

Multi-Source Data Fusion Pretreatment System in Distributed Hydrological Models

Lan Li^{1,2}, Pan Chen¹, Jie Meng¹

¹State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan

²Water Resources Security Collaborative Innovation Center in Hubei Province, Wuhan

Email: lilan@whu.edu.cn

Received: Mar. 31st, 2014; revised: Apr. 11th, 2014; accepted: Apr. 28th, 2014

Copyright © 2014 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

Abstract

Multi-source data fusion pretreatment system which is based on vast digital information, meteorological and hydrological monitoring data and so on is put forward for standardization and modularization needs of multi-purpose distributed hydrological models. The dynamic connection of spatial data and attribute data is created based on a four-level division framework of watershed. Besides, it also achieves seamless connections between GPS, GIS, RS systems and hydrological models from grid level to basin level, and can be quickly applied to different hydrological models. The system is used to build spatial information and topological relationship of distributed and semi-distributed hydrological models in Yangtze estuary, which suggests its convenience and adaptation.

Keywords

Multi-Source Data, Four-Level Division, Topological Relationship, Distributed Hydrological Models

分布式水文模型中的多源数据融合前处理

李 兰^{1,2}, 陈 攀¹, 孟 洁¹

¹武汉大学水资源与水电工程科学国家重点实验室, 武汉

²水资源安全保障湖北省协同创新中心, 武汉

Email: lilan@whu.edu.cn

作者简介: 李兰, 女, 武汉大学, 教授, 博士生导师, 主要从事水文、水资源、环境生态保护研究。

收稿日期：2014年3月31日；修回日期：2014年4月11日；录用日期：2014年4月28日

摘要

从构建多用途流域分布式水文模型的标准化、模块化需求出发，自主研发了基于海量数字化信息和气象水文监测数据等多源数据融合前处理系统。该系统基于一个流域四级划分框架体系，建立了空间数据与属性数据的动态关联。此外，其实现模型与GPS、GIS、RS系统从流域层次到网格层次的无缝耦合，能快速移用于不同的流域分布式水文模型构建中。该系统应用于长江河口地区分布式和半分布式水文模型的空间信息和拓扑关系构建中，证明其方便性和适应性。

关键词

多源数据，四级划分，拓扑关系，分布式水文模型

1. 引言

分布式水文模型是研究洪涝、干旱、水污染和水资源管理等的强有力工具，其构建不仅需要我们对水文物理机制的理解，而且需要大量的流域空间分布信息和海量数据信息及相关技术的支撑。数字高程模型(DEM)、土壤质地空间分布的数字化信息、土地利用的遥感数字信息，在GIS软件平台上可实现地形地貌信息的非专业户化前处理，仍然需要我们与专业数学模型匹配进行前处理，以便可以利用更多的数据信息十分方便地进行水文模拟，以认识水文循环下的水文过程[1]。YE等[2]提出了一个基于遥感信息的时变水文模型(DTVGM)，模型所有输入资料都来源于遥感影像信息，这为一种无资料地区水文及水资源问题的解决提供了一种途径。Kabir等[3]构建了一个空间分布的水文模型(WetSpa)，模拟了Gorganrood流域的日径流变化，模型中充分利用了地形、土地利用及土壤类型等数字信息。此外水资源供、用、耗、排等人类活动影响信息也逐渐被应用到分布式水资源模型的构建中，如吴俊秀和郭清[4]利用MIKE BASIN模型对大凌河流域水资源状况进行了模拟，模型中水资源分配模块考虑了区域水资源的利用信息。

