WeeklyNote

2019.11.01

張慕琪

Statistical downscaling of precipitation using machine learning techniques

D.A. Sachindra[a,*], K. Ahmed[b], Md. Mamunur Rashid[c], S. Shahid[d], B.J.C. Perera[a]

[a] Institute for Sustainability and Innovation, College of Engineering and Science, Victoria University, P.O. Box 14428, Melbourne, Victoria 8001, Australia
[b] Faculty of Water Resources Management, Water and Marine Sciences, Lasbela University of Agriculture, Uthal, Balochistan, Pakistan
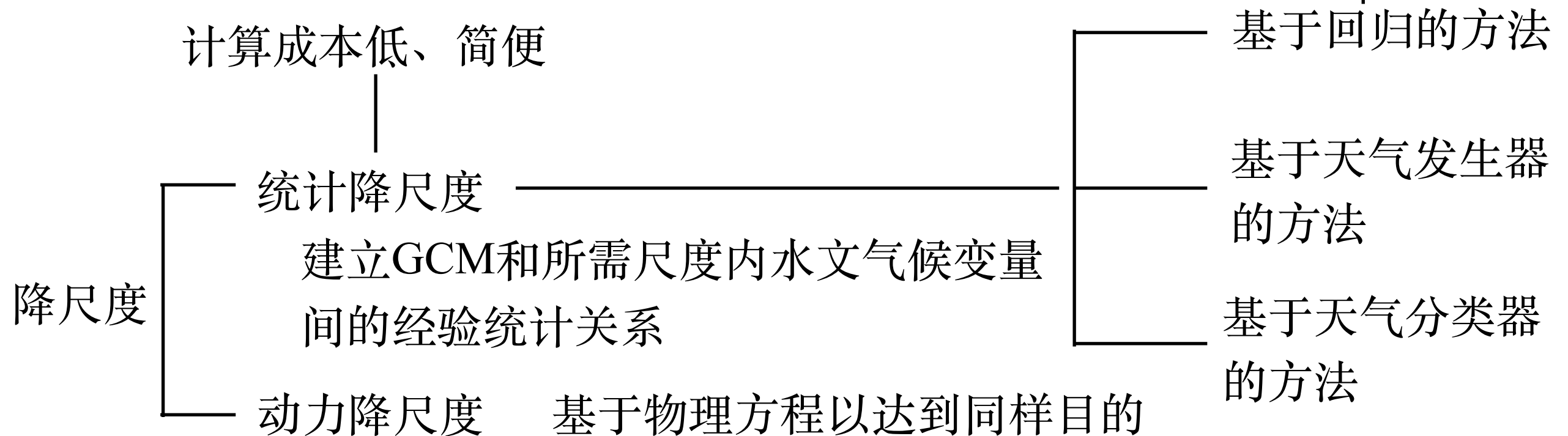[c] Civil, Environmental, and Construction Engineering Department, University of Central Florida, Orlando, Florida 32816-2450, USA
[d] Faculty of Civil Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia

Introduction

机器学习方法 —— MLR、GLMs、
ANNs、SVM、RVM、GP、GEP……

计算成本低、简便

统计降尺度 ———————— 基于回归的方法

建立GCM和所需尺度内水文气候变量
间的经验统计关系

基于天气发生器
的方法

降尺度

基于天气分类器
的方法

动力降尺度　基于物理方程以达到同样目的

相较于MLR，基于GP的降尺度模型可以更好的模拟日最高温和日最低温；
相较于ANN和MLR，基于SVM的降尺度模型可以更好的模拟月最高温和最低温；
相对于其他所有方法，基于SVM的降尺度模型对降水的月平均模拟最好；
——根据上述研究，在众多传统的统计回归方法中，依据机器学习方法建立的降尺度模型比其他方法更加有优势。

