

信息熵理论在水系统中的应用

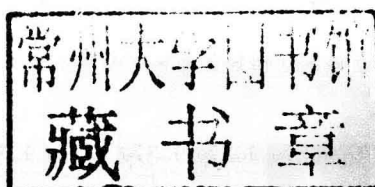
王栋 吴吉春 等著



中国水利水电出版社
www.waterpub.com.cn

信息熵理论在水系统中的应用

王 栋 吴吉春 王远坤 桑燕芳
曾献奎 尚晓三 童 心 丁 昊 著



中国水利水电出版社
www.waterpub.com.cn

内 容 提 要

本书论述了信息熵（Shannon 熵、Tsallis 熵、多尺度熵、样本熵、互熵等）及耦合理论方法（工程模糊集、小波分析、云模型等）在水系统中的不确定性研究与应用，内容包括水质评价模糊优化耦合建模，数据随机观测误差不确定性及影响，大型水库对泥沙和径流的影响，水文序列周期识别、消噪、复杂性分析与随机模拟，地下水流数值模拟不确定性定量评价等。

本书可供水文、水资源、水环境、水文地质以及水利、地理、环境、生态、地质等领域科研人员、工程技术人员与高校师生使用和参考。

图书在版编目（C I P）数据

信息熵理论在水系统中的研究与应用 / 王栋等著
— 北京：中国水利水电出版社，2012. 12
ISBN 978-7-5170-0399-1

I. ①信… II. ①王… III. ①信息熵—应用—给排水
系统—研究 IV. ①TU991

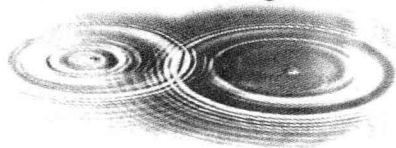
中国版本图书馆CIP数据核字(2012)第290377号

书 名	信息熵理论在水系统中的研究与应用
作 者	王栋 吴吉春 等 著
出版发行	中国水利水电出版社 (北京市海淀区玉渊潭南路1号D座 100038) 网址: www.waterpub.com.cn E-mail: sales@waterpub.com.cn
经 售	电话: (010) 68367658 (发行部) 北京科水图书销售中心 (零售) 电话: (010) 88383994、63202643、68545874 全国各地新华书店和相关出版物销售网点
排 版	中国水利水电出版社微机排版中心
印 刷	北京瑞斯通印务发展有限公司
规 格	184mm×260mm 16 开本 10.75 印张 255 千字
版 次	2012 年 12 月第 1 版 2012 年 12 月第 1 次印刷
印 数	0001—1000 册
定 价	32.00 元

凡购买我社图书，如有缺页、倒页、脱页的，本社发行部负责调换

版权所有·侵权必究

序



水文学是研究地球上的水的发生、循环、分布和时空分布与运动规律的科学，是水资源开发利用与保护的基础学科，是地球科学和水利科学的一部分。不确定性是水文学研究中的一个前沿问题，也是一个难题。特别是在当今气候变化和人类活动影响下，不确定性的问题越来越显著，不确定性研究越来越引起了国内外学者的广泛关注和高度重视。

信息科学的创始人 C. Shannon 为了研究信息的不确定性，把通信过程中信源讯号的平均信息量称为信息熵，使得热力学熵概念的应用领域获得了史无前例的扩展。在信息技术快速发展的今天，信息熵理论和方法在很多学科领域得到了广泛的研究与应用。

对于水文的不确定性，信息熵理论与方法亦显示了独有优势和独特魅力。在运用信息熵理论与方法研究水文不确定性问题方面，欧美学者做了大量工作，取得了很多突出成果。我国在该领域的研究人员较少，但相关研究人员近年不断增加，也取得了一些独具特色的学术成果，有些达到了国际领先的高水平。

本书作者之一王栋教授的博士论文主题是熵及其在水系统中的研究与应用，10 多年来，他坚持不懈地开展这方面工作，已建立起一套较系统的研究模式和研究框架，确定了多个研究方向并均取得了阶段性进展和突破。近年又与国家杰出青年基金获得者吴吉春教授合作，开展了信息熵在地表水和地下水耦合研究中的运用。本书作者结合多年研究和体会，较系统深入地介绍了他们在信息熵及耦合理论方法在水文不确定性研究与应用方面的探索性成果，内容涵盖了水质评价模糊优化耦合建模，数据随机观测误差不确定性影响，大型水库对泥沙和径流影响，水文序列周期识别、消噪、复杂性分析与随机模拟，地下水流数值模拟不确定性定量评价等多方面，有些原创性成果已经发表在 Water Resources Research, Advances in Water Resources, Journal of Hydrology, Stochastic Environmental Research & Risk Assessment, Human and Ecological Risk Assessment 等国际知名杂志和《水利学

