

# A Statistical Downscaling Method Based on Least Squares Support Vector Machines

Yukun Hou, Hua Chen, Xiao Huang, Chongyu Xu

State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan  
Email: [417680591@qq.com](mailto:417680591@qq.com)

Received: Sep. 30<sup>th</sup>, 2013; revised: Nov. 20<sup>th</sup>, 2013; accepted: Nov. 26<sup>th</sup>, 2013

Copyright © 2014 Yukun Hou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for Hans and the owner of the intellectual property Yukun Hou et al. All Copyright © 2014 are guarded by law and by Hans as a guardian.

**Abstract:** The statistical downscaling method has been more and more utilized in the climate change study for its simplicity and flexibility. A statistical downscaling method based on LS-SVM (least squares support vector machines) was developed and compared with SDSM (Statistical Downscaling Model) to test its ability in downscaling precipitation and temperature in Xiangjiang Basin. The results showed that the method based on LS-SVM has the similar performance with the SDSM method in simulating precipitation, while it was superior to SDSM in simulating temperature. The proposed method still needs to be applied to more regions to make it more suitable for studying the impact on water resources under climate change.

**Keywords:** Statistical Downscaling; LS-SVM; SDSM

## 基于 LS-SVM 的改进统计降尺度方法

侯雨坤, 陈 华, 黄 道, 许崇育

武汉大学水资源与水电工程科学国家重点实验室, 武汉  
Email: [417680591@qq.com](mailto:417680591@qq.com)

收稿日期: 2013 年 9 月 30 日; 修回日期: 2013 年 11 月 20 日; 录用日期: 2013 年 11 月 26 日

**摘 要:** 统计降尺度方法作为一种计算量小、使用灵活的降尺度模型, 被越来越多应用到气候变化研究当中。本文以湘江流域为例, 开发了一种基于 LS-SVM 回归的改进统计降尺度算法, 并与经典统计降尺度模型 SDSM (Statistical Downscaling Model) 进行比较。结果表明, 在湘江流域, 无论是降水模拟和温度模拟, 基于 LS-SVM 回归算法的改进统计降尺度方法都能达到 SDSM 的效果, 而温度的模拟, LS-SVM 回归降尺度方法模拟结果更好。为了使得这种方法能更适合气候变化对水资源的影响研究, 还需要在更多的区域进行应用证明。

**关键词:** 统计降尺度; LS-SVM; SDSM

### 1. 引言

降尺度方法就是将大尺度的预报因子经过一系

作者简介: 侯雨坤(1991-), 男(汉), 江苏南京, 武汉大学水文学及水资源系, 硕士研究生, 主要从事气候变化对水循环影响。

列处理, 降解到小尺度区域上, 使之能够与水文模型输入的尺度匹配。由于发展具有更高分辨率的 GCM 在计算量上过于庞大, 对于尺度的降解作用十分有限, 目前主要采用统计降尺度与动力降尺度两种方法。动

力降尺度方法通过内嵌全球气候模式, 具有较为明确的物理意义<sup>[1]</sup>, 但是相比较而言, 统计降尺度方法的计算量较小, 并且模型算法的构建相对简单, 有众多形式较为灵活的模型可供选择<sup>[2]</sup>。

目前, 统计降尺度方法越来越多地作为有效的工具应用到研究中<sup>[3-7]</sup>。Guo 等应用 ASD 降尺度方法黄河流域降雨变化进行预测, 得出 2010~2099 年 A2、B2 不同情景下的年降雨量趋势变化; Chen 等采用改进的平滑支持向量机, 利用 CGCM2 与 HadCM3 两种模式在 A2 情景下的输出, 预测汉江流域未来的降雨趋势变化, 为南水北调工程提供了有效的科技支持; 褚健婷等将 SDSM 模型应用到海河流域进行未来降雨、蒸发的预测, 取得了较好的效果; Chen 等通过将多种 GCM 模式与水文模型耦合进行预测的结果对比, 得出 NSC 与 RSR 是评价 GCMs 输出因子与水文模型耦合效果优劣的关键统计指标的结论; Jeong 等采用多元回归、稳健回归、岭回归及人工神经网络方法对于加拿大不同省份进行气温的预测, 通过比较不同方法取得的效果, 认为多元回归在加拿大大部分地区最为适用, 而稳健回归、岭回归的效果相对较差。

