

WeeklyNote

2019.06.02

張慕琪

INTERNATIONAL JOURNAL OF CLIMATOLOGY

Int. J. Climatol. **21**: 773–790 (2001)

DOI: 10.1002/joc.655

DOWNSCALING TEMPERATURE AND PRECIPITATION: A COMPARISON OF REGRESSION-BASED METHODS AND ARTIFICIAL NEURAL NETWORKS

J.T. SCHOOF* and S.C. PRYOR

*Atmospheric Science Program, Department of Geography, Indiana University Student Building, 701 East Kirkwood Ave.,
Bloomington, IN 47405-7100, USA*

Received 9 May 2000

Revised 22 January 2001

Accepted 26 January 2001

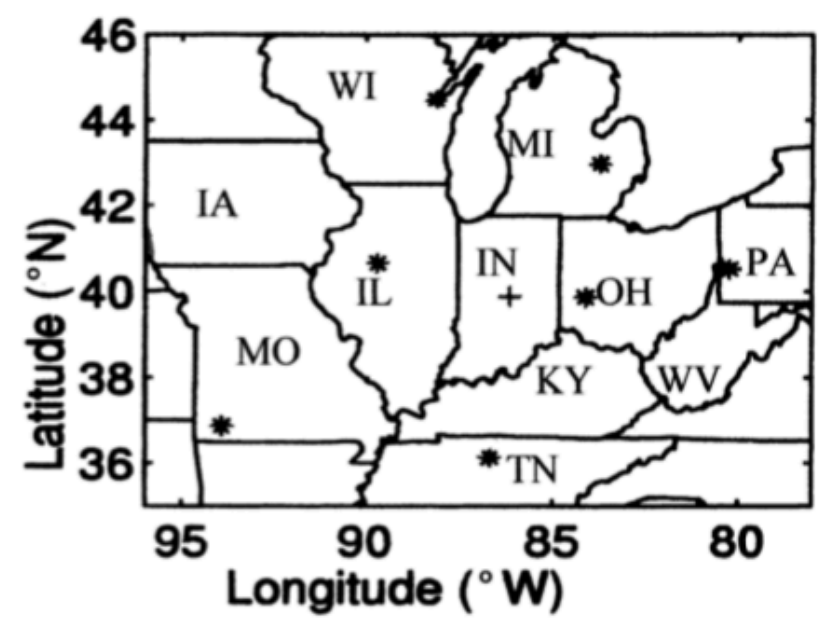
Introduction

农业生产所需要的天气尺度往往比现有模式的最高分辨率还要高。

高温时间的增加和低于正常水平的降水量很可能导致农作物的水汽环境改变（例如小麦和大豆），并因此对农业生产起到副作用。

Methods

1. 研究地区： from 35° to 45° N and from 80° to 95° W, includes Indiana (IN), Illinois (IL), Ohio (OH), Kentucky (KY), Tennessee (TN), Michigan (MI), Missouri (MO), western Pennsylvania (PA), and southeastern Wisconsin (WI)



该地区对于美国农业生产有很大作用

Figure 1. Map of study area showing rawinsonde stations (*) and Indianapolis, IN station (+)

Table I. Station information for rawinsonde stations

Station name	WMO station ID	Latitude (°N)	Longitude (°W)	Distance to Indianapolis (km)	Maximum length of data record (years)
Monett, MO	UMN 77349	36.88	93.90	720	26
Nashville, TN	BNA 72327	36.12	86.68	449	33
Flint, MI	ENT 72637	42.97	83.73	421	40
Peoria, IL	PIA 72532	40.67	89.68	325	37
Green Bay, WI	ORB 72645	44.48	88.13	529	51
Pittsburgh, PA	PIT 72520	40.53	80.23	520	32
Dayon, OH	DAY 72429	39.87	84.12	194	40

Methods

2. 研究时间：1973-1990（共18yr）

为确保这18年的时间段可以代表较长时间的特征（即18年是一个较长的时间范围），每个无线电探空台站地表温度和500hPa高度都和较长时间序列的数据一一比较；

使用Kolmogorov-Smirnov(K-S) test: 用以检验两个分布是否不同。

3. 研究变量：500hPa高度，700hPa高度，850hPa高度，海平面气压（SLP），柱平均相对湿度（column average relative humidity, RH）