报》、《水科学进展》、《水文》等高水平刊物上并获得好评，多个开创性成果使得我国该领域研究在国际上占有了重要的一席之地。更令人欣喜的是，在他们的引领和带动下，一支对运用信息熵理论方法富有热情、朝气蓬勃的科研队伍在南京大学不断发展、壮大。

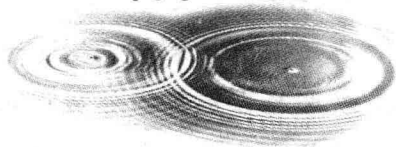
水文不确定性问题的研究任重而道远。希望本书的出版进一步推动信息熵理论在我国的研究与应用，促进水科学及相关领域的科技发展。

中国工程院院士



2012年9月于南京

前言



水是生命之源、生产之要、生态之基。就其本质而言，水文水资源水环境系统（本书称为水系统）是一个开放的复杂巨系统，其复杂性在于：一方面，水系统自身具有一定的确定性变化规律和随机性、模糊性等种种不确定性；另一方面，由于人类活动的介入和影响，进一步加大了水系统的不确定性和风险。

水系统的研究学科——水科学是一门古老的学科，其基本研究内容是自然界中水的存在、运动、变化和分布。从某种意义上来说，水科学是一门有关水信息（采集、传输、整理、分析、研究）的学科。由于自然过程本身的复杂性，观测资料信息的不完备性，认知水平的局限性，目前在水循环机理与规律、水资源开发利用、防洪抗旱减灾、水环境治理修复等方面还存在着一些亟待解决的难题。

熵理论为解决这些难题打开了一扇大门。

熵源于19世纪经典热力学，是一个极其重要的物理量，又以其抽象隐晦、难于理解而著称。实际上，热力学第二定律就是著名的熵定律。100多年来，经过诸多学者的不懈钻研，熵已经成为一个在自然科学、工程技术、社会科学和人文科学中得到全面应用的概念。

熵是衡量系统不确定性程度的一个量度。1948年，信息科学的创始人C. Shannon为研究信息的不确定性，依据热力学中熵的概念，把通信过程中信源讯号的平均信息量称为熵，从而使得熵概念的应用领域获得了史无前例的扩展。这就是信息熵。1957年，E. T. Jaynes发表了一篇里程碑性的论文，首次明确提出了最大熵原理（Principle of Maximum Entropy, POME），并以其成功地解决了信息科学中广为存在的不适定问题（ill-posed problem），由此开创了最大熵原理发展的先河。此后，信息熵理论方法在诸多研究领域得到了广泛应用。

对于解决水系统的诸多不确定性问题,信息熵理论极有前景。近20年,国外特别是欧、美水文学者在这方面做了大量工作。相比之下,国内研究工作要少得多,但近年来也不断增加并相继取得了一些成果。

在河海大学朱元铎教授的指导下,本书著者中的王栋自1998年攻读博士起,就坚持在水科学领域引入和应用信息熵。特别是近几年王栋又与本书著者中的吴吉春进行合作,尝试开展信息熵在地表水和地下水耦合研究中的运用;王远坤、桑燕芳、曾献奎、尚晓三、童心、丁昊等多位年轻教师、科研人员和在读研究生也积极参与到信息熵理论在水系统中的研究与应用。在诸多专家的关心和支持下,目前已有30余篇运用信息熵理论于水科学领域的论文在国内外刊物《水利学报》、《水科学进展》、《水文》、《环境科学学报》、《应用基础与工程科学学报》和 Water Resources Research、Advances in Water Resources、Journal of Hydrology、Stochastic Environmental Research & Risk Assessment、Human and Ecological Risk Assessment、Journal of Coastal Research、Entropy 上发表并获得好评。