统计降尺度由于所需数据较动力降尺度少很多, 找出气象因子与气候数据之间的合适的回归关系较为灵活, 所以具有巨大的发展空间。我们可以通过模型改进等多种手段对统计降尺度算法进行优化, 同时对于不同地区及流域建立不同参数的统计降尺度模型, 从而达到用少量气象数据获得较好降尺度效果的目标, 其研究意义十分明显。而在众多统计降尺度模型中, Chen 等的研究表明, SVM 回归比一般回归模型拥有更强的学习能力, 在长序列气象数据情况下基于 SVM 回归的模型降尺度效果较一般回归更好, 具有进一步研究的潜力。因此, 在此基础上, 本文主要讨论以 SVM 算法作为模型核心的统计降尺度的原理方法及应用改进。

## 2. 基于 SVM 回归的统计降尺度算法

### 2.1. 统计降尺度方法介绍

统计降尺度方法利用长序列观测资料, 构建大尺度 GCM 输出因子与区域气候要素之间的线性或非线性的统计关系, 通过验证所构建统计关系的合理性, 将这种统计关系用于 GCM 的尺度降解。较为复杂的统

计降尺度一般分为三类: 回归分析、天气分型技术及天气发生器<sup>[8]</sup>。尽管各类方法采用不同的步骤及算法, 但都是建立在区域气候是大尺度大气状态的函数这一重要假设的基础上的。大尺度气候变量(GCM 预报因子)与区域气候变量(预报量)之间的统计关系可以用随机或确定性的方法表达, 总的来说其关系如下所示:

$$Y = F(X) \quad (1)$$

式中,  $Y$  代表需要预测的区域气候变量,  $X$  为一系列 GCM 大尺度输出气候因子,  $F$  为建立两者的函数表达式。

### 2.2. 基于 LS-SVM 回归的降尺度算法

支持向量机以其模型简单、具有好的推广能力和全局最优解等特点, 被应用月时间序列建模与预测、智能控制等方面, 已经在统计降尺度上得到了应用。然而, 由于其基于二次规划的优化计算方法不适合海量数据训练, 采用最小二乘支持向量机进行函数估计, 改进其优化计算方法以提高其实用性。

本文采用 LS-SVM 回归算法为基础对流域进行降尺度分析。LS-SVM 的基本思想为: 将寻求最优超平面的算法定义为求最优化解的规划问题, 利用 Mercer 核函数进行数据处理, 将样本空间映射至高维乃至无穷维的特征空间, 从而达到用线性学习机处理非线性数据的能力, 以达到更好的拟合效果。与标准 SVM 相比, LS-SVM 回归只需要确定核函数的形状参数和惩罚系数, 而不需要不敏感损失函数的值, 这既简化了计算, 又便于实际使用。具体如下:

#### (1) 降雨状态序列的生成 0

##### ■ 率定期: 降雨状态模型率定

对于率定期的降雨数据, 规定一个降雨状态的判断阈值  $threshold$ , 认为雨量大于该阈值的状态为湿天, 小于该阈值的状态为干天, 如下所示。

$$w_t = \begin{cases} 1 & R_t > threshold \\ 0 & R_t \leq threshold \end{cases} \quad (2)$$

式中,  $t$  代表天数;  $R_t$  为第  $t$  天的实测降雨量;  $w_t$  为第  $t$  天的降雨状态, 为 0-1 状态, 0 代表无雨, 1 代表有雨。

采用 LS-SVM 回归, 构建降雨状态与大尺度气象

因子间的统计关系:

$$w_t = F_{1t}(\hat{u}_t^1, \hat{u}_t^2, \dots, \hat{u}_t^j) + \varepsilon \quad (3)$$