Introduction

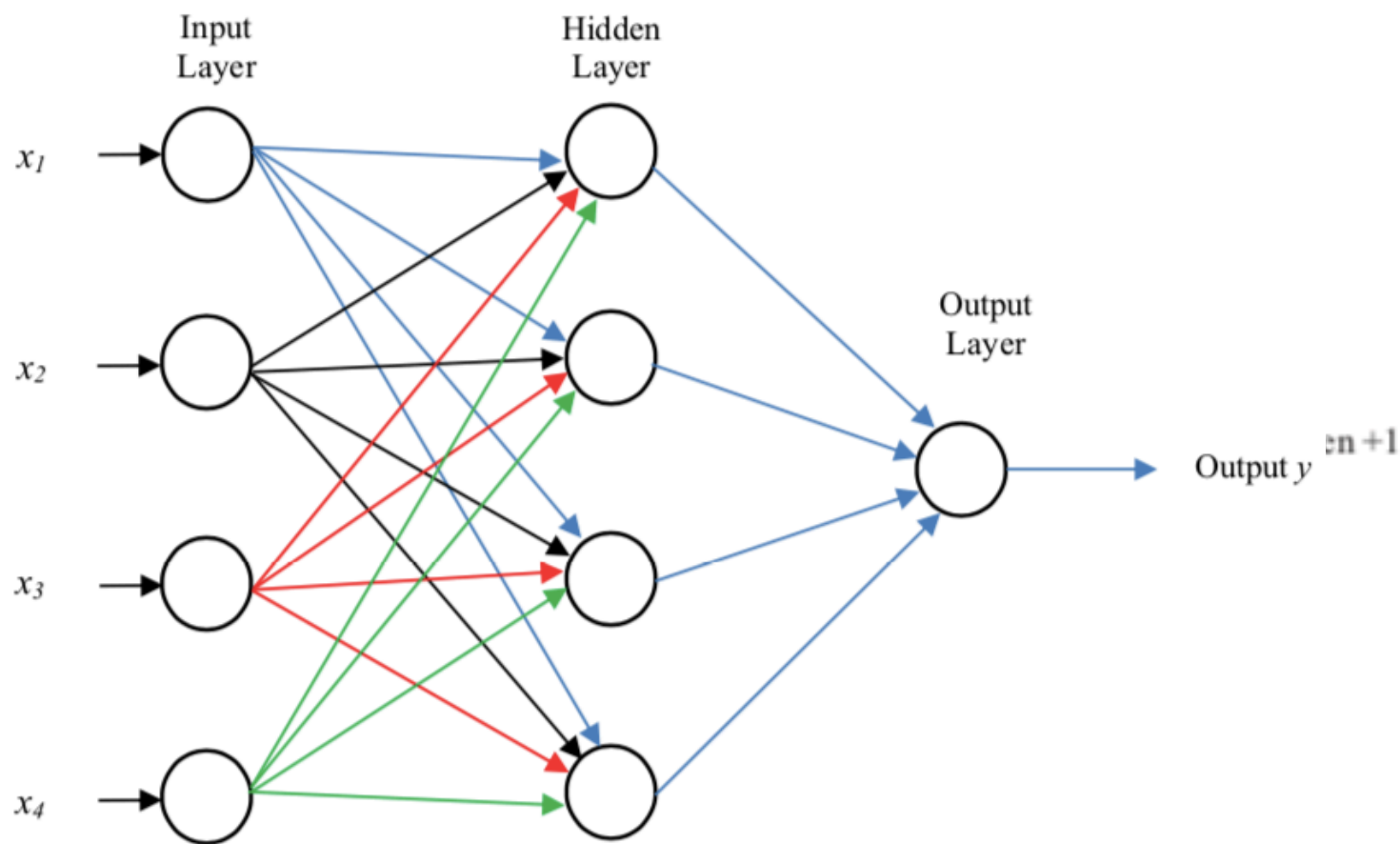1. GP（Genetic programming）

2. Artificial neural networks（ANNs）

Fig. 4. A typical artificial neural network.

Fig. 3. Genetic programming algorithm.