因此,我们感到有必要结合自己的研究和体会,撰写此书供大家参考,并希望本书的出版对信息熵理论在我国水科学领域的深入研究与推广应用起到一定的促进作用。

本书共10章。第1章基于最大熵原理和工程模糊集理论,综合考虑了评价过程中随机性和模糊性以及水质状况特点,建立了两种水环境质量评价的模糊优化耦合模型。第2章采用了3种模型,基于Shannon熵、最大熵原理和Tsallis熵,通过两种情况(小观测误差情况和大观测误差情况)探讨了数据随机观测误差不确定性及其影响。第3章基于多尺度熵理论,分析了内嵌在时间序列中的系统复杂性的变化程度,研究了葛洲坝水库对长江泥沙和径流的影响。第4章首先将样本熵与消噪方法耦合,提出了一种自适应确定阈值的小波消噪分析方法,然后运用样本熵理论分析了长江流域干流径流序列系统内部复杂性的变化特征。第5章首先提出了周期识别的两种新方法(分别将熵谱分析方法与小波消噪方法和经验模态分解方法耦合),然后定量描述了序列长度变化时水文序列主周期值和复杂特性的变化规律和特征。第6章基于信息熵理论,提出了水文序列消噪小波系数的阈值优选熵准则,研究了4个关键因素对熵值计算结果的影响,对比了4个常用小波熵指标的性能。第7章基于互熵分析、频率分析、逐步回归分析和分类树,识别了地下水流模型输出的概率

分布、数值特征的关键影响因子和输出概率分布类型的关键驱动因子。第8章基于信息熵方法和地下水概念模型的不确定性分析,对地下水流数值模拟输出变量的预测不确定性进行了定量评价。第9章基于最大熵原理,构造了月径流序列最大熵模型,随机模拟了黄河花园口52年月径流序列。第10章基于云模型,运用期望、熵和超熵,生成了水体富营养化程度各评价因子隶属于各级别的综合云模型。

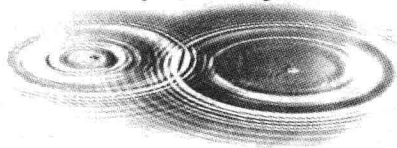
本书研究工作得到了国家自然科学基金(41071018、41030746)、南京大学青年骨干教师和优秀中青年学科带头人培养计划项目的资助。

由于作者水平有限,书中难免出现疏漏甚至错误,恳请读者批评指正。

作者

2012年8月于南京大学

目 录



序

前言

第 1 章 基于最大熵原理和工程模糊集理论的模糊优化耦合模型	1
1.1 引言	1
1.2 基本概念与理论方法	3
1.3 水环境质量评价的模糊优化耦合模型	5
1.4 应用实例	10
1.5 结论	16
第 2 章 基于 Shannon 熵和 Tsallis 熵的数据随机观测误差不确定性及其影响	17
2.1 引言	17
2.2 基于 Shannon 熵、最大熵原理 (POME) 和 Tsallis 熵的实测数据的概率分布	19
2.3 两种观测误差下随机观测误差 (SOE) 的影响	21
2.4 结论	28
本章附录 截尾正态分布	29
第 3 章 基于多尺度熵理论的大型水库对泥沙和径流影响	31
3.1 引言	31
3.2 研究区域概况	31
3.3 资料与方法	32
3.4 结果分析与讨论	33
3.5 结论	41
第 4 章 基于样本熵理论的水文序列小波消噪与内部复杂性变化特征分析	43
4.1 基于样本熵理论的自适应小波消噪分析方法	43
4.2 基于样本熵理论的径流序列复杂性分析	49
第 5 章 基于耦合熵谱分析的水文时间序列周期识别方法与主周期变化分析	55
5.1 水文序列周期识别方法——I (MSSA)	55
5.2 水文序列周期识别方法——II (EEMD-MESA)	68
5.3 水文序列主周期变化原因分析	73