式中,  $\varepsilon$  为服从标准正态分布的随机误差项, 用以衡量由于降雨状态阈值设置而带来的降雨状态划分误差;  $\hat{u}_t^1, \hat{u}_t^2, \dots, \hat{u}_t^j$  为挑选出来的  $j$  个大尺度气象因子;  $F_{1t}$  为 LS-SVM 回归得到的预报因子与降雨状态之间的多元非线性回归关系。最终率定期通过实测降雨资料与对应时段的大尺度气象因子, 获得降雨状态的 LS-SVM 回归关系。

#### ■模拟期: 降雨状态模拟

利用模拟期大尺度因子与率定期获得的回归关系, 结合随机天气发生器原理, 进行模拟期降雨状态的模拟, 如下:

$$W_t^{sim} = \begin{cases} 1 & rand \leq F_{1t}(\hat{u}_t^1, \hat{u}_t^2, \dots, \hat{u}_t^j) \\ 0 & rand > F_{1t}(\hat{u}_t^1, \hat{u}_t^2, \dots, \hat{u}_t^j) \end{cases} \quad (4)$$

式中,  $W_t^{sim}$  为模拟期  $t$  天回归模拟得到的降雨状态;  $rand$  为服从[0,1]均匀分布的随机项;  $\hat{u}_t^1, \hat{u}_t^2, \dots, \hat{u}_t^j$  为挑选出来的  $j$  个大尺度气象因子。

#### (2) 雨量的模拟

##### ■率定期: 降雨雨量模型率定

首先, 依据经验习惯, 对实测降雨数据进行转换, 如下:

$$R_t^* = \ln(R_t) \quad R_t > \text{threshold} \quad (5)$$

式中,  $R_t$  为  $t$  天的实测降雨量;  $R_t^*$  为实测雨量的转化形式。

然后采用 LS-SVM 进行回归, 构建降雨雨量与大尺度气象因子间的统计关系:

$$R_t^* = F_{2t}(\hat{u}_t^1, \hat{u}_t^2, \dots, \hat{u}_t^j) + \varepsilon \quad (6)$$

式中,  $\varepsilon$  为服从标准正态分布的随机误差项, 用以衡量在数据转换过程中降雨状态阈值设置而带来的误差;  $F_{2t}$  为 LS-SVM 回归得到的预报因子与降雨雨量之间的多元非线性回归关系。

##### ■模拟期: 降雨序列的模拟

利用模拟期大尺度因子与率定期获得的回归关系, 进行模拟期降雨雨量的模拟, 最终得到完整的降雨时间序列, 如下:

$$R_t^* = F_{2t}(\hat{u}_t^1, \hat{u}_t^2, \dots, \hat{u}_t^j) + e \quad w_t = 1 \quad (7)$$

式中,  $R_t^*$  为模拟期第  $t$  天回归模拟得到的降雨雨量转换量;  $\hat{u}_t^1, \hat{u}_t^2, \dots, \hat{u}_t^j$  为挑选出来的  $j$  个大尺度气象因子;  $e$  为一个假设服从高斯分布的误差随机项, 参照 ASD 算法<sup>[9]</sup>, 其计算如下:

$$e = \sqrt{V_{obs} - V_{calibration}} \cdot \varepsilon + b \quad (8)$$

式中,  $V_{obs}$  为率定期实测  $R_t^*$  的方差;  $V_{calibration}$  为率定期通过回归得到的  $R_t^*$  确定性部分方差;  $\varepsilon$  为服从标准正态分布的随机数;  $b$  值在使用 NCEP 数据进行检验时, 一般取值为 0, 当采用 GCM 数据生成未来情景时,  $b$  值计算如下:

$$b = M_{obs} - M_{calibration} \quad (9)$$

式中,  $M_{obs}$  为率定期实测  $R_t^*$  的均值;  $M_{calibration}$  为率定期通过回归得到的  $R_t^*$  确定性部分均值。

最后, 将  $R_t^*$  值还原, 得到完整的时间序列, 如下:

$$R_t = \begin{cases} 0 & w_t = 0 \\ \exp(R_t^*) & w_t = 1 \end{cases} \quad (10)$$

