1.047)	1.037(1.032 ,1.043)
东南诸河区	年	1.005(1.001 ,1.009)	1.004(1.000 ,1.008)
珠江区	年	0.993(0.990 ,0.995)	0.994(0.992 ,0.996)
珠江	年	1.003(1.000 ,1.005)	1.001(0.999 ,1.004)
西南诸河区	年	1.028(1.025 ,1.030)	1.030(1.027 ,1.033)
西北诸河区	年	1.088(1.084 ,1.091)	1.064(1.060 ,1.067)
所有江河或江河区水质的相互比较和随年份变化			
自变量		相对风险(Efron)	相对风险(Breslow)
松花江 vs. 松花江区		1.072(1.063 ,1.082)	1.058(1.049 ,1.068)
辽河区 vs. 松花江区		0.784(0.775 ,0.793)	0.816(0.807 ,0.825)
辽河 vs. 松花江区		0.613(0.604 ,0.623)	0.653(0.643 ,0.663)
海河区 vs. 松花江区		0.681(0.675 ,0.687)	0.718(0.712 ,0.725)
海河 vs. 松花江区		0.616(0.610 ,0.621)	0.656(0.650 ,0.662)
黄河区 vs. 松花江区		0.876(0.868 ,0.883)	0.903(0.895 ,0.911)
淮河区 vs. 松花江区		0.814(0.808 ,0.821)	0.840(0.833 ,0.847)
淮河 vs. 松花江区		0.839(0.832 ,0.846)	0.862(0.855 ,0.869)
长江区 vs. 松花江区		1.449(1.439 ,1.459)	1.378(1.369 ,1.388)
太湖 vs. 松花江区		0.586(0.578 ,0.593)	0.627(0.619 ,0.636)
东南诸河区 vs. 松花江区		1.409(1.394 ,1.425)	1.347(1.333 ,1.363)
珠江区 vs. 松花江区		1.461(1.450 ,1.473)	1.373(1.362 ,1.384)
珠江 vs. 松花江区		1.369(1.357 ,1.380)	1.300(1.289 ,1.311)
西南诸河区 vs. 松花江区		2.490(2.468 ,2.511)	2.033(2.016 ,2.051)
西北诸河区 vs. 松花江区		3.354(3.323 ,3.385)	2.720(2.695 ,2.745)
年		1.011(1.010 ,1.012)	1.009(1.008 ,1.010)

2732 , 536 , 866) 、 (106 , 1387 , 3070 , 3864 , 498 , 1662) 、 (75 , 962 , 4018 , 2767 , 940 , 1923) 和 (75 , 891 , 5032 , 2210 , 408 , 2114) ,水质似有变坏的趋势。显然在水质污染严重的流段,监测点少了。由此可见,用表 1 中的数据(2) 可以弥补数据(1) 的不足。

将所有江河和江河区数据放在一起分析,发现它们按水质由好到坏的排序为: 16 - 西北诸河区 , 15 - 西南诸河区 , 13 - 珠江区 , 10 - 长江区 , 12 - 东南诸河区 , 14 - 珠江 , 2 - 松花江 , 1 - 松花江区 , 7 - 黄河区 , 9 - 淮河 , 8 - 淮河区 , 3 - 辽河区 , 5 - 海河区 , 6 - 海河 , 4 - 辽河 , 11 - 太湖 , 其中,西北诸河区的水质显著好于其他江河和江河区,太湖的水质显著比其他江河和江河区差。经调整所有江河之间的差异后,随着年份的增加,每一年这些江河和江河区比上一年不更坏的相对风险显著大于 1,即总体平均来说,这些江河流域水质逐年显著变好,这与按监测断

面数据得出的结论一致。通过调整不同的比较基线(结果未列在表中)发现:辽河与海河、长江区和珠江区的水质差异均不显著。

主要江河水质由好到坏的次序与按监测断面数据得出的结论也基本一致。

(三) 近岸海域水质评估

利用近岸海水水质监测点数据(3)(见表1)和Cox回归模型得到表4。结果表明用Efron和Breslow两种方法得到的结论一致:除了渤海海域的水质变化随年份增加不显著外,其他三个海域水质显著变好。

表4 近岸海域水质不更坏的相对风险(Relative Risk,按水质监测点)

每个近岸海域水质随年份变化			
名称	自变量	相对风险(Efron)	相对风险(Breslow)
渤海	年	1.042(0.965, 1.125)	1.039(0.962, 1.123)
黄海	年	1.124(1.046, 1.207)	1.104(1.026, 1.187)
东海	年	1.066(1.004, 1.132)	1.062(1.000, 1.127)
南海	年	1.138(1.076, 1.202)	1.100(1.041, 1.162)
所有近岸海域水质的相互比较和随年份变化			
自变量		相对风险(Efron)	相对风险(Breslow)
黄海 vs. 渤海		1.802(1.514, 2.144)	1.583(1.330, 1.884)
东海 vs. 渤海		0.691(0.592, 0.806)	0.725(0.621, 0.846)
南海 vs. 渤海		1.934(1.656, 2.258)	1.680(1.438, 1.961)
年		1.104(1.070, 1.140)	1.083(1.049, 1.118)