随着大量数据信息被融入到分布式水文模型构建中，弄清这些数据信息在水文模型中的拓扑结构就显得尤为必要，而缺乏能体现数据信息拓扑结构的有效编码方式，将在一定程度上阻碍了分布式水文模型进一步的发展和应用[5]。当前比较通用的子流域编码方法主要有Horton法[6]、Strahler法[7]、Shreve法[8]和Pfafstetter法[9]，基于这几种编码方法的一些编码软件主要有WMS[10]、TOPAZ[11]、Arc Hydro Tools[12]，GRASS[13]等。其中Strahler法是在Horton法基础上，从形态与水文要素综合分析中提出的，并且划分的河道结构具有拓扑关系，因而应用十分广泛，但该编码方式只是一种河网编码，并未将流域坡地汇水网络和多源数据的拓扑关系考虑进去，因此有必要对其进行进一步深化处理。本文就是为解决考虑多源数据拓扑关系的全流域编码问题，提出了一种流域-子流域-网格单元-土地利用分类的四级划分框架体系，其主要适用于基本单元为网格的全分布式水文模型。针对半分布式水文模型的构建，编码方式可简化为流域-子流域-土地利用分类的三级划分框架体系。该体系通过网格单元和土地利用分类的匹配实现了空间数据与属性数据的动态关联。基于该体系构建了流域范围内多源数据融合的前处理系统，该系统使得模型前处理模块化、标准化，方便快捷实现了模型与GIS信息的无缝连接，同时减少了分布式水文模型的存储单元和提高了模型的运算速度。该系统为LILAN分布式水文模型的一个有机组成部分，LILAN分布式水文模型中集合了大量的地形地貌特征数据、人类活动的土地利用和水资源利用

数据信息,该系统为模型的构建提供了坚实的基础[14]。本文最后以长江河口地区的分布式水文模型多源数据融合前处理系统的构建作为例子,说明了该系统广泛的适用性。

2. 分布式水文模型中的多源数据融合前处理系统

分布式水文模型对空间数据的需求量越来越大,如在水文分析基础上提取的 DEM 洼地、流向、集水面积、子流域、河网水系、流域边界等流域地形几何(GIS)信息,雷达测雨、土地利用和土壤质地等遥感(RS)信息,水文与水质站点、水库、湖泊、河流边界、取排水口、城镇等经纬度(GPS)信息,降雨、径流、蒸发等水文气象观测资料,野外流域观测实验数据,人口、工业产值、农业产值、林业产值、牧业产值、水产养殖等社会经济数据,水资源的供、用、耗、排数据,水利工程特征数据和运行管理过程记录数据等,因此经常要进行大量的空间数据融合提取。然而这些数据具有语义差异、时空差异、尺度差异以及空间参考异化等特征,为满足分布式水文模型不同格式和算法的需求,有必要进行多源数据的融合,以实现空间数据与属性数据的匹配,消除不同数据源之间的歧义,从而将多源数据快速有效地应用到分布式水文模型中。

分布式水文模型的前处理系统功能就是要实现多源数据的融合以提供模型输入信息,该系统应具有良好的扩展性、通用型、适应性和可控性等特征,能够方便实现多种复杂时空数据的无缝融合,避免模型重复计算,节省机时和存储空间。系统构建时需明确分布式水文模型结构与算法的需求,以数据融合为切入点,在数据标准化、空间数据与属性数据关联、时间序列的匹配等方面进行规范化和统一化处理。拟解决的关键问题如下: 1) 基于 DEM 的流域地形几何数据与土地利用信息的融合; 2) 土地利用数据与社会经济数据的融合; 3) 土壤质地与流域试验、测量数据、水文物理参数之间的融合; 4) 雷达测雨数据与 DEM 数据融合; 5) 气象水文时间序列与 DEM 数据融合; 6) 水利工程运行管理与 DEM 数据融合。

3. 一种流域四级划分框架体系

分布式水文模型的前处理系统构建的一个核心问题就是确定多源数据信息拓扑结构的编码,基于此本文提出了一种流域-子流域-网格单元-土地利用分类的四级划分框架体系(或流域-子流域-土地利用分类的三级划分框架体系)。该框架体系是在 Strahler 法基础上,为进一步解决多源数据拓扑结构的编码提出的。Strahler 法将流域和研究区域划为很多较小的网格单元和子流域,对流域的网格单元、子流域进行从上游到下游逐级划分和编码,并且每个子流域和网格单元都被赋给唯一的编码,从而确定了河网和子流域的拓扑关系。本文中提出的四级划分框架体系进一步将土地利用遥感信息加入到流域拓扑关系的编码中,从而可以根据土地利用类型的差异将多源数据匹配到各个网格单元及子流域单元中。