4. 数据来源：Solar and Meteorological Surface Observation Network(SAMSON) dataset issued by the National Climatic Data Center(NCDC/NREL, 1993)

Methods

1. Synoptic classification:

- 1) 去除年循环
- 2) 根据指标进行季节划分

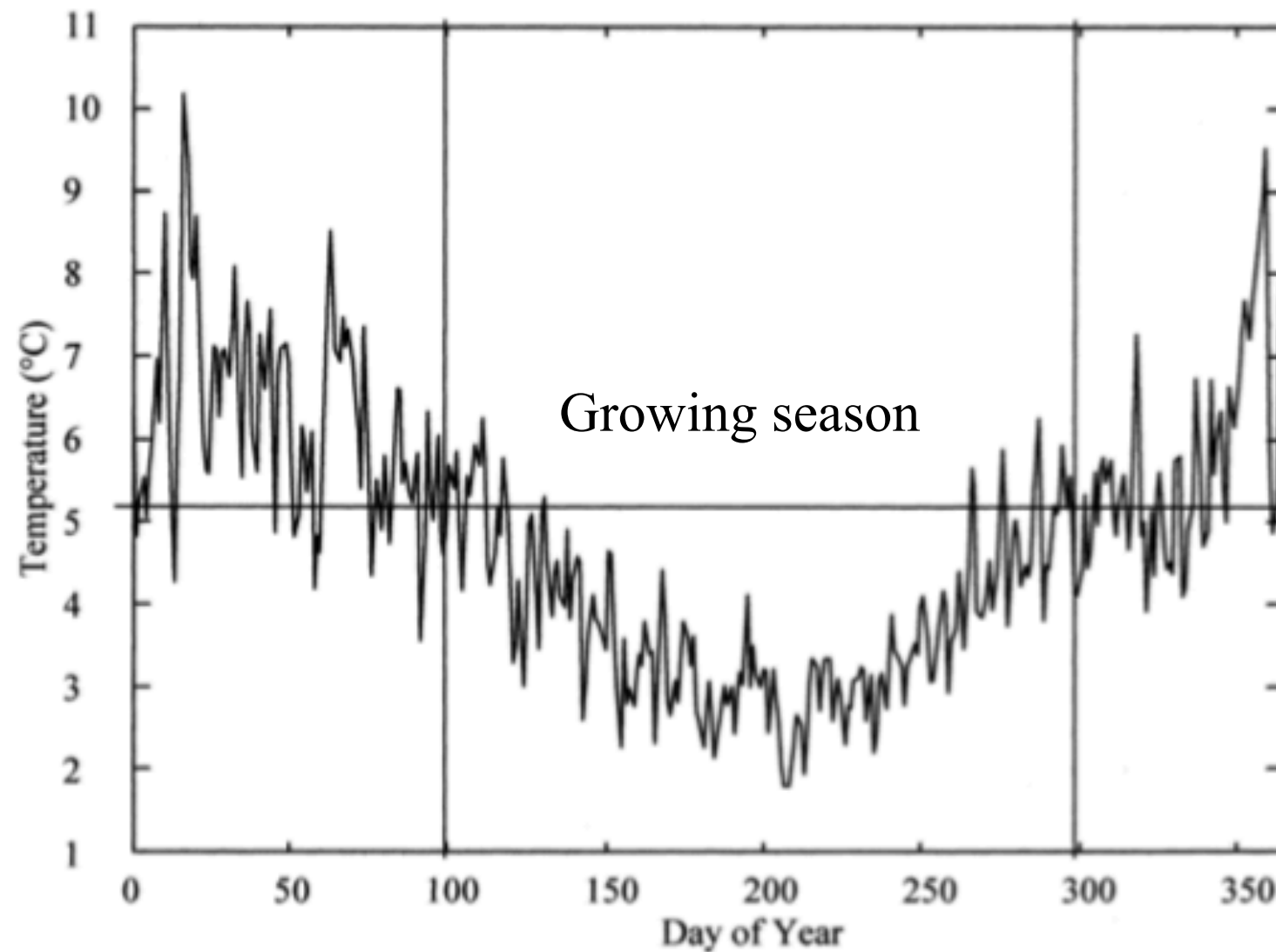


Figure 2. Standard deviation of daily surface air temperature for Indianapolis, IN. The horizontal line is the average standard deviation over the entire year. The seasons are determined by whether the mean daily standard deviation is above or below the annual mean