5.4 小结	79
第 6 章 基于信息熵理论的水文序列消噪与复杂性表征	80
6.1 水文序列分析中基于信息熵理论的消噪方法	80
6.2 水文序列复杂特性定量描述	89
第 7 章 基于互熵分析的地下水流模型输出的概率分布特征分析	100
7.1 频率分析	100
7.2 敏感性分析	103
7.3 算例研究	105
7.4 结论	116
第 8 章 基于信息熵方法的地下水流数值模拟不确定性的定量评价	118
8.1 地下水数值模拟不确定性定量评价概述	118
8.2 基于方差方法的不确定性评价	120
8.3 基于信息熵方法的不确定性评价	121
8.4 算例研究	122
8.5 结果的对比与分析	125
8.6 结论	131
第 9 章 基于最大熵原理的大河月径流序列随机模拟	132
9.1 引言	132
9.2 最大熵模拟方法	133
9.3 实例应用	135
9.4 结论	139
第 10 章 基于云模型的水体富营养化程度评价方法	140
10.1 引言	140
10.2 云模型简介	141
10.3 基于云模型的水体富营养化程度评价方法	143
10.4 实例分析与比照	148
10.5 结论	151
参考文献	152

第1章 基于最大熵原理和工程模糊集理论的模糊优化耦合模型

水环境质量评价不可避免随机性和模糊性问题。基于最大熵原理和工程模糊集理论,综合考虑了评价过程中随机性和模糊性以及水质状况特点,建立了两种水环境质量评价的模糊优化耦合模型,将所建模型应用于我国12个代表性湖库营养化等级的综合评判。实例评价结果表明,所提出的方法是针对复杂水环境质量评价问题的优化解决方法,所建模型灵活有效且可靠性较高,既可以提供更多信息供环境决策,又可以降低水环境质量评价中的不确定性。

1.1 引言

随着我国经济的快速发展,水污染问题日益引起广泛关注。水环境质量评价,即对水环境质量优劣的定量描述和评定,目的是正确反映水环境质量和污染状况,指出将来的发展趋势(陆雍森等,1990)。水环境评价是防治水污染的前提和基础,也是进行水质规划的依据,具有重要意义。

如今,在人口增长和人类活动影响下,水体富营养化已成为世界上许多地区一个严重的问题。过高的营养物质输入而产生富营养化问题已被欧洲环境署(EEA)关注:一个生态系统中有机的供给率提高,常会引起系统中养分富集的加强(EEA,2001)。因此,湖库富营养化的有效控制被广泛重视(Somlyódy,1998;Pei and Wang,2003;McIntyre et al.,2003)。

湖库富营养化是复杂的物理、化学和生物过程共同产生的结果。水质不仅取决于自然过程,如降水、地壳物质的侵蚀和风化、生物种群的相互关系,也取决于人为的影响,如城市、工业和农业活动(Papatheodorou et al.,2006)。水体及其周边环境之间的相互作用是复杂的,在流域尺度上研究水体富营养化评价和水环境管理是十分重要的(Hession and Storm,2000)。不确定性、非均质性等是环境研究领域中最重要和最具挑战性的课题(Anderson et al.,2000;Medina et al.,2002;Kirchner et al.,2004;Neal and Heathwaite,2005)。

更要指出,包括随机性和模糊性在内的不确定性,在水环境评价和管理中普遍存在,其重要性已被广泛认可并研究(例如,Beck,1987;Van der Perk,1997;Beven and Freer,2001;Beven,2002;Vrugt et al.,2002;Harris and Heathwaite,2005;Zheng and Keller,2006)。

一方面,由于污染物在水环境中进行的物理、化学、生物过程是随机过程,因而水环



境评价特征值在监测、实验、数据分析和处理过程中不可避免随机性；另一方面，由于评价中的分级标准、评价类别、污染程度等是客观存在的模糊概念和模糊现象，因而水环境评价不可避免模糊性。这两种基本并重要的不确定性在水环境评价模型的建立中应该同时被考虑。

随机性主要表现在基本数据上。以富营养化评价为例，随机性主要反映在如叶绿素 a (Chl-a)、总磷 (TP)、总氮 (TN)、化学需氧量 (COD) 和透明度 (SD) 等富营养化相关数据的监测、实验、数据分析和处理中。

模糊性主要表现在标准的确定上，例如，分类标准、评价标准等的确定。以富营养化评价为例，营养状态的分类在两种级别分界附近就是模糊的 (Vollenweider et al., 1998)。

一般情况下，可以采用 4 种途径研究包括水环境评价在内的不确定性问题：①统计和随机方法；②模糊集方法；③人工智能 (AI) 方法；④耦合方法。这 4 种途径都是基于数据的建模方法。这就意味着各种途径都必须使用高质量的能够如实反映水文和环境过程时空非均质性的数据，以 “get right answers for right reasons” (Kirchner, 2006)。

