

李亚飞, 王磊斌, 毛慧琴, 等. 2016. 使用岭回归对哈萨克斯坦月平均气温的统计降尺度研究 [J]. 气候与环境研究, 21 (5): 567-576. Li Yafei, Wang Leibin, Mao Huiqin, et al. 2016. Statistical downscaling of monthly mean temperature for Kazakhstan using ridge regression [J]. Climatic and Environmental Research (in Chinese), 21 (5): 567-576, doi: 10.3878/j.issn.1006-9585.2016.16027.

# 使用岭回归对哈萨克斯坦月平均气温的统计降尺度研究

李亚飞<sup>1</sup> 王磊斌<sup>1</sup> 毛慧琴<sup>2</sup> 延晓冬<sup>1</sup>

<sup>1</sup> 北京师范大学地表过程与资源生态国家重点实验室, 北京 100875

<sup>2</sup> 环境保护部卫星环境应用中心, 北京 100094

**摘要** 哈萨克斯坦是世界最大的内陆国家, 拥有典型的大陆性气候和多样的地理环境及生态系统, 同时哈萨克斯坦的自然环境和人类社会对于气候变化这一全球性问题是敏感的、脆弱的, 需要运用科学的研究方法应对气候变化的挑战。通常, 区域或局地尺度的气候变化影响研究需要对气候模式输出或再分析资料进行降尺度以获得更细分辨率的气候资料。近年来, 大量验证统计降尺度方法在各个地区能力的研究见诸文献, 然而在哈萨克斯坦地区验证统计降尺度方法的研究非常少见。本文使用了岭回归的方法对哈萨克斯坦地区 11 个气象站点 1960~2009 年的月平均气温进行了统计降尺度研究。结果显示, 使用前 30 年数据和岭回归模型建立大尺度预报因子和观测资料的统计关系可以较好地预测后 20 年的月平均气温, 预测能力在各站各月均有不同程度的差异, 地形复杂的站点预测效果较差, 夏季预测结果好于冬季; 此外, 将哈萨克斯坦地区平均来看则与观测数据相吻合。

**关键词** 哈萨克斯坦 统计降尺度 岭回归 月平均气温

文章编号 1006-9585 (2016) 05-0567-10

中图分类号 P468.0

文献标志码 A

doi:10.3878/j.issn.1006-9585.2016.16027

## Statistical Downscaling of Monthly Mean Temperature for Kazakhstan Using Ridge Regression

LI Yafei<sup>1</sup>, WANG Leibin<sup>1</sup>, MAO Huiqin<sup>2</sup>, and YAN Xiaodong<sup>1</sup>

<sup>1</sup> State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875

<sup>2</sup> Satellite Environment Center, Ministry of Environmental Protection, Beijing 100094

**Abstract** Kazakhstan is the largest landlocked country in the world with a typical continental climate, and its natural environment and human society are sensitive and vulnerable to climate change. In climate change impact studies, one widely-used approach to obtain future climate change scenario at regional or local scale is to downscale future climate projection from General Circulation Model (GCM). So far, to our knowledge, no statistical downscaling study has been carried out in Kazakhstan region. In this study, the authors explored and validated the ability of a statistical downscaling model that is based on ridge regression to predict monthly mean temperature at 11 stations in Kazakhstan from NCEP/NCAR monthly mean reanalysis. The 30-year dataset for the period from 1960 to 1989 was used to train the downscaling model

收稿日期 2016-01-26; 网络预出版日期 2016-03-07

作者简介 李亚飞, 男, 1990 年出生, 博士研究生, 主要从事气候模式统计降尺度研究。E-mail: yafei.li@mail.bnu.edu.cn

通讯作者 延晓冬, E-mail: yxd@bnu.edu.cn

资助项目 北京师范大学地表过程与资源生态国家重点实验室自由探索项目 2015-ZY-13, 北京师范大学环境演变与自然灾害教育部重点实验室研究生自主基金 2015jzhz15

**Funded by** The free exploration program of State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University (Grant 2015-ZY-13), Key Laboratory of Environmental Change and Natural Disaster, Ministry of Education, Beijing Normal University (Grant 2015jzhz15)

and the next 20-year data for the period of 1990–2009 was used for validation of the downscaling model. The result shows that despite certain disagreements with observations at several stations, the ridge regression model generally is able to reasonably reproduce monthly mean temperature over Kazakhstan region. The authors also find that the performance of the ridge regression model is better in the summer than in the winter and better in flat terrain areas than in complex terrain areas.