Methods

1. Synoptic classification:

3) 主成分分析 (PCA) : PCA的思想就是将n维特征映射到k维上 ($k < n$), 这k维是重新构造出来的全新维度特征, 而不是简单的从n维特征去除n-k维特征, 这k维就是主成分。

4) 聚类分析

2. Downscaling

1) Multiple regression

1. Multiple linear regression(MLR)用于日最高温和日最低温的降尺度, 温度变量为预测值, PC scores为用于预报的值;

2. Poisson regression model用于降水的降尺度。

2) Artificial neural networks(ANNs)

往往隐藏层的数量是预测变量树木的1.5倍

后向传播学习算法有两个参数: 学习速率和动量因子; 学习速率决定在一次学习中权重可以改变多少, 动量因子决定当前时间片内有权重受到前一次权重的影响

$$w_{i,j}(\text{new}) = w_{i,j}(\text{old}) + \eta \delta_i o_j + \alpha [\Delta w_{i,j}(\text{old})]$$

$$\delta_i = (t_i - o_i) o_i (1 - o_i)$$

Results

1. PCA: 根据主成分分析得出，在生长季节有7个主成分，而非生长季节有5个主成分，相应的，这些主成分分别可以解释探空台站数据集86%和83%的变化。

PC1和温度相关，其中在500hPa、700hPa高度和850hPa温度表现最佳；PC2和压力相关，在海平面气压变量上表现良好；PC3和湿度成分相关，因此在相对湿度变量上表现最佳；其余的四个成分同样和温度、压力和湿度相关，但是表现为不同方面。P.S. 非生长季节的数据并不重要，不做考虑。