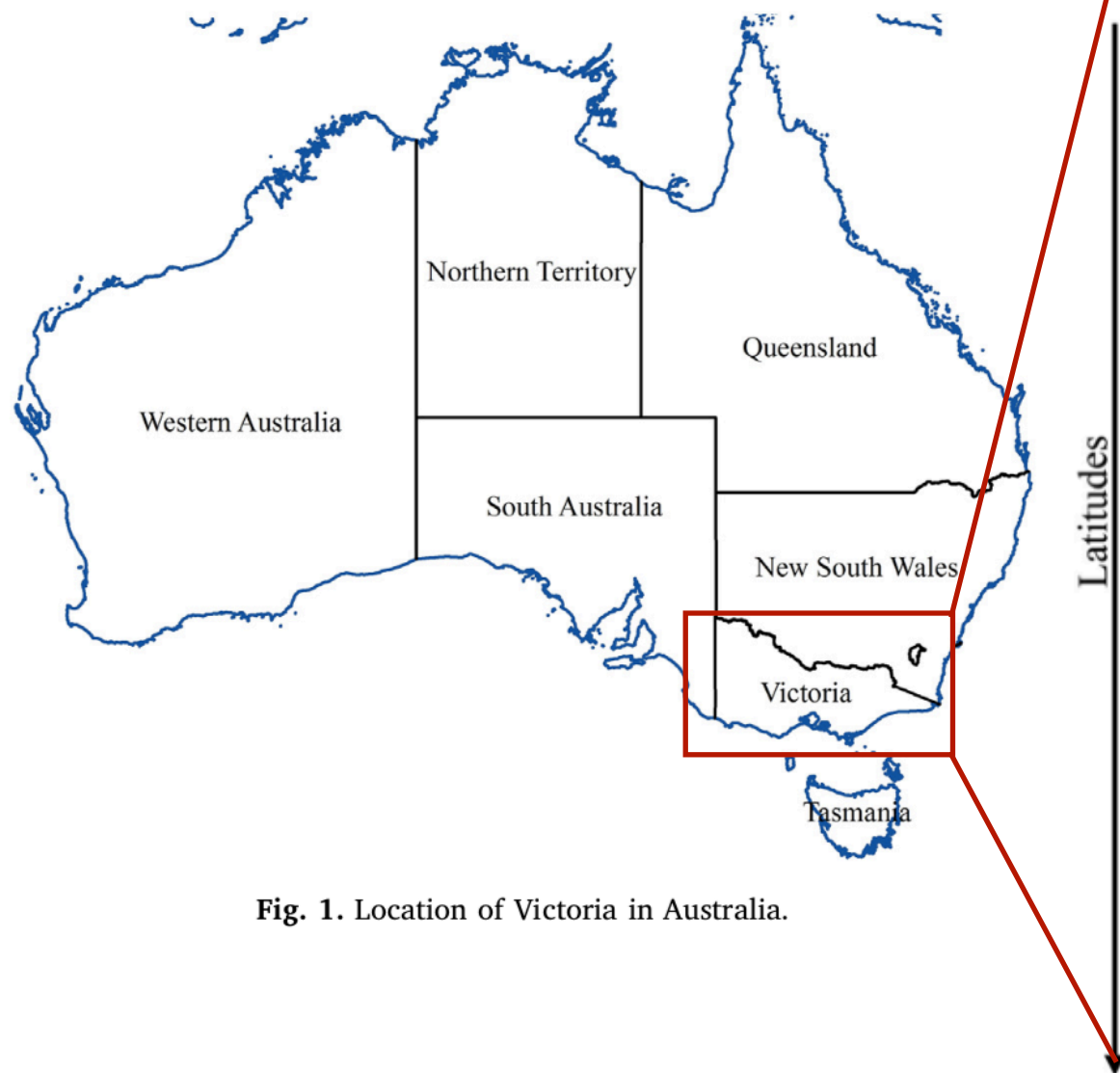3. Support vector machine（SVM）

4. Relevance vector machine（RVM）

Method

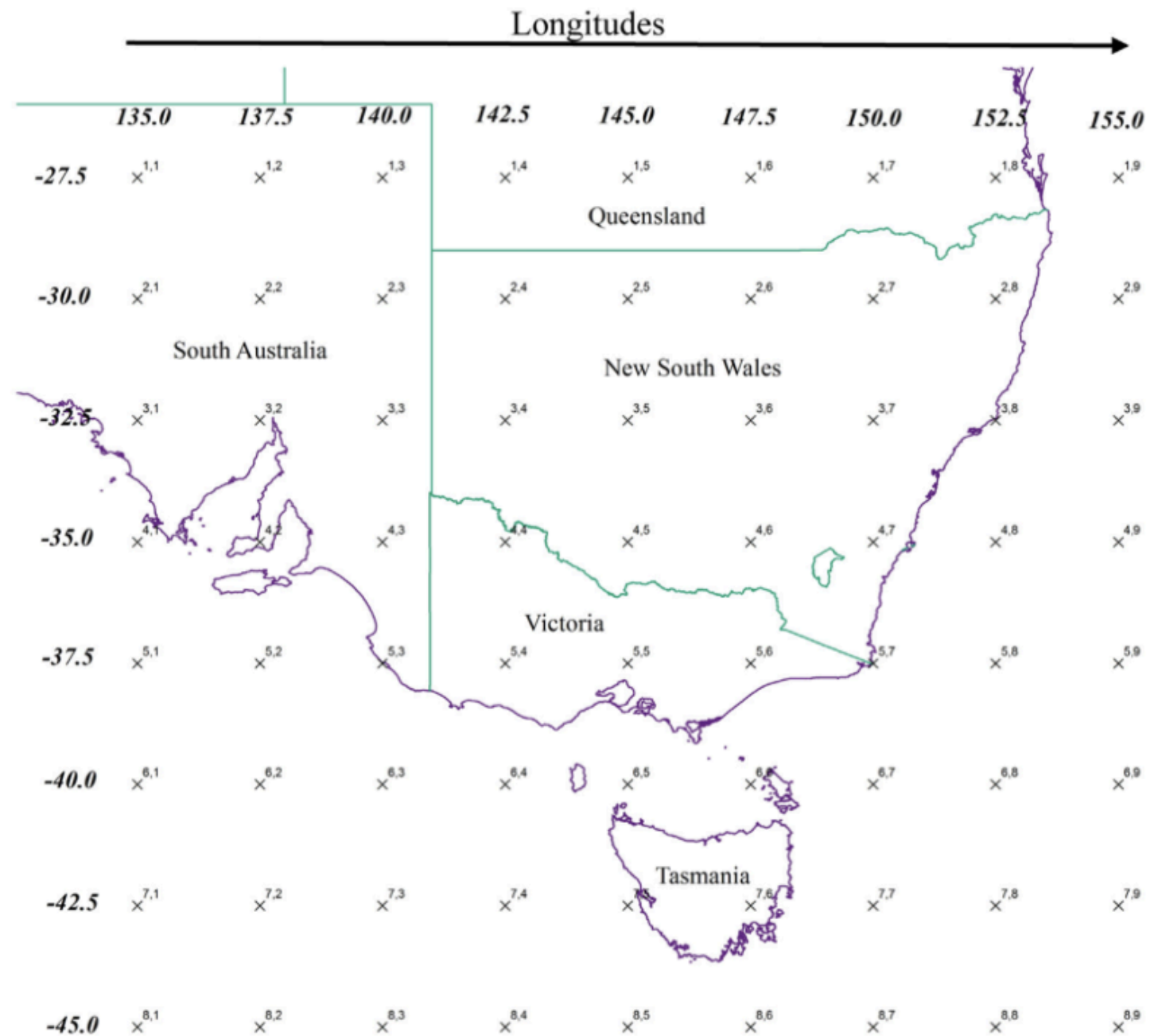