**Keywords** Kazakhstan, Statistical downscaling, Ridge regression, Monthly mean temperature

## 1 前言

哈萨克斯坦是中亚地区（中亚五国：哈萨克斯坦、乌兹别克斯坦、土库曼斯坦、吉尔吉斯斯坦、塔吉克斯坦）面积最大的国家，也是世界上最大的内陆国家，其 270 万平方公里的面积几乎可与整个西欧相当。因其独特的地理特征和经济地位，哈萨克斯坦的自然环境和人类社会对于全球气候变化的响应是敏感而脆弱的(de Beurs and Henebry, 2004; Giorgi, 2006; Lioubimtseva and Henebry, 2009)。

这一区域分布着迥异的自然环境：哈萨克斯坦的西部是低于海平面的里海低地，中部横亘着世界最大的干旱草原——哈萨克草原，北部森林葱郁，南部沙漠无垠，而东部则矗立着海拔近 7 km 的雪域高原。这种自然地理条件，使得未来的气候变化对哈萨克斯坦地区存在较大影响。例如，哈萨克斯坦东部地区水资源中有很高比例来自于西部帕米尔高原上的冰雪融水，高原上的温度变化将会造成融水体量的快速而巨大改变，进而可能造成下方的干旱或洪涝灾害(龙爱华等, 2010; Malsy et al., 2015)。20 世纪 90 年代建国之后，哈萨克斯坦经济飞速发展，2012 年其人均国民生产总值达 11900 美元(李同昇和龙冬平, 2014)，居中亚国家之首。中国提出“一带一路”战略后，哈萨克斯坦作为中亚区域经济大国又地处亚欧大陆中心，是联通中国与欧洲的重要陆上纽带，是我国向西开放的必经之地，也是建设丝绸之路经济带的重要一环。综合相关研究及以上分析，研究气候变化对哈萨克斯坦地区的自然环境及人类社会的影响，对当地水资源管理、生态环境保护、粮食安全决策、灾害与风险评估、基础设施投资建设，对中国“一带一路”项目的评估预测与顺利实施，具有重要意义(Wright et al., 2009; Qi et al., 2012; Sommer et al., 2013; Bobojonov and Aw-Hassan, 2014; Eisfelder et al., 2014; Hu et al., 2014)。

在气候变化影响评估的研究中，月平均温度作

为一个最常用的参数，在各种评估模型中扮演着重要的角色，故对未来平均温度的预估的质量极大程度上影响着气候变化影响研究的可靠性和准确性。目前，科学的未来气候变化预估依据于全球气候模式(GCMs)在不同温室气体排放情景下对气候的模拟。然而，GCM 重在在大尺度气候的模拟，其空间分辨率往往较粗，很难满足小尺度气候变化影响研究的要求(Wilby et al., 2004)。鉴于此，降尺度手段应运而生，以期构建大尺度环流场(预报因子)与小尺度地表气候变量(预报量)的关系，进而将 GCM 的气候预估转化到区域尺度、甚至站点尺度上去。近年来，气候模式的降尺度技术备受重视，发展迅猛，见诸历次 IPCC 政府间气候变化评估报告(Flato et al., 2013)。

降尺度技术目前主要分为两类，即动力降尺度(Laprise, 2008; Orskaugo et al., 2011; Mannig et al., 2013; de Sales and Xue, 2013; 王树舟和于恩涛, 2014)和统计降尺度(Hanssen-Bauer et al., 2005; 范丽军等, 2007; Wetterhall et al., 2009; Goyal and Ojha, 2011)。动力降尺度方法因其需要运行内嵌于 GCM 的区域气候模式(RCM)，对于计算量的需求很大，并且，降尺度的表现受外套的提供边界条件的 GCM 的影响。而统计降尺度方法，相较于动力降尺度有自身的优势。例如，统计降尺度所实现需要的计算量小，难度低，易应用，还能提供动力降尺度所不能的站点尺度的气候信息(Wilby et al., 2002, 2004)。