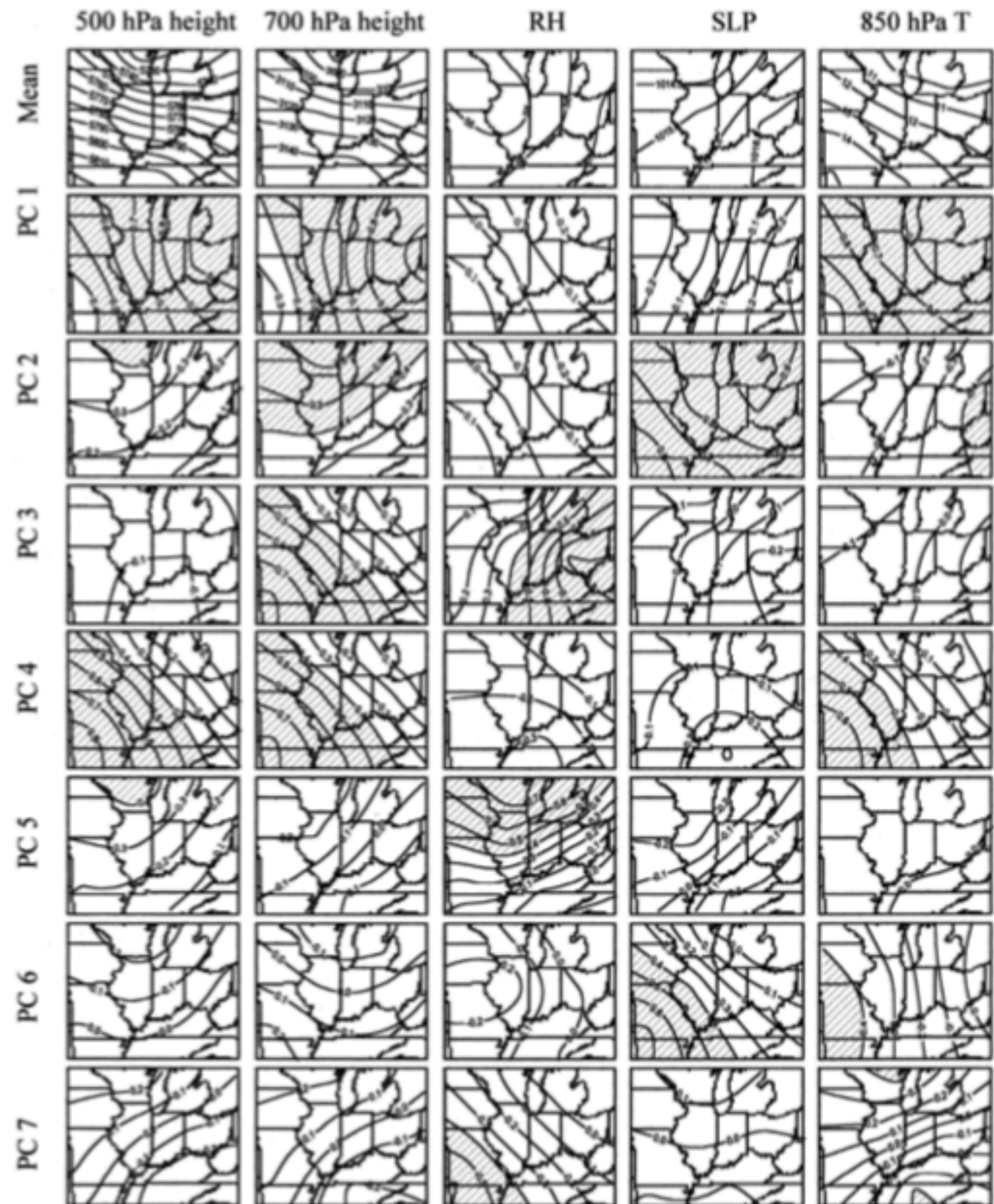


Figure 4. Growing season PC loadings. Values larger than ± 0.4 are shaded to aid in interpretation

Results

2. Downscaling: 75%的数据用于训练，25%用于测试。

1) Regression-based methods.

Table IV. Description of the regression models, input values and model forms^a

Predictand	Equation	Predictors
T_{\max} GS (°C)	$T_{\max}\text{GS} = 1.86P_1 - 0.36P_2 + 0.63P_3 + 1.24P_4 - 0.25P_5 + 0.31P_6 - 0.83P_7 + 24.95$	00:00 GMT PC scores
T_{\max} GSAU (°C)	$T_{\max}\text{GSAU} = 1.07P_1 - 0.41P_2 + 0.14P_3 + 0.52P_4 - 0.30P_5 + 0.07P_6 - 0.34P_7 + 0.76\text{AU} + 5.99$	00:00 GMT PC scores and an autoregressive term
T_{\min} GS (°C)	$T_{\min}\text{GS} = 2.79P_1 - 0.54P_2 + 1.16P_3 + 1.18P_4 + 0.39P_5 - 0.66P_6 - 0.47P_7 + 13.70$	12:00 GMT PC scores
T_{\min} GSAU (°C)	$T_{\min}\text{GSAU} = 1.17P_1 + 0.20P_2 - 0.15P_3 + 0.23P_4 + 0.54P_5 - 0.78P_6 + 0.21P_7 + 0.80\text{AU} + 2.74$	12:00 GMT PC scores and an autoregressive term
T_{\max} NS (°C)	$T_{\max}\text{NS} = 1.93P_1 - 2.37P_2 - 0.16P_3 + 0.10P_4 - 0.53P_5 + 7.16$	00:00 GMT PC scores
T_{\max} NSAU	$T_{\max}\text{NSAU} = 1.22P_1 - 1.25P_2 - 0.22P_3 + 0.24P_4 - 0.10P_5 + 0.69\text{AU} + 2.30$	00:00 GMT PC scores and an autoregressive term
T_{\min} NS (°C)	$T_{\min}\text{NS} = 3.80P_1 - 3.10P_2 - 0.15P_3 + 0.36P_4 - 0.32P_5 - 2.87$	12:00 GMT PC scores
T_{\min} NSAU	$T_{\min}\text{NSAU} = 2.15P_1 - 1.12P_2 - 0.21P_3 - 0.32P_4 + 0.11P_5 + 0.68\text{AU} - 0.94$	12:00 GMT PC scores and an autoregressive term
PREC_dGS (mm)	$\text{PREC}_d\text{GS} = 0.04P_{12}1 - 0.30P_{12}2 + 0.71P_{12}3 - 0.09P_{12}4 + 0.20P_{12}5 - 0.10P_{12}6 - 0.45P_{12}7 - 0.10P_{00}1 - 0.07P_{00}2 + 0.19P_{00}3 + 0.14P_{00}4 - 0.09P_{00}5 + 0.02P_{00}6 + 0.04P_{00}7 + 0.84$	00:00 GMT PC scores ($P_{00} \#$) and 12:00 GMT PC scores ($P_{12} \#$)
PREC_dNS (mm)	$\text{PREC}_d\text{NS} = 0.31P_{12}1 - 0.29P_{12}2 + 0.42P_{12}3 + 0.59P_{12}4 + 0.29P_{12}5 - 0.09P_{00}1 - 0.07P_{00}2 + 0.04P_{00}3 - 0.08P_{00}4 + 0.02P_{00}5 + 0.57$	00:00 GMT PC scores ($P_{00} \#$) and 12:00 GMT PC scores ($P_{12} \#$)