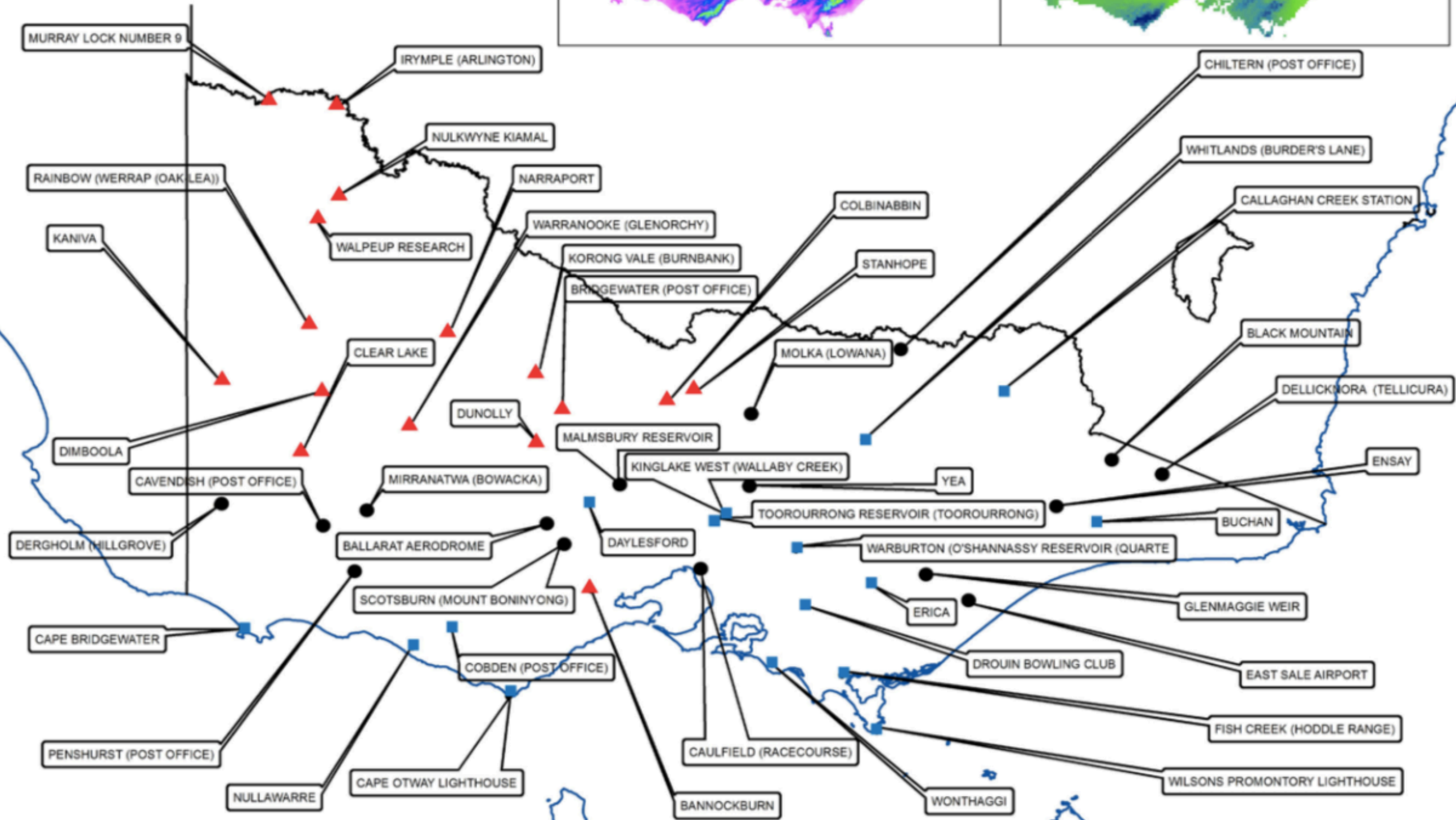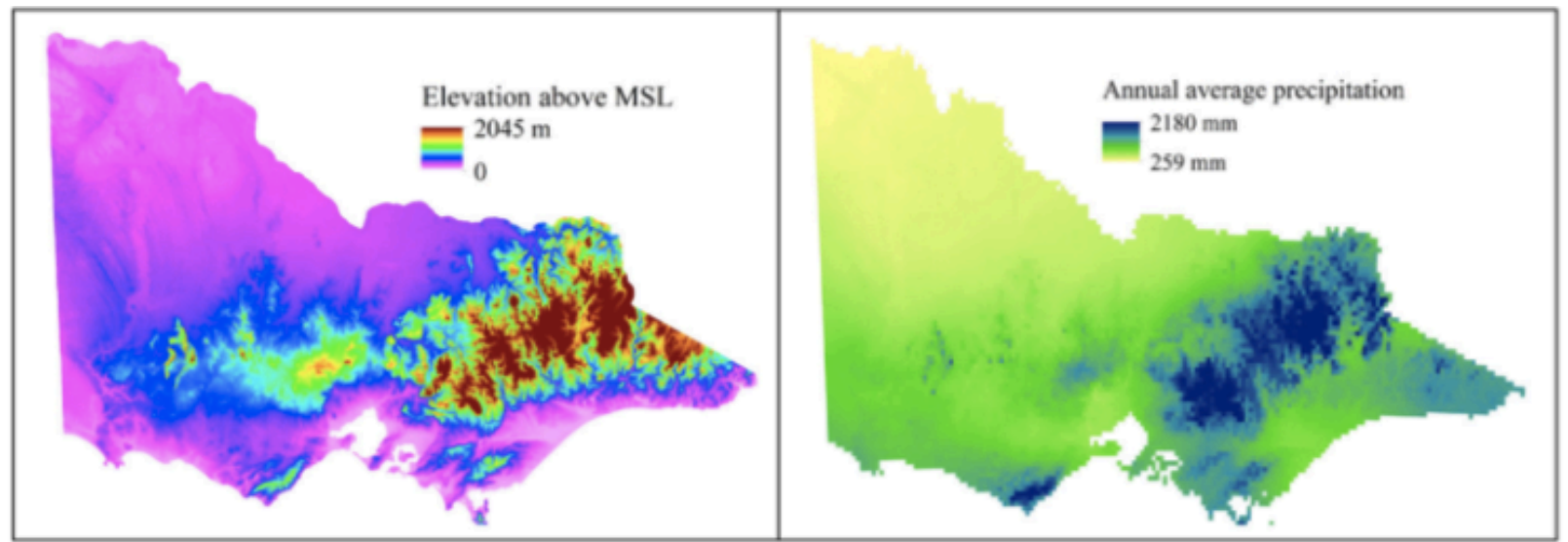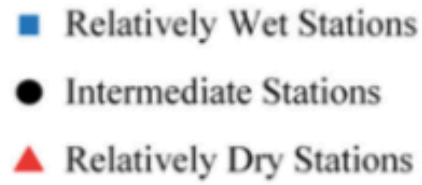Fig. 1. Location of Victoria in Australia.