具体操作步骤为,首先应用 ArcGIS9.0 软件中的水文分析模块进行河网水系及子流域信息的提取;然后将水文分析后划分的子流域及河网成果转换成*.txt 文件,带入前处理系统中,生成子流域、坡地网格和河网的排序文件,其中子流域的排序主要基于河网的汇流顺序,坡地网格及河网的排序中则主要记录汇入及流出本子流域的网格单元及其他子流域单元信息,最后将土地利用遥感信息带入前处理系统中,生成网格单元或子流域单元上土地利用类型文件。这样几个输入文件就组成了一个流域四级划分框架体系,多源数据信息便可以根据各个网格单元或子流域位置及相应的土地利用情况进行匹配。多源数据融合前处理系统所生成的长江河口的子流域排序、坡地网格排序、河网排序编码方法见表 1~表 3。表 4 为子流域上各种土地利用类型所占的比例信息。

表 1 用于体现了模型计算中子流域的汇流演算顺序;表 2 说明了以流域坡地网格单元为基础的几何信息和数值计算优先演算顺序,如 1 号子流域里面第一个网格由(2,15)网格流入,(11,25)网格流出;(2,15)表示 DEM 数据的第 2 行第 15 列,(11,25)表示 DEM 数据的第 11 行第 25 列。表 3 体现了流域河网网格

Table 1. Order of sub-basin

表 1. 子流域排序

子流域排序	1	2	4	12	14	10	...	...	...	13	16	18	19	21	20
-------	---	---	---	----	----	----	-----	-----	-----	----	----	----	----	----	----

Table 2. Order of slope grid

表 2. 坡地网格排序

子流域序号	网格 1		网格 2		网格 3	
	行号	列号	行号	列号	行号	列号
1	2	15	3	15	4	15
	11	25	10	25	11	26
2	2	10	3	11	4	12
	3	16	4	16	5	15
...	...	...	...	...	...	...
	...	...	...	...	...	...
21	17	26	17	25	18	24
	22	27	21	27	21	28
20	10	18	11	18	12	18
	16	27	15	27	15	28

Table 3. Order of river grid

表 3. 河网排序

计算子流域 序号	河道网格序号		
	行号	列号	所在子流域序号
1	3	17	1
	5	15	4
2	5	13	2
	5	15	4
...	...	...	...
	...	..	...
21	18	23	18
	21	27	21
20	14	18	13
	15	27	20

单元的几何信息和数值计算排序，如第 1 个子流域内的河网由(3,17)网格单元流入，从(5,15)网格单元流到第 4 个子流域。这样就把流域内各个网格单元、河网水系、子流域之间的连接以及数值计算时的优先演算顺序清楚的、有条理的组织在表 2 与表 3 这两个输入文件中。表 1 的子流域排序在两个文件都有涉及，不需要另列文件输入到模型中。对应上述的每个网格单元或子流域单元，又有着不同的土地利用类型的数字化遥感信息(表 4)，这样就实现了分布式水文模型与 GIS、RS 信息的对接，方便了多源数据的融入。

研究区域被划分为很多较小的子流域，子流域又包含很多网格单元，并且每个网格单元、每个子流

Table 4. Land use proportion of sub-basin
表 4. 子流域土地利用类型比例

子流域编号	水域(%)	高植被(%)	低植被(%)	城镇用地(%)	旱田(%)	水田(%)
1	13.85	0	0.72	0.18	1.61	83.64
2	29	0	1.71	0	1.14	68.14
4	57.73	0	2.93	0	1.47	37.87
12	12.39	0	0.39	0.77	1.29	85.16
14	40.88	0	0.75	2.5	1.88	54
10	11	0	0	0	1	88
...	...	...	...	...	...	...
...	...	...	...	...	...	...
13	49.88	0	3.25	0.38	0.25	46.25
16	8.14	0	4.78	6.48	2.8	77.81
18	65.26	0	0.46	0	0.23	34.06
19	22.78	0	0	0	0	77.22
21	100	0	0	0	0	0
20	60.77	0.13	1.55	0	0.77	36.77