(1) 统计和随机方法。信息论的创建者 C. Shannon (1948) 开创性地提出：信息具有 (随机) 统计性质。如今，许多统计方法，如多元统计方法 (Yu et al., 2003; Papa-theodorou et al., 2006) 已经被广泛用于水环境评价。主成分分析 (PCA)，也被广泛应用于确定水体富营养化程度 (Vega et al., 1998; Parinet et al., 2004)。

(2) 模糊集方法。模糊集理论的创建者 Zadeh (1965) 提出了后来在许多领域都得到广泛应用的模糊集，借助隶属函数详细表征模糊不确定性。模糊集理论已被用于评价环境质量状况。Silvert (2000) 指出，模糊逻辑可用于环境参数的研究，解决了一些常见问题。许多模糊方法已用于水质评价 (Lu and Lo, 2002; Liou and Lo, 2005; Ghosh and Mujumdar, 2006)。陈守煜 (1998) 扩充了 Zadeh 的模糊集理论，并把它命名为工程模糊集理论 (EFST)，为隶属度和隶属函数的确定提供了新途径。

(3) 人工智能 (AI) 方法。多种人工智能方法被应用在水质评价，如人工神经网络模型 (ANN) (Aguilera et al., 2001; Schulze et al., 2005)，遗传算法 (GA) (Gentry et al., 2003; Kuo et al., 2006) 等。

(4) 耦合方法。主要思路是结合上述两个或更多的途径以建立耦合模型，适用于复杂过程的建模，综合反映不确定性特质。例子包括：耦合模糊集理论和灰色系统理论 (Chang et al., 1996)；耦合信息熵理论和模糊集理论 (张成科, 1998)；耦合自组织映射和模糊集理论 (Lu and Lo, 2002)；耦合人工神经网络和蒙特卡洛方法 (Zou et al., 2002)；以及耦合人工神经网络和经验模型 (Jain and Jha, 2005)。另一种新的耦合模式是耦合信息熵理论和工程模糊集理论 (EFST) (王栋, 2001; Wang, et al., 2007)。

信息论的创建者 Shannon (1948) 开创了信息熵的数学理论。其后，Jaynes (1957) 提出了最大熵原理 (POME)。如今，信息熵，特别是最大熵原理和最大熵谱分析 (MESA) 已经被应用于众多领域。例如，最大熵原理已被用于推导各种分布，并进行参数估计 (Singh, 1998)。Wang 等 (2004) 将 MESA 用于长江口年最高潮位序列分析。熵在水科学领域的应用可参见综述 (Singh, 1997; 2011; 王栋, 朱元牲, 2001)。



本章目的是改进不确定性评价方法并进一步提升对不确定性评估研究必要性的认识。为此,提出了基于最大熵原理和工程模糊集理论的耦合方法,综合考虑了评价过程中随机性和模糊性以及水质状况特点,建立了两种水环境质量评价的模糊优化耦合模型,将所建模型应用于我国 12 个代表性湖库富营养化等级的综合评判。实例评价结果表明,所提出的方法是针对复杂水环境质量评价问题的优化解决方法,所建模型灵活有效且可靠性较高,既可以提供更多信息供环境决策,又可以降低水环境质量评价中的不确定性。

文中有关理论、方法与模型也可以推广应用于其他领域。

1.2 基本概念与理论方法

1.2.1 信息熵

Shannon (1948) 在信息论中提出:用熵 $I(f)$ 度量信息的不确定性,随机变量 X 的概率密度函数 (PDF) 为 $f(X)$

$$I(f) = - \int_a^b f(x) \ln[f(x)] dx \quad (1-1)$$

熵可以用来选择 $f(X)$, 根据约束条件最大限度地减少不确定性。

1.2.2 最大熵原理 (POME)

对于如何选择 $f(X)$, Jaynes (1957) 提出了最大熵原理。其基本思路是:给定约束条件下,在所有可能的(都符合实验数据)的先验概率分布中,以使得 $I(f)$ 取最大值的分布为最优(最为符合实验的物理背景)。