Fig. 5. Atmospheric domain used in this study.

NCEP/NCAR reanalysis data set；

48 precipitation observation stations located across Victoria (237,000km$_2$), Australia were selected；

时间：1950-1991(calibration)、1992-2014(validation)

Method

Results

Table 4
Percentage of selection of a kernel as the best for a given climate regime.

Machine learning technique	Climate regime	Kernel								
		Hyperbolic tangent	Polynomial	RBF	Laplacian	Bessel	ANOVA	Spline	String	Linear
SVM	Wet	0.0	34.4	13.0	25.0	18.8	8.9	0.0	0.0	0.0
	Intermediate	0.0	42.2	13.0	19.3	16.7	8.9	0.0	0.0	0.0
	Dry	0.0	39.1	10.4	25.0	14.6	10.9	0.0	0.0	0.0
RVM	Wet	1.0	33.9	13.0	25.0	18.2	8.9	0.0	0.0	0.0
	Intermediate	1.6	39.6	13.0	19.8	16.1	9.9	0.0	0.0	0.0
	Dry	0.5	38.0	10.4	25.5	15.1	9.9	0.5	0.0	0.0

无论是那种气候区域，SVM和RVM使用Polynomial kernel，Laplacian，Bessel，RBF和ANOVA都可以很好的模拟降水变量；同时，Spline，String，Linear和Hyperbolic tangent kernals无法得到降尺度模型所需的信息。

Results
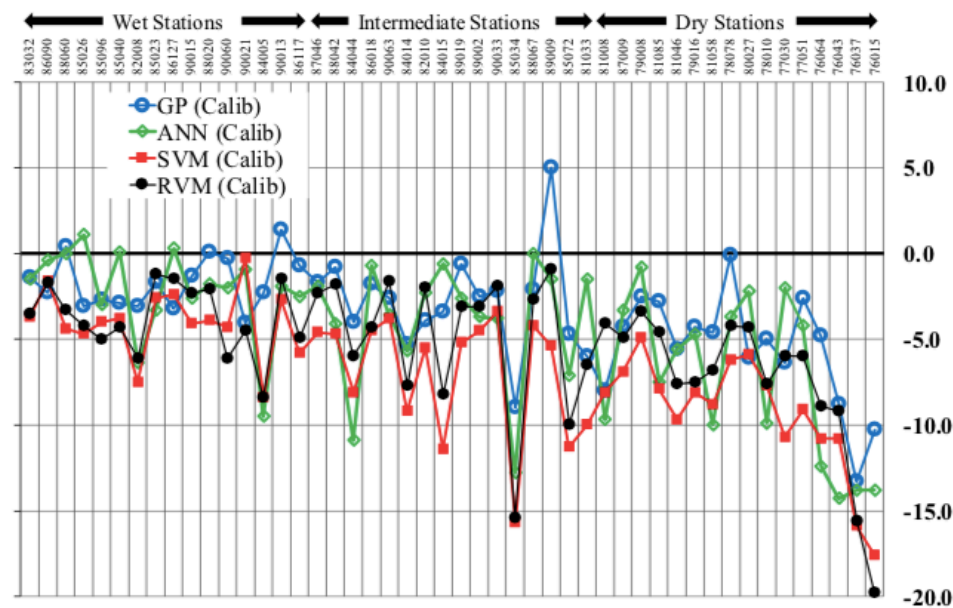
(a). Percentage of bias in average in Calibration
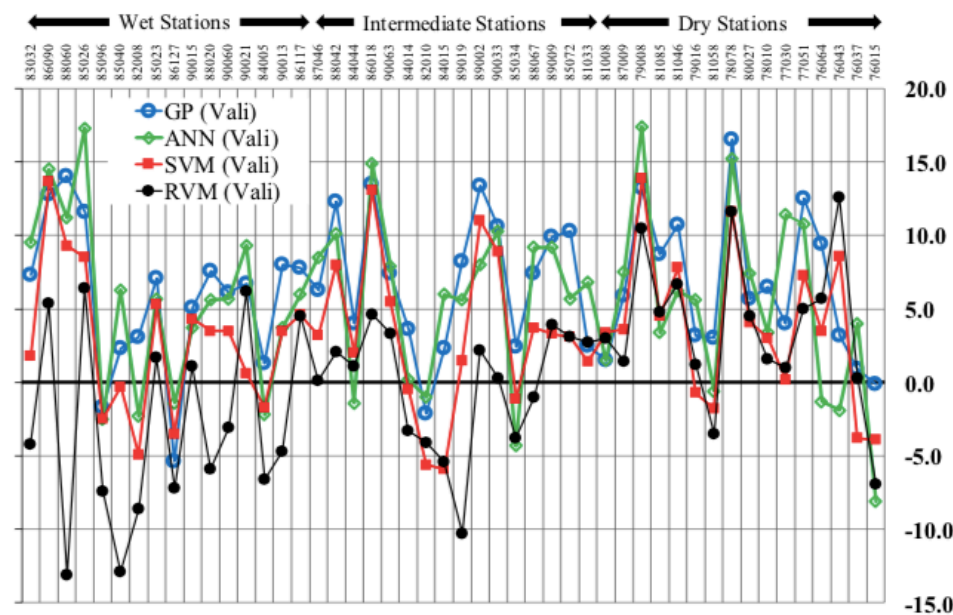
(b). Percentage of bias in average in Validation