对于温度等连续变量, 不用对其进行状态的模拟, 可直接仿照(2)步骤进行数量的模拟。

## 3. 湘江流域降尺度算法应用

### 3.1. 流域概况及研究数据

湘江流域位于北纬 24°~29°, 东经 110°30'~114°之间, 地处长江之南, 南岭山地之北, 属于副热带及暖温带地区<sup>[10]</sup>。本文选用湘江流域内 11 个气象站点和 5 个水文站点。气象站点的实测数据包括日最高气温、日最低气温、平均气温、日相对湿度、日平均风速以及逐日日照时间; 水文站点的实测数据为日径流量。降雨数据主要来源与湘江流域 261 个雨量站, 大部分站点具备 1970~2005 年完整的降雨数据, 且空间分布比较均匀。通过 SDSM、SVM 回归降尺度方法, 采用 NCEP 大尺度气候因子, 以 1975~1998 为模型率定期, 1999~2004 为模型检验期, 对湘江流域降水及气温两个对于水文模型最为重要的气象要素进行降尺度模拟和预测, 比较不同算法统计降尺度的模拟结果优劣。根据计算相关系数及传统经验, 选用 6 个 NCEP 大尺度因子, 如表 1 所示。

### 3.2. 降雨模拟

尽管统计降尺度技术可以将大尺度气候因子降

解到站点降雨尺度, 但是由于模型最终会以面雨量进行水文计算, 从后续耦合的便捷性和精度方面考虑, 论文选择将大尺度因子降解到子流域面降雨尺度。本文具体计算提取了湘江流域的 56 个子流域 1975~2004 年通过观测值处理得到的逐日面雨量, 子流域对应网格的 NCEP 因子, 应用 SVM 两阶段回归(考虑降雨状态与雨量两个部分)及 SDSM 进行统计降尺度。

在比较中, 本文提取了划分出的 56 个子流域中面积大于 1000 平方公里的 37 个子流域(占流域总面积 93%)计算其检验期的统计特征值, 选用的统计特征值及其意义如下表 2 所示。

为突出模拟的效果比较, 对不同方法下子流域 6 检验期各月的日平均雨量进行计算, 如图 1。

对 37 个子流域如表 2 中的统计特征值进行平均, 并且统计在各个子流域两种方法各项统计指标的优胜次数, 得到综合反映整个流域模拟效果的表 3。

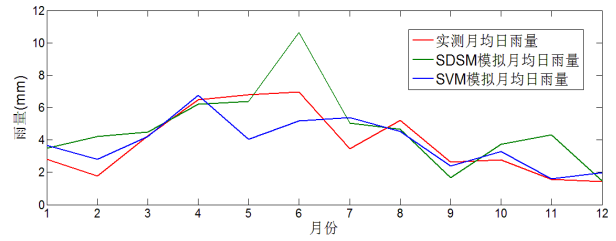
从表 3 可以看出, 除了 95%的分位数, 其它方面 SVM 改进的统计降尺度算法均好于 SDSM。特别值得关注的是, 改进 SVM 方法在均值、方差以及最大 5

**Table 1. Selection of predictors and their physical meaning**  
**表 1. 气象因子的选取及其意义**

气候因子简写	意义
msslpgl	平均海平面气压
p500gl	500 hPa 位势
s500gl	500 hPa 比湿
p8_fgl	850 hPa 风速
p8_zgl	850 hPa 涡量
tempgl	地面温度(2 m)

**Table 2. Statistical characteristics and their meaning**  
**表 2. 统计特征值名称及其意义**

名称	意义
Mean(mm/d)	多年日均雨量
Standard Deviation (mm/d)	多年日标准差
Percentile (95)	95%分位数
5-day maximum rainfall (mm)	连续 5 日最大降雨量
Percentage wet	降雨天数的比重
Average wet spell length (d)	多年连续降雨序列平均时长
PBIAS (%)	相对误差系数