^a The predictors in the regression equations ($P \#$) indicate the PC scores. GS indicates the growing season and NS indicates the non-growing season. The term AU refers to the autoregressive term.

Results

2. Downscaling: 75%的数据用于训练，25%用于测试。
- 1) Regression-based mothods.

Table V. Model evaluation statistics for the regression models^a

Model	μ		σ		r	MAE	RMSE	RMSE _s	RMSE _u
	OBS	PRED	OBS	PRED					
T_{\max} GS	24.65	24.92	6.03	2.67	0.51	4.15	5.22	4.69	2.30
T_{\max} GSAU	24.65	24.88	6.03	5.25	0.85	2.31	3.18	1.58	2.76
T_{\min} GS	13.78	13.80	6.49	3.40	0.60	4.35	5.21	4.44	2.72
T_{\min} GSAU	13.78	13.85	6.49	5.52	0.88	2.39	3.08	1.63	2.61
T_{\max} NS	8.25	7.71	6.67	2.93	0.34	6.68	8.16	7.68	2.75
T_{\max} NSAU	8.25	7.46	8.67	6.64	0.81	4.04	5.17	3.39	3.90
T_{\min} NS	-3.09	-2.93	9.38	5.08	0.55	4.87	7.84	6.59	4.25
T_{\min} NSAU	-3.09	-3.66	9.38	9.59	0.80	3.23	6.09	1.83	5.81
PREC _d GS	2.69	0.66	7.19	1.12	0.36	2.97	7.16	7.09	1.04
PREC _d NS	2.76	0.49	6.06	0.99	0.50	2.72	6.07	6.01	0.86

^a The observed values are indicated by OBS, and the model predictions by PRE. The table includes the predicted and observed means (μ) and S.D.s (σ), Pearson product-moment correlation coefficient (r), MAE, RMSE, and the systematic and unsystematic components (RMSE_s and RMSE_u). The models are as described in Table IV.

PC scores 对于 T_{\max} GS和 T_{\max} GSAU变量有较高的相关系数；然而， T_{\max} GS和PC3和PC4有更高的相关性， T_{\max} GSAU和PC1和PC4有更高的相关性。对于 T_{\min} GS和PC1、PC3有更大相关性。

Results

2. Downscaling: 75%的数据用于训练, 25%用于测试。
 2) ANNs.

Table VI. Learning rate, momentum factor, number of hidden nodes and sample sizes for the training and testing of the ANNs