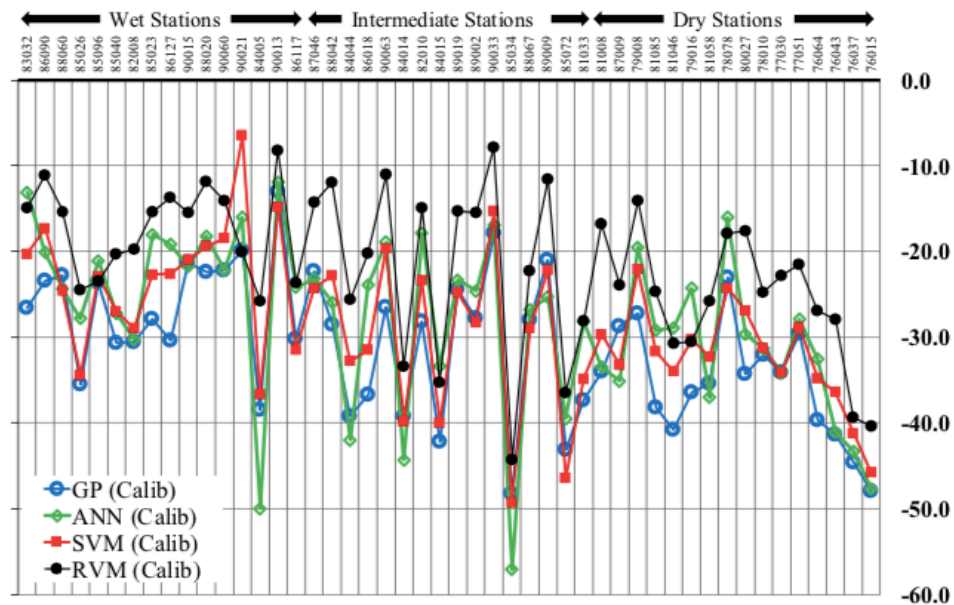在训练集中，模型对于降水的模拟在干燥地区偏差较大，在湿润地区偏差较小（图6a）；同样的，基于GP的方法偏差最小，而基于SVM方法的偏差最大。

（图6b）在验证集中，湿润地区的观测站中，基于RVM方法的偏差最大，然而在大部分干燥或是温和区域中，RVM方法却有较小的偏差；同样，对于大部分观测站，基于SVM的方法也具有较小的偏差，特别适合基于GP的方法比较。

Results

(c). Percentage of bias in standard deviation in Calibration

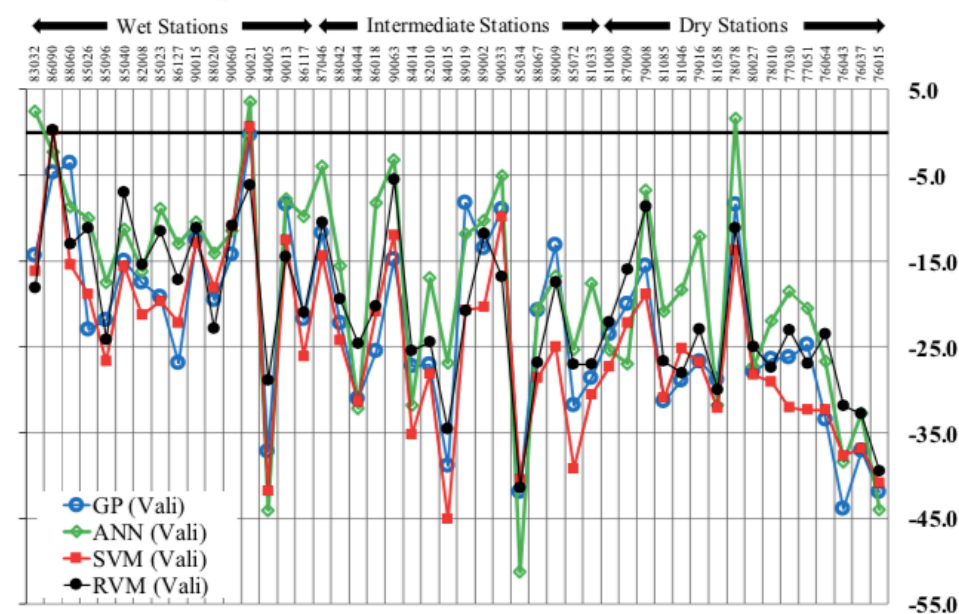(d). Percentage of bias in standard deviation in Validation