使用统计降尺度以连接大尺度环流和区域气候变量的方法繁多，大致可依 Fowler et al. (2007) 的综述分为 5 类，而其中每一类又有更为细致的实现层面上区分。不同的统计降尺度方法在不同的区域其降尺度能力或许截然不同，那么这种在不同的区域验证不同的统计降尺度方法的对比研究，就可以为统计降尺度方法自身的发展及当地的气候变化影响评估提供科学的指导。此类对比研究的研究区域多在欧美发达国家，中国、印度等国家也存在为数不少的研究(Goyal et al., 2012; Kazmi et al.,

2015; Zhang and Yan, 2015), 然而, 据作者所知, 并没有一例统计降尺度研究关注于哈萨克斯坦, 这与前文所述的该地区对于地区经济发展和我国“一带一路”建设的重要地位是不相匹配的, 因此, 对哈萨克斯坦地区在算法层面上的统计降尺度评估研究是有必要的。

本文应用了岭回归的方法和 NCEP/NCAR 再分析资料对哈萨克斯坦范围内 11 个气象站点的月平均气温进行了统计降尺度研究, 使用 1960~1989 年 30 年数据集建立统计降尺度模型并用 1990~2009 年 20 年的数据集进行验证, 探索了岭回归作为统计降尺度模型在哈萨克斯坦地区月平均气温的模拟能力。

## 2 数据和方法

### 2.1 数据

本研究采用全球历史气候网 (Global Historical Climatology Network, GHCN) 观测资料 (<https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/global-historical-climatology-network-ghcn>[2016-03-04]) 以获得站点观测的月平均气温序列作为预报量训练和验证降尺度模型, 并筛选了哈萨克斯坦境内有 1960~2009 年各月完整观测资料的站点共 11 个, 其位置和基本信息分别展示在图 1 和表 1 中。NCEP/NCAR 再分析资料 (Kalnay et al., 1996; <http://www.esrl.noaa.gov/psd/data/gridded/reanalysis/>[2016-03-04]) 来自于水平方向  $2.5^\circ$  (纬度)  $\times 2.5^\circ$  (经度) 空间分辨率、垂直方向 17 层的大气模式模拟, 被广泛认为可以当作理想的模式输出数据用在降尺度方法的发展和验证上, 因此本文采用 NCEP 再分析资料作为大尺度预报因子。本文中为配合可获取的哈萨克斯坦站点观测资料长度, NCEP 再分析资料亦被截取为 1960~2009 年。根据以往对月平均气温统计降尺度的研究, 本文选取了 NCEP 资料中 925 hPa、500 hPa 和 200 hPa 的大尺度温度场 ( $T_{925}$ 、 $T_{500}$ 、 $T_{200}$ , 其中,  $T$  表示温度场, 数字表示层高, 下同)、位势高度 ( $H_{925}$ 、 $H_{500}$ 、 $H_{200}$ )、纬向风场 ( $u_{925}$ 、 $u_{500}$ 、 $u_{200}$ ) 和经向风场 ( $v_{925}$ 、 $v_{500}$ 、 $v_{200}$ ), 以作为大尺度预报因子。

### 2.2 方法

建立统计降尺度模型首先需要数据集进行截取, 站点观测数据和大尺度预报因子都被分为前

表 1 站点基本信息

Table 1 Basic information for the selected Global Historical Climatology Network (GHCN) climatological stations

站点	海拔/m	纬度	经度	1 月平均 气温/ $^\circ\text{C}$	7 月平均 气温/ $^\circ\text{C}$
Aktobe	219	50.28°N	57.15°E	-13.39	22.62
Aralskoe More	62	46.78°N	61.67°E	-11.44	26.86
Uralsk	36	51.25°N	51.40°E	-11.55	22.63
Atyrau	-22	47.12°N	51.92°E	-7.30	26.42
Atbasar	304	51.82°N	68.37°E	-16.99	20.21
Balhash	350	46.80°N	75.08°E	-13.80	23.96
Semej	196	50.42°N	80.30°E	-14.83	21.71
Irtysk	94	53.35°N	75.45°E	-17.00	20.99
Petropavlovsk	142	54.83°N	69.15°E	-16.72	19.55
Zharkent	645	44.17°N	80.07°E	-6.90	24.43
Uil	128	49.07°N	54.68°E	-11.50	24.78