在数学上,它可以表述如下:给定 m 个线性约束条件 C_i

$$C_i = \int_a^b y_i f(x) dx \quad i = 1, 2, \dots, m \quad (1-2)$$

其中 $y_i(X)$ 是 $f(X)$ 的函数,根据方程式 (1-2) 给定条件下 $I(f)$ 的最大值为

$$f(x) = \exp[-\lambda_0 - \sum_{i=1}^m \lambda_i y_i(x)] \quad (1-3)$$

其中 $\lambda_i, i = 1, 2, \dots, m$, 是拉格朗日乘数。且

$$\int_a^b f(x) dx = 1 \quad (1-4)$$

1.2.3 经典模糊集理论:隶属度和隶属函数

1965 年,美国 L. A. Zedeh 教授提出的开创性论文——模糊集合,是对物质系统、现象、概念在发展演化过程中亦此亦彼或模糊性的科学描述,具有深刻的辩证思维和哲理内涵,并由此发展成为一门新兴数学分支学科——模糊集合论或模糊数学。隶属度和隶属函数概念是模糊集合论赖以建立的基石,是模糊集合论中最为重要的基本概念。

在普通集合论中,一个元素要么属于、要么不属于某一集合,用特征函数加以定义。



$$\left. \begin{aligned} \chi_A: U &\rightarrow \{0, 1\} \\ u &\mapsto \chi_A(u) \in \{0, 1\} \end{aligned} \right\} \quad (1-5)$$

$$\text{且} \quad \chi_A = \begin{cases} 1 & u \in A \\ 0 & u \notin A \end{cases} \quad (1-6)$$

在经典模糊集合论中,一般做法是将上述特征函数定义直接推广为隶属函数定义。

定义 1.1 设 u 是论域 U 中的一个元素, A 是 U 的一个模糊子集, 定义 A 的隶属函数 μ_A 为

$$\left. \begin{aligned} \mu_A: U &\rightarrow [0, 1] \\ u &\mapsto \mu_A(u) \in [0, 1] \end{aligned} \right\} \quad (1-7)$$

例如, 令 $A = \{x \mid x \gg 1\}$ (即 A 是实际数值远远大于 1 的模糊集)。那么, 这样的集可以如方程式 (1-8) 所示由隶属函数主观定义:

$$\mu_A(x) = \begin{cases} 0 & x \leq 1 \\ [1 + (x-1)^{-2}]^{-1} & x > 1 \end{cases} \quad (1-8)$$

如果 $x \leq 1$, 例如 $x=1$, 根据方程式 (1-8), $\mu_A(x=1) = 0$, 表示 $x(=1)$ 在 $A(=\{x \mid x \gg 1\})$ 中的隶属度为 0。如果 $x > 1$, 例如 $x=10$, 根据方程式 (1-8), $\mu_A(x) = 0.987805$, 表示 $x(=10)$ 在 $A(=\{x \mid x \gg 1\})$ 中的隶属度为 0.987805。当 $x=1000$, 根据方程式 (1-8), $\mu_A(x) = 0.999999$, 表示 $x(=1000)$ 在 $A(=\{x \mid x \gg 1\})$ 中的隶属度为 0.999999。

模糊集理论的发展过程中, Zadeh 又成功引入模糊系统、模糊逻辑、语言变量和近似推理、模糊信息粒化、模糊逻辑和语言软计算、模糊逻辑和基于感知的理论、不确定性广义理论等 (Zadeh, 2005)。

如今, 模糊集理论已经在许多领域得到了极为广泛的应用, 取得了大量研究成果。

需要指出, 上述隶属函数的定义是以映射的形式从特征函数的定义推广而得的, 把隶属函数的概念绝对化, 且未完整地反映模糊集合论最基本的特征——亦此亦彼或模糊性 (陈宋煜, 1998)。为此, 我国大连理工大学的陈守煜教授、王本德教授等水文学家根据辩证思维与哲理关于差异、共维、中介、两极的观点, 对隶属度、隶属函数的概念进行重新定义, 提出了所谓绝对隶属度、隶属函数和相对隶属度、隶属函数的定义, 并由此建立了工程模糊集理论的框架与体系。