　　无论是在训练集还是在验证集，对于所有机器学习的方法，在大部分台站上，均低估了降水的标准差；而这一点在相对干燥的台站上表现更为明显。

　　机器学习在降尺度过程中对于变量标准差的低估已经在许多文献中提及过，这是由于大尺度气候信息所包含的方差为含有可解释所需水文气候变量方差的信息。
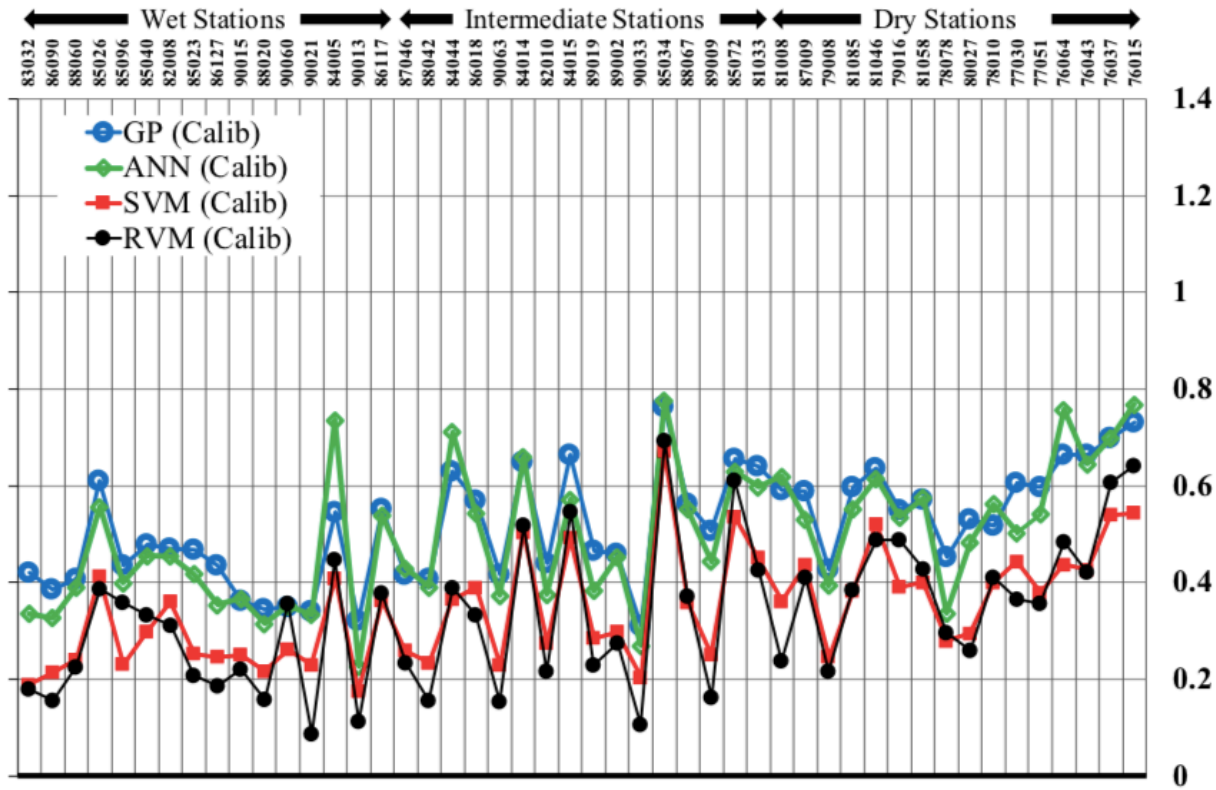
　　在验证集中，对于大多数的台站，ANN方法的标准差偏差比是最低的。