30 年的训练集, 从 1960~1989 年, 用以训练统计降尺度模型, 建立起预报变量和预报因子的关系; 以及 1990~2009 年的测试集, 用以验证统计降尺度模型的模拟能力; 进而, 预报因子需要进行标准化处理和主成分变换以减少预报因子维度和计算量; 随后, 统计模型建立起预报因子和预报变量的关系并使用测试集的预报因子得到预测值; 最后, 利用统计量衡量预测值与观测值之间的差距以量化衡量模型预测能力。

#### 2.2.1 预报因子变换

本文对各个站点选择距离其最近的 9 个 NCEP 格点作为预报因子区域(以 Aktobe 站为例, 见图 2)。对各站将有在水平上 9 个格点、垂直上 3 层的 4 个大尺度预报因子变量, 共 108 ( $9 \times 3 \times 4$ ) 个相互之间不独立的因子。为解决因子数量多而导致的过拟合问题 (即对训练集数据过度拟合而导致对测试集数据预测结果不稳定), 本文采用主成分分析 (Principal Component Analysis, PCA) 方法 (Benestad et al., 2015) 对大尺度预报因子降维。举例而言, 对 Aktobe 站点 1 月的大尺度预报因子 ( $T_{925}$ 、 $T_{500}$ 、 $T_{200}$ 、 $H_{925}$  等前文所述 12 个) 进行 PCA 变换, 首先需要各个变量标准化, 即减去各量平均值并除以标准差; 而后对该站对应的 9 个 NCEP 格点的区域的所有值使用主成分分析算法变换, 并选择变换结果中的第 1 个主成分(含有信息量最多, 大多数解释 95% 以上变异), 于是该组预报因子即被变换为 12 ( $1 \times 3 \times 4$ ) 个, 较之前大大减少。依此方法对各站各月大尺度预报因子循环处理, 直至全部完成。

#### 2.2.2 岭回归模型

对大尺度预报因子的 PCA 变换只能排除因子

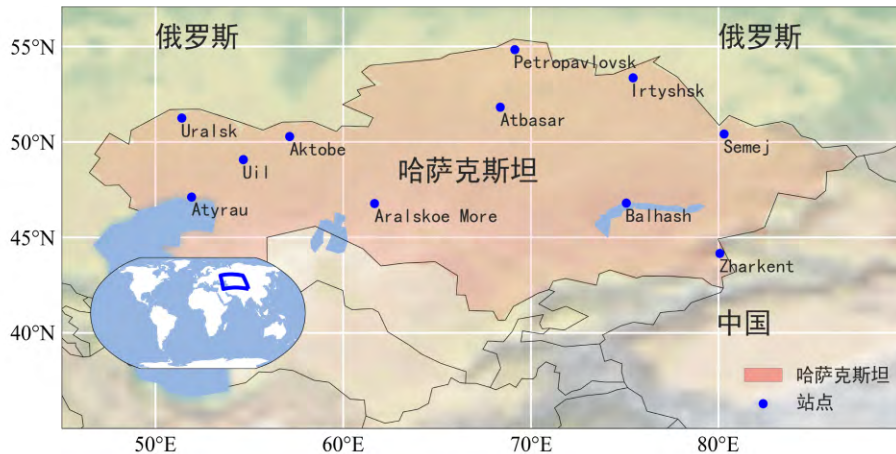


图 1 选取站点位置

Fig. 1 Locations of selected Global Historical Climatology Network (GHCN) climatological stations