域都被赋给唯一的土地利用类别信息。这样的编码方法具有诸多优点，如编码中包含有拓扑信息，方便与多源数据进行对号，编码规则富有规律性，便于计算机处理，适用范围广等。

4. 几何信息与拓扑关系

多源数据融合前处理系统应能充分利用空间几何信息及相应的拓扑关系信息将气象、水文、水资源利用等信息与地形、地貌、土壤质地信息对号入座。几何信息和拓扑关系是地理信息系统中描述地理要素的空间位置和空间关系的不可缺少的基本信息。其中几何信息主要涉及几何目标的坐标位置、方向、角度、距离和面积等信息。而空间关系信息主要涉及几何关系的“相连”、“相邻”、“包含”等信息，它通常用拓扑关系或拓扑结构的方法来分析。拓扑关系是明确定义空间关系的一种数学方法。从拓扑观点出发，关心的是空间的点、线、面之间的联接关系，而不管实际图形的几何形状。水文的空间数据信息包含两类，一类是目标本身的位置信息，即一些空间几何信息，如水文站、雨量站、取水工程等的位置分布等，一个流域上因水库的拦蓄作用，其下游流量将发生变化，不同于自然来水状态，将 GPS 确定的水库位置对应到所处的网格单元上，即可确定空间上哪个位置应考虑水库调度；另一类是地物间的空间关系信息，即一些拓扑关系信息，如通过提取每个单元的水域、高植被、低植被、建筑物、不透水面积等不同土地利用类型信息，从而确定不同下垫面条件应采用的产流计算模式，如降水落在水域上，可以直接转换成径流，不透水面积上下渗损失很小可以忽略不计；在水资源模拟中，通过不同土地利用类型及取排水设施的分布从而进行水资源取用水量的空间分配。本文提出的四级划分框架体系由于实现了分布式水文模型与 GPS、GIS 与 RS 信息的对接，能较好的处理空间几何信息及相应的拓扑关系信息。

5. 应用实例

长江河口上起徐六泾，下至口外 50 号灯标，全长约 181.8 km，位于北纬 31°00′~31°42′，东经 121°59′~122°43′。长江河口地区涉及上海市及江苏省的南通市和苏州市，上海市、南通市和苏州市面积分别为 5790

km²、6881 m²、5104 km²，长江河口地区总面积 17,775 km²。按全国水资源分区的成果，长江河口地区涉及到湖口以下干流的通南及崇明岛诸河区以及太湖水系的武阳区和黄浦江区共 3 个水资源三级区。将上海市、南通市和苏州市这三个地区在长江河口地区的范围分解到对应的水资源三级区作为长江河口地区的分析范围。为了模拟与预测长江河口地区水资源变化，拟在长江河口地区建立一个半分布式水文模型，模型构建过程中将应用该多源数据融合前处理系统。

长江河口地区 5 km × 5 km 的 DEM 图见图 1，应用 ArcGIS9.0 软件中的 Hydrology 水文分析模块对流域的 DEM 信息进行了水文信息提取，从而得到各个子流域的编码信息(见图 2)。

将水文分析结果转换成*.txt 文件，带入前处理系统中，该系统首先寻求河流拓扑关系，再对河流每个网格单元排序；然后根据河流拓扑关系获取子流域之间的拓扑关系，并对每个子流域的网格单元进行优先排序；最后分析各个子流域及网格单元上土地利用类型信息，得到流域 - 子流域 - 网格单元 - 土地利用分类的四级划分框架体系(表 1~表 4)。由于长江河口地区所构建的是半分布式水文模型，因此最终构建了流域 - 子流域 - 土地利用分类的三级划分框架体系，该框架体系由表 4 和表 5 两个输入文件组成。