**Figure 1. Comparison of monthly simulated precipitation amounts in June**

**图 1. 子流域 6 月雨量模拟比较**

**Table 3. Mean statistical characteristics of precipitation in Xiang-jiang basin**  
**表 3. 湘江流域降雨平均统计特征值**

Indicators	Seasons	Average Value			Number for excellence	
		Observe	SVM	SDSM	SVM	SDSM
Mean (mm/d)	DJF	1.99	2.50	2.53	21	16
	MAM	6.07	5.75	6.45	20	17
	JJA	5.92	5.04	6.73	15	22
	SON	2.47	2.32	2.76	21	16
	Annual	4.19	3.96	4.66	22	15
Standard Deviation (mm/d)	DJF	4.6	6.6	5.0	4	33
	MAM	12.0	13.2	10.3	22	15
	JJA	13.8	13.5	12.8	20	17
	SON	7.0	6.9	6.5	12	25
	Annual	10.7	11.6	9.6	26	11
Percentile (95)	DJF	10.4	12.1	13.8	28	9
	MAM	30.9	25.5	28.5	13	24
	JJA	33.7	24.3	35.7	9	28
	SON	15.3	12.0	18.1	17	20
	Annual	23.6	18.7	25.5	9	28
5-day maximum rainfall (mm)	DJF	49.7	66.8	51.0	8	29
	MAM	127.1	132.2	103.6	24	13
	JJA	151.0	138.1	131.5	20	17
	SON	84.1	67.1	66.0	17	20
	Annual	173.9	182.7	138.8	28	9
Percentage wet	DJF	0.488	0.496	0.479	×21	16
	MAM	0.588	0.624	0.584	17	20
	JJA	0.485	0.487	0.467	20	17
	SON	0.405	0.402	0.364	12	25
	Annual	0.487	0.504	0.470	17	20
Average wet spell length	DJF	2.55	2.05	2.00	26	11
	MAM	3.27	2.77	2.72	24	13
	JJA	2.67	2.16	2.22	21	16
	SON	2.32	1.68	1.74	23	14
	Annual	2.78	2.30	2.21	25	12
PBIAS (%)	monthly		5	12	25	12
In Total					587	560

日降雨量的模拟相对 SDSM 效果提升很多。可以认为, 采用 SVM 改进的算法对于降水的降尺度计算, 相比 SDSM 更适用于气候变化下的水文模型耦合。

### 3.3. 温度模拟

区别于降雨降尺度, 温度降尺度将 NCEP 大尺度因子降解到站点尺度的最高温、最低温要素上。考虑到温度作为连续无需考虑状态, 只需直接与大尺度气象因子进行多元线性回归回归即可。由于 SDSM 与 SVM 在温度降尺度方面效果皆较为良好, 论文直接对 SVM 回归的统计降尺度效果进行评判, 而并不与 SDSM 进行比较。对于湘江流域 11 个气象站点检验期的日温度 Nash-Sutcliffe 效率系数如表 4 所示。

以邵阳站检验期的日温度数据为例, 计算各月的多年日均温度, 如图 2 所示。

Table 4. Coefficient of efficiency of daily temperature for all meteorological stations  
表 4. 检验期各站日温度效率系数

站点名称	站点编码	效率系数 Ens (%)	
		最高温度	最低温度
马坡岭	57679	76.5	90.9
长沙	57687	77.1	91.2
邵阳	57766	86.5	92.5
双峰	57774	85.6	91.2
南岳	57776	84.2	90.7
株洲	57780	77	91.3
永州	57866	78.3	90.1
衡阳	57872	79.4	91
常宁	57874	85	92.2
道县	57965	80.1	90.8
郴州	57972	76.5	90.8
平均		80.6	91.2

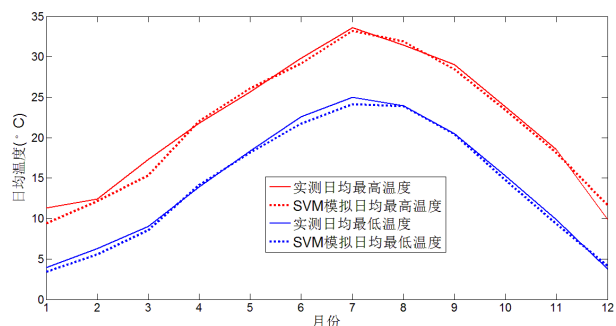


Figure 2. Comparison of observed and simulated temperature at Shaoyang station  
图 2. 邵阳站检验期实测及模拟温度比较