Results

(a). NMSE in Calibration
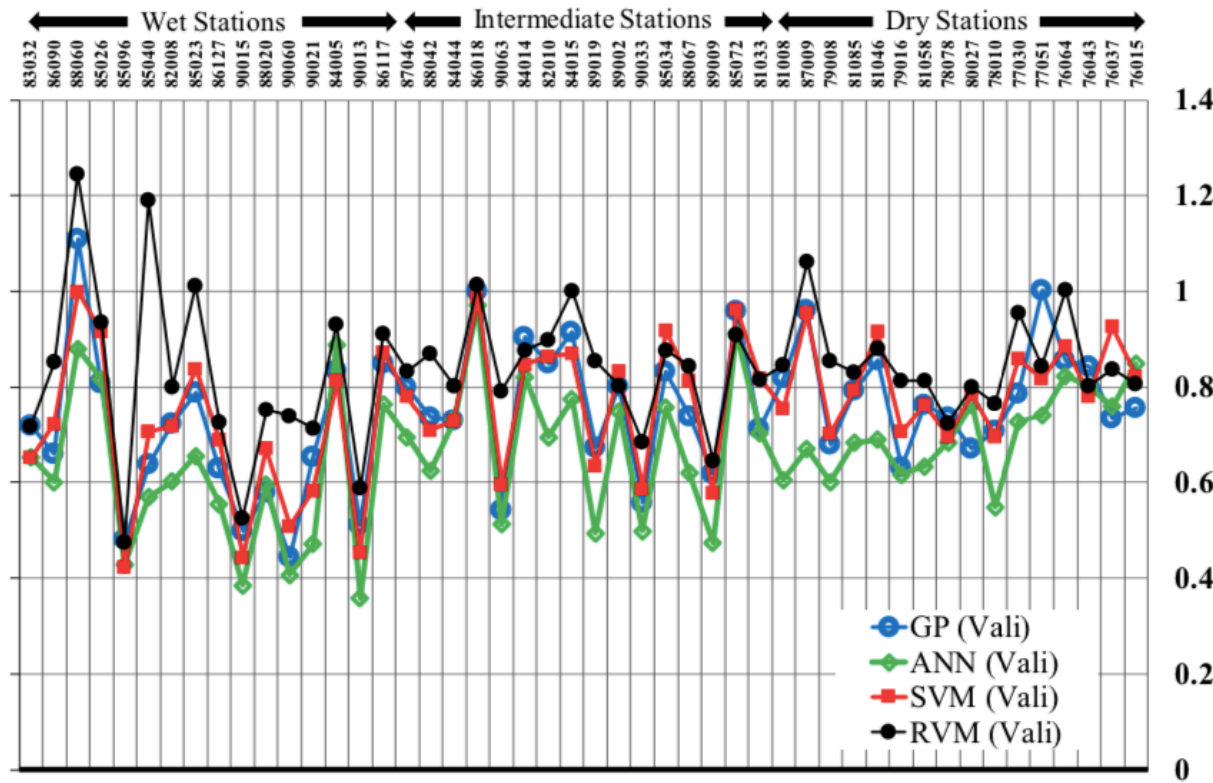
(b). NMSE in Validation
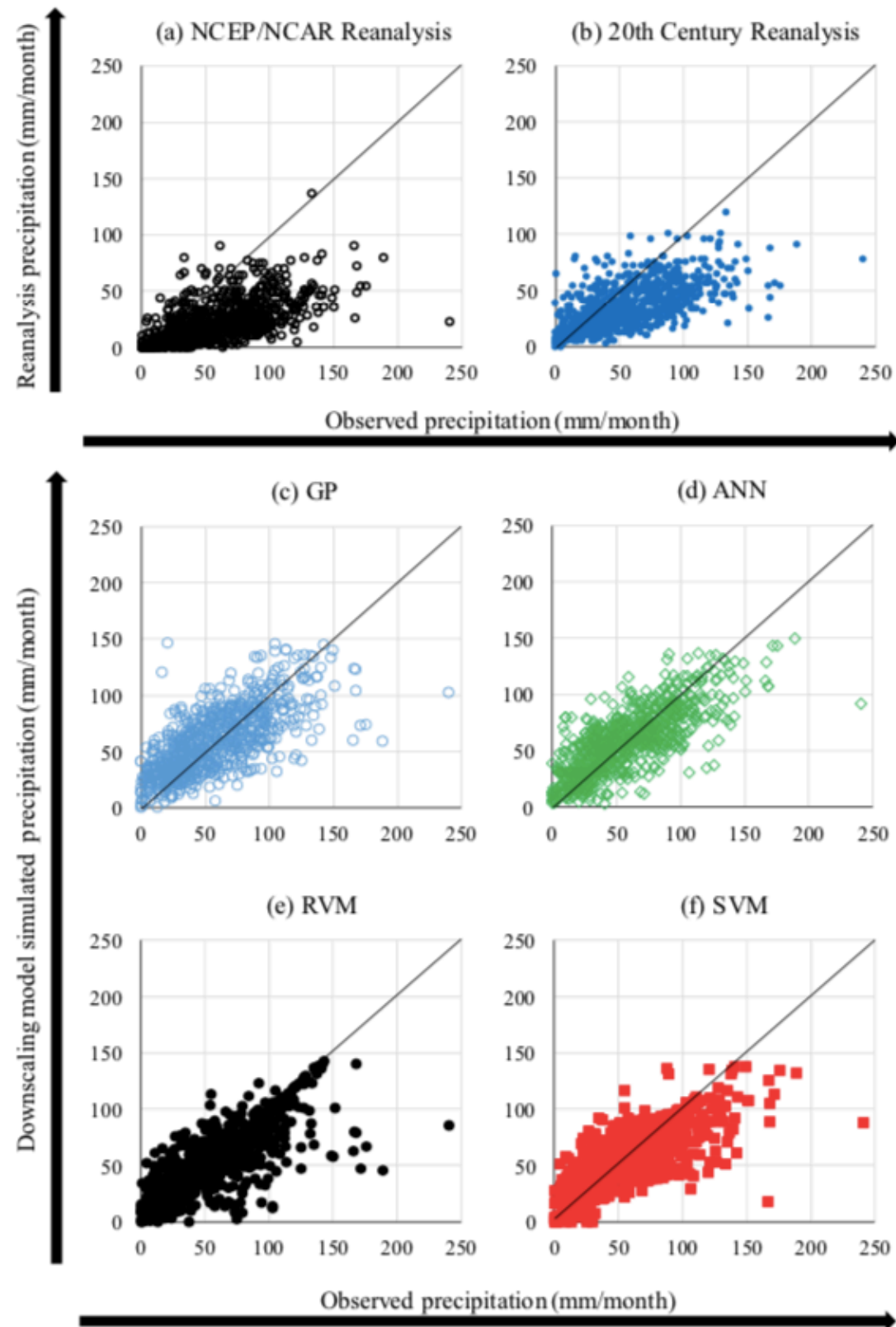
Fig. 9. Normalised mean square error of precipitation simulated by downscaling models.

　　在训练集中，无论是哪个地区，基于RVM方法的降尺度模型的NMSE最小，而GP和ANN的NMSE较大。在验证集中，基于RVM方法的降尺度模型在大部分潮湿地区的NMSE较大。

Results

Fig. 10. Agreement between observed precipitation, reanalysis precipitation, and downscaled precipitation.

a significant underestimation

Far more accurate than the raw precipitation

Conclusions

1. 基于RVM和ANN方法的降尺度模型在潮湿地区训练集和验证集上都具有较小的偏差，因此对于洪水的预报上更加推荐使用RVM和ANN方法；
2. 对于干燥地区的降水，RVM方法具有较小的偏差，因此在做干旱地区的分析时更加推荐使用RVM方法；
3. 无论是那种气候区域，SVM和RVM使用Polynomial kernel，Laplacian，Bessel，RBF和ANOVA都可以很好的模拟降水变量；同时，Spline，String，Linear和Hyperbolic tangent kernals无法得到降尺度模型所需的信息。

谢谢