表 5 第 1 列子流域编码体现了子流域空间关系和先后演算顺序，第 2 列反应该子流域上游有无输入的河段，无填 0，有填 1，第 3 列和第 4 列给出上游有子流域输入的编码，第 5 列和第 6 列为该子流域的

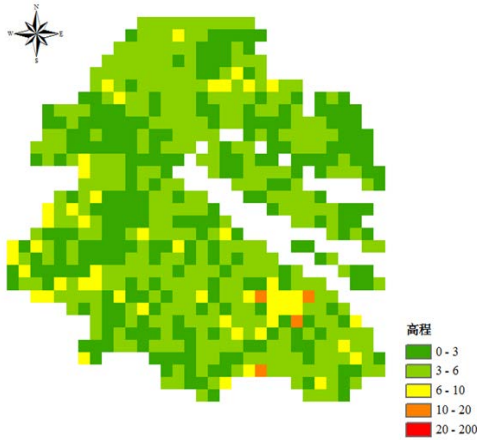


Figure 1. DEM in Yangtze estuary
图 1. 长江河口 DEM 图

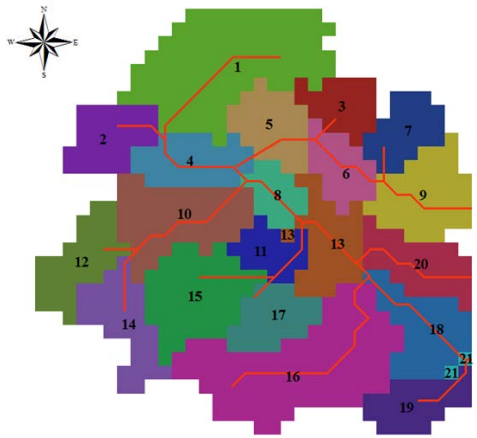


Figure 2. Order of sub-basins in Yangtze estuary
图 2. 长江河口子流域编码

Table 5. Input information of sub-basins
表 5. 子流域输入信息表

子流域编码	有无子流域输入	输入第 1 个子流域	输入第 2 个子流域	河长(km)	子流域面积(km ²)
1	0	0	0	55.36	2375.00
2	0	0	0	15.00	700.00
4	1	1	2	42.07	750.00
12	0	0	0	5.00	775.00
14	0	0	0	15.00	800.00
10	1	12	14	55.36	1200.00
8	1	4	10	28.28	400.00
5	1	4	0	24.14	700.00
3	0	0	0	7.07	525.00
6	1	3	5	31.21	550.00
7	0	0	0	10.00	600.00
9	1	6	7	39.14	1000.00
15	0	0	0	20.00	1200.00
17	0	0	0	5.00	625.00
11	1	15	17	31.21	450.00
13	1	8	11	28.28	800.00
16	0	0	0	91.57	2400.00
18	1	13	16	54.50	875.00
19	0	0	0	15.14	625.00
21	1	19	18	7.07	50.00
20	1	13	0	44.14	775.00

面积及河长；表 4 则给出了各个子流域上土地利用类型的比例信息。这种拓扑关系给半分布式水文模型构建提供了足够的空间信息，便于其他模型输入资料的融入，再加上降水、蒸发、径流等其他输入信息，构成了基于 GIS 和 RS 的半分布式水文模型的主要输入文件。

分布式水文模型由于超存储容量而导致不能编译是一个普遍存在的问题。这里提出的网格或者子流域单元的拓扑关系每次能给出每个计算单元的输入单元和输出单元，将流域化整为零，方便数值求解，这三个单元可以反复使用，极大节省了流域计算的空间存储单元。