1.2.4 工程模糊集理论中的相对隶属度和相对隶属函数

陈守煜 (1998) 扩展了 Zadeh 的模糊集理论并将它命名为工程模糊集理论 (EFST), 提供了一种新的方法来确定隶属度和隶属函数。

定义 1.2 设 A 为论域 U 上的一个模糊子集/模糊概念, 分别赋给 A 处于共维差异的中介过渡段的两个极点以 0 和 1 的数, 构成 $[0, 1]$ 闭区间数的连续统。在该连续统的数轴上建立参考系, 使其中的任两个点定为参考系坐标上的两极, 赋给参考系的两极以 0 和 1 的数, 并构成参考系 $[0, 1]$ 数轴上的参考连续统。对于任意 $u \in U$, 在参考连续统上指定了一个数 $\mu_A(u)$, 称为 u 对 A 的相对隶属度, 映射



$$\left. \begin{array}{l} \mu_{\tilde{A}}: U \rightarrow [0, 1] \\ u \mapsto \mu_{\tilde{A}}(u) \in [0, 1] \end{array} \right\} \quad (1-9)$$

称为 $\tilde{A}$ 的相对隶属函数。

应用模糊集合理论时,根据实际问题的不同性质和要求,常常只需确定相对隶属度、隶属函数。计算相对隶属度、隶属函数要远比确定绝对隶属度、隶属函数容易,更为重要的是,可以在基本理论与概念上消除在应用领域长期存在的所谓确定隶属度、隶属函数中的“主观任意性”的困惑。这种困惑在很大程度上影响着模糊集合论的应用,尤其在工程领域的应用。

例如,在水环境质量评价领域,富营养化评价依据的分类指标可分为两种类型。一种类型指标是下降型的分类指标,如叶绿素 a (chl-a),总磷 (TP),总氮 (TN),化学需氧量 (COD),这种类型指标的数值越大,相应的营养化水平等级越高。另一种类型指标是上升型的分类指标,如透明度 (SD),这种类型指标的数值越大,相应的营养化水平等级越低。

对于下降型的分类指标,如果使用相对隶属度的基本概念,可分别表示如下:

(1) 对应 1 级,相对隶属度 $s_{i1}=0$,这是根据评价概念 $\tilde{A}$ 的分类标准 i 的标准浓度 y_{i1} 。

(2) 对应 c 级,相对隶属度 $s_{ic}=1$,这是根据评价概念 $\tilde{A}$ 的分类标准 i 的标准值 y_{ic} 。

相对隶属度记为 s_{ih} ,是在级别 h 上的分类标准 i 的标准浓度 y_{ih} ,可以用线性关系表示

$$s_{ih} = \frac{y_{ih} - y_{i1}}{y_{ic} - y_{i1}} \quad (1-10)$$

上升型的分类指标也同样可以表示。

1.3 水环境质量评价的模糊优化耦合模型

1.3.1 数据处理

设水环境评价标准分级数为 c ,评价指标有 m 项,各评价指标的分级标准浓度值为 y_{ih} ,则有水环境评价分级标准浓度矩阵 Y :

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1c} \\ y_{21} & y_{22} & \cdots & y_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mc} \end{bmatrix} = [y_{ih}]_{m \times c} \quad (1-11)$$

现有 n 个待评价水体样本,每一样本有 m 项评价指标的实测值 x_{ij} ,则有待评价水体实测浓度矩阵 X :

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} = [x_{ij}]_{m \times n} \quad (1-12)$$



若规定 1 级评价标准状态中评价指标 i 的标准浓度 y_{i1} 对于某评价概念 $\tilde{A}$ 的相对隶属度 $s_{i1} = 0$; c 级评价标准状态中评价指标 i 的标准浓度 y_{ic} 对于该评价概念的相对隶属度 $s_{ic} = 1$, 则可根据线性公式确定 h 级评价标准状态中指标 i 的标准浓度 y_{ih} 的相对隶属度 s_{ih} :

$$s_{ih} = \frac{y_{ih} - y_{i1}}{y_{ic} - y_{i1}} \quad (1-13)$$

从而将水环境分级标准浓度矩阵 Y 变换为分级标准浓度相对隶属度矩阵 $\tilde{S}$:

$$\tilde{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1c} \\ s_{21} & s_{22} & \cdots & s_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mc} \end{bmatrix} = [s_{ih}]_{m \times c} \quad (1-14)$$

将指标值越大评价等级越大的指标实测浓度值 x_{ij} 转化为相应的相对隶属度 r_{ij} :

$$r_{ij} = \begin{cases} 1 & x_{ij} > y_{ic} \\ \frac{x_{ij} - y_{i1}}{y_{ic} - y_{i1}} & y_{i1} \leq x_{ij} \leq y_{ic} \\ 0 & x_{ij} < y_{i1} \end{cases} \quad (1-15)$$