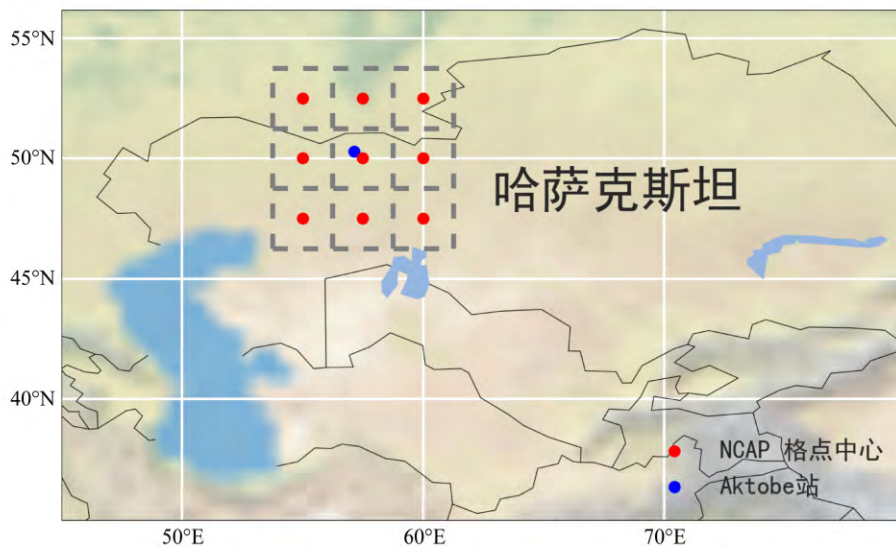


图 2 Aktobe 站点对应的预报因子区域示意图

Fig. 2 Illustration of the domain within which NCEP/NCAR grids are used as the predictor for Aktobe station

之间空间分布上的相关性，但是变量与变量之间、不同气压高度之间的预报因子仍然要求相互之间没有显著的相关性 (Hewitson and Crane, 1996)，以避免回归模型出现过拟合的情况。鉴于此，本文采用岭回归模型，通过引入正则项处理过拟合问题。岭回归模型系数的估计过程即求最小化正则化模型的离差平方和：

$$\omega = \arg \min \|X\omega - y\|_2^2 + \alpha \|\omega\|_2^2, \quad (1)$$

此处， $X$ 、 $y$ 、 $\omega$  分别为预报因子、预报变量和回归系数， $\alpha$  为控制正则项作用效果的参数，该值越大则系数对多重共线性越稳定，不易出现过拟合；而该系数如果趋近于 0 即为实质上的普通最小二乘回

归模型。 $\alpha$  使用多次试验的方法进行确定，即多次尝试不同的  $\alpha$  值建立模型，取测试集上的预测结果最好时的值使用。

### 2.2.3 模型评估方法

为量化评估降尺度模型的模拟效力，本文采用了均方根误差 (root-mean-square error, RMSE)、平均绝对误差 (mean absolute error, MAE) 和可决系数  $R^2$  三种统计量描述模型对测试集数据的预测误差和模拟能力，计算公式参见 Goyal and Ojha (2011)。