6. 小结

随着大量数据信息被应用到分布式水文模型构建中，我们迫切需要一种能体现数据信息拓扑结构的有效编码方式。编码方式的优劣往往直接影响模型的运算速率和计算结果的准确性，因此不恰当的编码方式在某种程度上将阻碍了分布式水文模型的进一步发展和应用。本文中提出了一种流域 - 子流域 - 网格单元 - 土地利用分类的四级划分框架体系，其中包含流域 - 子流域 - 土地利用分类的三级划分框架体系。基于该体系的多源数据融合前处理系统能实现分布式水文模型与 GPS、GIS、RS 以及水资源的供、用、耗、排等信息的对接，较好地处理了空间几何信息及相应的拓扑关系信息，为基于 GIS 和 RS 的分

布式水文模型的构建提供了一种合理的数据处理平台。

基金项目

水利部公益性行业科研专项项目(201201066)。

参考文献 (References)

- [1] CHORMANSKI, J., VOORDE, T., ROECK, T., et al. Improving distributed runoff prediction in urbanized catchments with remote sensing based estimates of impervious surface cover. *Sensors*, 2008, 8(2): 910-932.
- [2] YE, A. Z., DUAN, Q. Y., ZENG, H. Z., et al. A distributed time-variant gain hydrological model based on remote sensing. *Journal of Resources and Ecology*, 2010, 1(3): 222-230.
- [3] KABIR, A., MAHDAVI, M., BAHREMAND, A., et al. Application of a geographical information system (GIS) based hydrological model for flow prediction in Gorganrood river basin, Iran. *African Journal of Agricultural Research*, 2011, 6(1): 35-45.
- [4] 吴俊秀, 郭清. 大凌河流域 MIKE BASIN 水资源模型[J]. *水文*, 2011, 31(1): 71-76.
WU, Jun-xiu, GUO, Qing. MIKE BASIN Model for water resources management of Dalinghe river basin. *Journal of China Hydrology*, 2011, 31(1): 71-76. (in Chinese)
- [5] LI, T., WANG, G. and CHEN, J. A modified binary tree codification of drainage networks to support complex hydrological models. *Computers & Geosciences*, 2010, 36(11): 1427-1435.
- [6] HORTON, R. E. Erosional development of streams and their drainage basins: Hydro-physical approach to quantitative morphology. *Bulletin of the Geological Society of America*, 1945, 56(2): 275-370.
- [7] STRAHLER, A. N. Quantitative analysis of watershed geomorphology. *Transactions of the American Geophysical Union*, 1957, 38(6): 913-920.
- [8] SHREVE, R. L. Statistical law of stream number. *Journal of Geology*, 1966, 74(1): 17-37.
- [9] PFAFSTETTER, O. Classification of hydrographic basics: Coding methodology. Departamento Nacional de Obras de Saneamento, 1989: 25-28. (unpublished manuscript)
- [10] Environmental Modeling Research Laboratory. Environmental modeling research laboratory watershed modeling system WMS 7.0 tutorial. Provo: Brigham Young University, 1998.
- [11] GARBRECHT, J., MARTZ, L. W. TOPAZ (TOpographic PARAMeteri Zation) overview. Oklahoma: Grazinglands Research Laboratory, USDA ARS (US Department of Agriculture, Agriculture Research Service), 1999: 26.
- [12] MAIDMENT, D. R. Arc hydro: GIS for water resources. ESRI Press, California, 2002.
- [13] JASIEWICZ, J., METZ, M. A new GRASS GIS toolkit for Hortonian analysis of drainage networks. *Computers & Geosciences*, 2011, 37(8): 1162-1173.
- [14] 李兰, 朱瑞瑞, 钟名军. 基于 GIS 的 LL-II 分布式降水径流模型的应用[J]. *水电能源科学*, 2004, 22(1): 8-11.
LI Lan, ZHU Rui-rui and ZHONG Ming-jun. Application of LL-II distributed rainfall-runoff model based on GIS. *Water Resources and Power*, 2004, 22(1): 8-11. (in Chinese)