可以看出, 由于最高温度相对最低温度的变幅较大, SVM 改进的统计降尺度算法对于最高温度的模拟略差于对最低温度的模拟。但是其模拟的效果总体令人满意。

### 4. 结论

综上所述可以看出, 无论是基于 LS-SVM 回归的降尺度方法, 还是传统的 SDSM 方法, 对于降雨这种状态不连续序列的模拟依旧存在比较大的改进空间, 无法获得很好的效率系数。但是相比于经典的 SDSM 算法而言, LS-SVM 回归降尺度方法能实现对降雨的模拟, 能适用于气候变化下的水文灾害分析。而对于温度这种连续序列, SVM 可以利用其非线性回归的特点, 很好的对温度进行模拟。总的来说, 建立在 SDSM 模型原理基础上, 采用 SVM 算法进行改进的统计降尺度模型, 其模拟结果不亚于传统的 SDSM 模型, 能够进一步应用于气候变化对水资源的影响研究。

### 致谢

感谢国家自然科学基金项目的支持, 感谢第十一届中国水论坛推荐!

### 基金项目

国家自然科学基金(51279138)。

### 参考文献 (References)

- [1] 赵芳芳, 徐宗学. 统计降尺度方法和 Delta 方法建立黄河源区气候情景的比较分析[J]. 气象学报, 2007, 65(4): 653-662. ZHAO Fangfang, XU Zongxue. Comparative analysis on downscaled climate scenarios for headwater Catchment of Yellow River using SDSM and delta methods. Acta Meteorologica Sinica, 2007, 65(4): 653-662.
- [2] 刘永和, 郭维栋, 冯锦明, 张可欣. 气象资料的统计降尺度方法综述[J]. 地球科学进展, 2011, 26(8): 837-847. LIU Yonghe, GUO Weidong, FENG Jinming and ZHANG Kexin. A summary of methods for statistical downscaling of meteorological data. Advances in Earth Science, 2011, 26(8): 837-847.
- [3] GUO, J., CHEN, H., XU, C. Y., et al. Prediction of variability of precipitation in the Yangtze River Basin under the climate change conditions based on automated statistical downscaling. Stochastic Environmental Research and Risk Assessment, 2012, 26(2): 157-176.
- [4] CHEN, H., GUO, J., et al. Downscaling GCMs using the Smooth Support Vector Machine method to predict daily precipitation in the Hanjiang Basin. Advances in Atmospheric Sciences, 2010, 27(2): 274-284.
- [5] 褚健婷, 夏军, 许崇育. SDSM 模型在海河流域统计降尺度研究中的适用性分析[J]. 资源科学, 2008, 30(12): 1825-1832. CHU Jianting, XIA Jun and XU Chongyu. Suitability analysis of

- SDSM Model in the Haihe River Basin. *Resources Science*, 2008, 30(12): 1825-1832.
- [6] CHEN, H., Xu, C. Y. and GUO, S. Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff. *Journal of Hydrology*, 2012, 434: 36-45.
- [7] JEONG, D. I., ST-HILAIRE, A., OUARDA, T., et al. Comparison of transfer functions in statistical downscaling models for daily temperature and precipitation over Canada. *Stochastic Environmental Research and Risk Assessment*, 2012, 26(5): 633-653.
- [8] 范丽军, 符淙斌, 陈德亮. 统计降尺度法对未来区域气候变化情景预估的研究进展[J]. *地球科学进展*, 2005, 20(3): 320-329.
- FAN Lijun, FU Congbin and CHEN Deliang. Review on creating future climate change scenarios by statistical downscaling techniques. *Advances in Earth Science*, 2005, 20(3): 320-329.
- [9] HESSAMI, M., GACHON, P., OUARDA, T. B. M. J., et al. Automated regression-based statistical downscaling tool. *Environmental Modelling & Software*, 2008, 23(6): 813-834.
- [10] 杨纫章. 湘江流域水文地理[J]. *地理学报*, 1957, 23(2): 161-182.
- YANG Renzhang. Hydrography of the Hsiang-Kiang Basin, Hunan Province. *Acta Geographica Sinica*, 1957, 23(2): 161-182.