Model	Learning	Momentum	# Hidden nodes	Train	Test
T_{\max} GS (°C)	0.009	0.9	10	1896	632
T_{\max} GSAU (°C)	0.01	0.5	12	1896	632
T_{\min} GS (°C)	0.012	0.9	10	2455	671
T_{\min} GSAU (°C)	0.012	0.8	10	2455	671
PREC _d GS (mm)	0.05	0.3	18	1813	604
PREC _m GS (mm)	0.3	0.05	12	68	22
T_{\max} NS (°C)	0.007	0.9	8	1316	435
T_{\max} NSAU (°C)	0.007	0.7	9	1316	435
T_{\min} NS(°C)	0.012	0.9	7	1995	611
T_{\min} NSAU (°C)	0.012	0.9	8	1995	611
PREC _d NS (mm)	0.05	0.3	15	1230	409
PREC _m NS (mm)	0.001	0.1	15	68	22

Results

2. Downscaling: 75%的数据用于训练，25%用于测试。

1) Regression-based methods.

Table VII. Model evaluation statistics for the ANNs. Entries are analogous to those in Table V

Model	μ		σ		r	MAE	RMSE	RMSE _s	RMSE _u
	OBS	PRED	OBS	PRED					
T_{\max} GS	24.77	24.84	6.07	3.12	0.58	3.86	4.98	4.27	2.55
T_{\max} GSAU	24.77	24.66	6.07	5.13	0.85	2.40	3.18	1.70	2.68
T_{\min} GS	13.69	13.73	6.51	3.59	0.63	4.24	5.08	4.25	2.79
T_{\min} GSAU	13.69	13.87	6.51	5.60	0.89	2.27	2.95	1.53	2.53
T_{\max} NS	7.63	7.60	8.16	3.10	0.36	6.12	7.61	7.03	2.89
T_{\max} NSAU	7.63	7.18	8.16	6.49	0.80	3.81	4.87	2.98	3.85
T_{\min} NS	-2.84	-2.73	7.59	5.00	0.64	4.52	5.81	4.37	3.83
T_{\min} NSAU	-2.84	-2.82	7.59	6.92	0.87	2.76	3.76	1.57	3.42
PREC _d GS	2.84	3.43	8.17	0.05	0.09	4.71	8.19	8.19	0.05
PREC _d NS	2.79	2.63	6.01	0.06	0.11	3.80	6.01	6.00	0.06
PREC _m GS	80.27	90.98	66.6	40.38	0.65	42.18	52.00	41.89	30.82
PREC _m NS	57.18	68.02	28.78	13.28	0.54	22.88	26.60	24.15	11.16

ANNs日降水无法和泊松回归的结果相比较。

神经网络的月平均降水数据表现的比日降水模型好很多

Conclusions

1. 对于日最高温变量来说，ANNs的表现要比MLR模型表现的好（table V和table VII比较），ANNs中的预报值和用于预报的值之间有更高的相关系数；
2. 同样，ANNs的温度模型有更小的偏差，因此具有更大的预报性；并且，神经网络往往对于平均值和标准差的预估更加准确；
3. 在table V和table VII中，非生长季节的日最高温模型并没有生长季节的模型表现的好；然而，线性回归模型比神经网络能更好的预估日最高温的平均值；无论是线性回归模型还是神经网络模型都无法捕捉方差；
4. 关于PC scores的降尺度分析表明其同时含有两个季节的日最低温信息；
5. 神经网络的 $T_{\min}GS$ 变量要比线性回归模型更好；在测试集中，神经网络的相关系数要比线性回归模型略高一些；

Conclusions

6. 无论是线性回归还是神经网络都无法成功捕捉日降水信息，两个模型各有各的缺点导致；
7. 2000年，Cannon and Lord提出了一种直方图均等过程（histogram equalization process）可能可以帮助神经网络中这类情况的提高；
8. ANNs的月平均降水要远比日平均降水好很多；这说明ANNs可以学习到数据中普遍存在的规律，但无法捕捉到与数据相关的规律（这些规律偏离了月降水量的平均值）。

谢谢

附录： 主要参考页面

PCA：

1. <https://blog.csdn.net/u014182497/article/details/78663353>

K-S test：

2. <https://www.cnblogs.com/sddai/p/5737408.html>