相反地, 对于指标值越小评价等级越大的指标实测浓度值 x_{ij} 而言, 以下式进行转化:

$$r_{ij} = \begin{cases} 1 & x_{ij} < y_{ic} \\ \frac{x_{ij} - y_{i1}}{y_{ic} - y_{i1}} & y_{i1} \geq x_{ij} \geq y_{ic} \\ 0 & x_{ij} > y_{i1} \end{cases} \quad (1-16)$$

从而将待评价水体实测浓度矩阵 X 变换为水体实测浓度相对隶属度矩阵 $\tilde{R}$:

$$\tilde{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} = [r_{ij}]_{m \times n} \quad (1-17)$$

待评价水体实测浓度相对隶属度矩阵 $\tilde{R}$ 具有明确的物理意义, 它表示了 n 个待评价水体样本的 m 个指标实测浓度值对于模糊概念 $\tilde{A}$: 营养化的作用大小或影响程度, 从数学上表达了所有样本全部指标实测浓度值对于模糊概念 $\tilde{A}$: 营养化的相对隶属度, 给出了样本集的全体指标超过 1 级评价标准浓度的相对隶属度值。 $\tilde{R}$ 在模糊数学中也称为超标权重矩阵。

此外, 各指标对模糊概念 $\tilde{A}$: 营养化的影响也不同, 即每个分类指标在决定营养级别的过程中都有不同的作用。因此, 每个评价分类指标的影响权重被引入。例如, 在 5 个指标中叶绿素 a 起着最重要作用, 所以叶绿素 a 的影响权重应该是最大的。各评价分类指标的影响权重可以用分类指标权重向量 $\vec{v}$ 表示:

$$\begin{cases} \vec{v} = (v_1, v_2, \dots, v_m) \\ \sum_{i=1}^m v_i = 1 \end{cases} \quad (1-18)$$

综合考虑超标权重和评价指标权重,可建立 n 个样本 m 项指标的综合权重矩阵 A :

$$A = \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_m \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} v_1 r_{11} & v_1 r_{12} & \cdots & v_1 r_{1n} \\ v_2 r_{21} & v_2 r_{22} & \cdots & v_2 r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_m r_{m1} & v_m r_{m2} & \cdots & v_m r_{mn} \end{bmatrix} = [v_i r_{ij}]_{m \times n} \quad (1-19)$$

将矩阵 A 的元素按列归一化,得到环境指标综合权重矩阵 W :

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{bmatrix} = [w_{ij}]_{m \times n} \quad (1-20)$$

$$\begin{cases} w_{ij} = \frac{v_i r_{ij}}{\sum_{i=1}^m v_i r_{ij}} \\ \sum_{i=1}^m w_{ij} = 1 \quad j = 1, 2, \cdots, n \end{cases} \quad (1-21)$$

矩阵 W 既考虑了指标的超标权重,又考虑了评价指标自身权重,综合表达了样本集各项指标对于模糊概念 A : 营养化的作用大小或影响程度,称为环境指标综合权重矩阵。

待评价水体样本以不同的相对隶属度隶属于评价各级标准。设 n 个样本对于 c 级水的相对隶属度矩阵为 $\tilde{U}$:

$$\tilde{U} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{c1} & u_{c2} & \cdots & u_{cn} \end{bmatrix} = [u_{hj}]_{c \times n} \quad (1-22)$$

模糊矩阵 $\tilde{U}$ 必须满足的首要约束条件为概率归一化条件:

$$\begin{cases} \sum_{h=1}^c u_{hj} = 1 \\ u_{hj} \geq 0 \quad j = 1, 2, \cdots, n \end{cases} \quad (1-23)$$

显然,这样的矩阵有无数个。进行水环境评价的目的就是确定满足式 (1-23) 的优化模糊矩阵 $\tilde{U}$ 。

1.3.2 模糊优化耦合模型 I (模型 I)

如本章引言所述,由于水环境评价具有随机性和模糊性,因而, $[u_{hj}]_{c \times n}$ 的确定也具有随机性和模糊性等不确定性。

随机性包括了两个方面的含义:其一是应用相对隶属度的概念进行评价过程中的随机性;其二是实测指标观测过程中的观测误差的随机性。对于后者的分析请见第 2 章。现重