将四个近岸海域的数据放在一起分析,发现近岸海域的水质从好到坏排序为:南海、黄海、渤海、东海,东海水质显著地比其他海域差。调整海域之间的差异之后,随着年份的增加,每一年近岸海域的水质比上一年不更坏的相对风险显著大于1,即总体平均来说,近岸海域的水质逐年显著变好。通过更换不同的比较基线(结果未列出),可知黄海与南海水质差异不显著。

(四) 近海水质评估

假设这套数据中每个海域的水质数据都是按每平方公里测量或观测得到的,将所测量的面积作为权重(单位:平方公里),利用Cox回归模型,得到表5。结果表明,利用Efron和Breslow两种方法得到的结论基本一致:渤海、东海、南海近岸海域的水质是随年显著变坏的,而黄海近海域的水质是逐年显著变好的。

将所有海域数据放在一起分析,得到污染程度由轻到重的顺序是:南海、渤海、黄海、东海,东海水质的污染程度显著地比其他海域严重。调整海域之

表5 近岸海域水质不更坏的相对风险(按照海域污染面积)

每个近岸海域水质随年份变化			
名称	自变量	相对风险(Efron)	相对风险(Breslow)
渤海	年	0.951(0.949, 0.953)	0.961(0.959, 0.963)
黄海	年	1.014(1.012, 1.016)	1.011(1.009, 1.013)
东海	年	0.979(0.978, 0.980)	0.982(0.981, 0.983)
南海	年	0.936(0.934, 0.938)	0.950(0.949, 0.952)
所有近岸海域水质的相互比较和随年份变化			
自变量		相对风险(Efron)	相对风险(Breslow)
黄海 vs. 渤海		0.835(0.829, 0.841)	0.866(0.860, 0.872)
东海 vs. 渤海		0.598(0.595, 0.602)	0.656(0.652, 0.660)
南海 vs. 渤海		1.119(1.112, 1.127)	1.093(1.086, 1.101)
年		0.969(0.968, 0.969)	0.975(0.974, 0.976)

间的差异之后,随年份的增加,每一年近岸海域的水质比上一年不更坏的相对风险都是显著小于1,由此,从总体平均意义来说,近岸海域的水质是逐年显著变坏的。可以看出,用按海域污染面积数据得到的结论更加符合目前人们对近岸海域污染状况的认识。

(五) 全海域水质评估

同样地,假设这套数据中每个全海海域的水质数据都是按每平方公里测量或观测得到的,将面积数作为权重(单位:平方公里),利用Cox回归模型,得到表6。可见用Efron和Breslow两种方法得到了一致的结论:随年份的增加,每一个全海海域的水质比上一年不更坏的相对风险都显著小于1,即每一个全海海域水质都逐年显著变坏。

表6 全海海域水质不更坏的相对风险(按海域污染面积)

每个全海海域水质随年份变化			
名称	自变量	相对风险(Efron)	相对风险(Breslow)
渤海	年	0.967(0.965, 0.968)	0.975(0.973, 0.977)
黄海	年	0.981(0.980, 0.982)	0.985(0.984, 0.987)
东海	年	0.982(0.981, 0.983)	0.985(0.984, 0.986)
南海	年	0.943(0.941, 0.944)	0.956(0.954, 0.957)
所有全海海域水质的相互比较和随年份变化			
变量		相对风险(Efron)	相对风险(Breslow)
黄海 vs. 渤海		0.904(0.899, 0.909)	0.927(0.922, 0.932)
东海 vs. 渤海		0.692(0.689, 0.696)	0.750(0.746, 0.753)
南海 vs. 渤海		0.980(0.974, 0.986)	0.983(0.977, 0.989)
年		0.973(0.972, 0.973)	0.979(0.978, 0.979)

将所有全海海域水质数据放在一起分析,得到全海域水质由好到坏的排序为:渤海、南海、黄海、东海,其中,渤海全海的水质污染程度最轻,显著轻于其他三个海域;调整海域之间的差异后,平均每年全海域水质比上一年不更坏的相对风险都显著小于1,即全海域的水质是逐年显著变坏的。

四、结论和建议

利用广义相对风险的概念和Cox比例风险模型,本文量化分析了五套有关主要江河流域以及海域水质评价的年度数据,有以下主要结论:

1. 按监测断面和所评价的河长两套数据,所得到的结论基本一致,如:从总体平均的意义上说,七大主要江河和十六条江河与江河区水质都逐年变好。特别地,淮河水质逐年显著变好;长江和珠江差别不显著,是七大主要江河中水质最好的;海河、珠江的水质随年份变化不显著;辽河与海河差别不显著,它们是七大主要江河中水质最差的。特别值得关注的是,松花江、黄河区、长江区和珠江区水质逐年显著变坏,而太湖是十六条江河和江河区中水质最差的。另外,松花江上的监测断面取点偏重污染不太严重的流域,造成了水质逐年变好的假象。因此,按河长评价水质的数据是监测断面数据的重要补充。

2. 按近岸海域监测点和近岸污染面积两套数据,得到的有些结论一致,如:黄海近岸海域的水质是逐年显著变好的,东海近岸海域是污染最严重的。按近岸海域监测点数据,从总体平均意义上说,近岸海域水质逐年变好,特别地,东海和南海近岸海域的水质是逐年显著变好的;但是按污染面积数据,近岸海域逐年变坏,特别地,渤海、东海、南海近岸海域的水质是逐年显著变坏的。这也说明,近岸污染严重的海域水质监测点较少。

3. 通过对全海海域污染面积数据的分析,可以看出全海海域的污染程度逐年显著加剧,渤海全海的水质污染程度最轻而东海污染最严重。

4. 近岸海域中,南海比渤海的水质好,而在全海海域范围内比较时,南海水质比渤海水质差。可能的原因是南海深海开发相对于渤海更为严重。

从总体平均的意义上说,我国主要江河流域水质逐年变好,但近岸海域和全海海域的水质却逐年变坏。工业和生活污水排向江河,江河污水最终入海,致使我国海域污染越来越严重。

参考文献

[1] Chen H., etc. Source Apportionment of Water Pollution in the Jinjiang River (China) Using Factor Analysis With Nonnegative Constraints and Support Vector Machines [J]. Environmental Forensics 2012(13): 175 - 184.

- [2] Zhang Q., etc. Assessment of surface water quality using multivariate statistical techniques in red soil hilly region: a case study of Xiangjiang watershed [J]. Environmental Monitoring Assess 2009(152): 121 - 131.
- [3] Zhang H., Wei L. River water quality assessment based on CWQII in Hua county [J]. China Journal of Environmental Sciences 2013 (3): 1477 - 1484.
- [4] 柯丽娜等. 基于可变模糊识别模型的海水环境质量评价 [J]. 生态学报 2013 33(6): 1889 - 1899.
- [5] 楼文高. 海水水质评价的人工神经网络模型研究 [J]. 海洋环境科学 2001 20(4): 49 - 53.
- [6] 史晓新, 夏军. 水环境质量评价灰色模式识别模型及应用 [J]. 中国环境科学 1997 17(2): 127 - 130.
- [7] 杨磊磊等. 改进内梅罗污染指数法和模糊综合法在水质评价中的应用 [J]. 水电能源科学 2012 30(6): 41 - 44.
- [8] 潘晓辉, 吴敏, 王悦. 黄浦江水环境污染风险的模糊数学综合评价 [J]. 四川环境 2008 27(6): 60 - 67.
- [9] 李名升等. 常用水环境质量评价方法分析与比较 [J]. 地理科学进展 2012 31(05): 617 - 624.
- [10] Zhao B. Generalized Concept of Relative Risk and Wider Applications of the Proportional Hazards Model and the Kaplan-Meier Estimator [J]. Communications in Statistics-Theory and Methods. forthcoming.
- [11] 国家海洋局. 2009年中国海洋环境质量公报 [N]. 北京: 国家海洋局 2010.
- [12] Cox, D. R. Regression models and life-tables (with discussion) [J]. Journal of the Royal Statistical Society: Ser. B, 1972(34): 187 - 220.
- [13] Cox, D. R. Partial Likelihood [J]. Biometrika, 1975(62): 269 - 276.
- [14] Breslow, N. E. Analysis of survival data under the proportional hazards model [J]. International Statistical Review 1975(43): 45 - 57.
- [15] Efron, B. The efficiency of Cox's likelihood function for censored data [J]. Journal of the American Statistical Association, 1974(72): 557 - 565.

作者简介

赵博娟,女,天津人,1995年毕业于南开大学,获数理统计专业博士学位,现为天津财经大学教授。研究方向为生物统计、人口统计、应用统计。

崔东华,男,河南商丘人,现为天津财经大学统计学硕士研究生。研究方向为统计。

秦成,男,河南南阳人,现为天津财经大学统计学硕士研究生。研究方向为统计。

(责任编辑:曹麦)