## 3 结果

在进行统计降尺度之前，需要对大尺度预报因

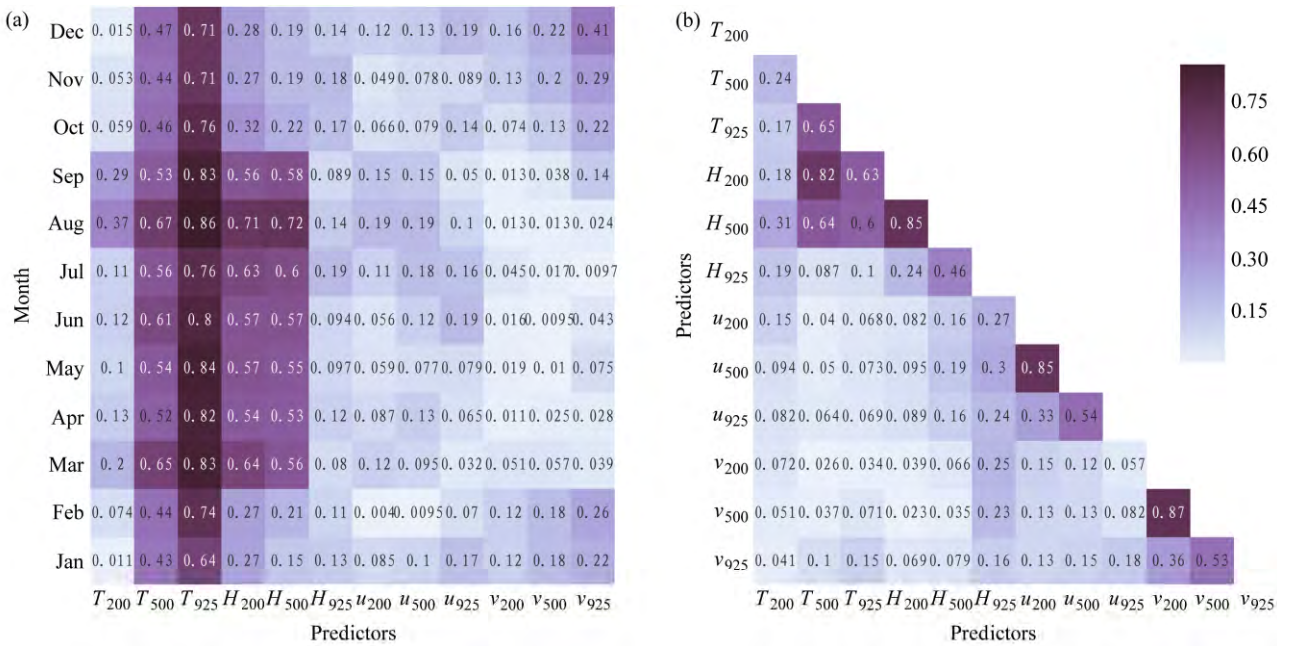


图3 (a) 预报因子与预报量之间及 (b) 预报因子相互之间的 Pearson 相关系数的平方  
Fig. 3 Squares of Pearson's correlation coefficient between (a) potential predictors and predictands and (b) potential predictors themselves

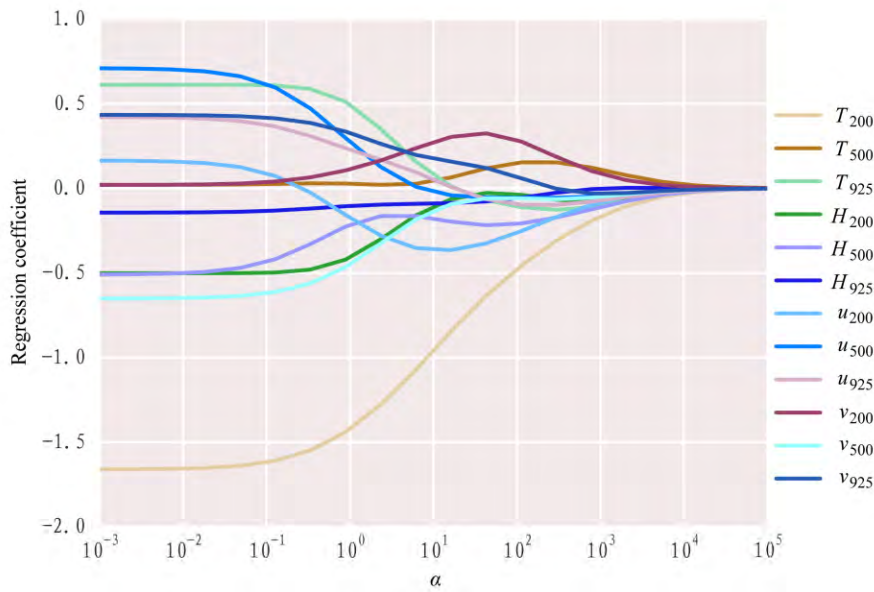


图4 Aktobe 站 1 月回归系数随  $\alpha$  变化的情况  
Fig. 4 Regression coefficients in Jan as a function of  $\alpha$  at Aktobe station

子做相关性分析以了解预报因子的特征，以及需要确定岭回归模型的参数  $\alpha$ 。图 3 所示为预报因子变量之间的 Pearson 相关系数的平方，其值的相对大小可反映出变量之间的相关性大小 (Pearson, 1896)。图 3a 显示了对所选的哈萨克斯坦 11 站平均之后预报因子与预报量之间各月的相关关系。结果可见，各预报因子中， $T_{925}$  与预报量相关性最高，除 12 月以外各月 Pearson 相关系数的平方均能达到 0.7 以上。

环流因子中，位势高度场与预报量相关程度在大多数月份高于风场，并且可以看出较为明显的冬、夏半年的差别。4~10 月，位势高度场与预报量的相关系数的平方均高于 0.5，而在 11 月至次年 3 月该值则大多为 0.3 以下。风场的结果更加明显，纬向风场 ( $u_{925}$ 、 $u_{500}$ 、 $u_{200}$ ) 在冬半年与预报量相关性低而在夏半年高，而经向风场 ( $v_{925}$ 、 $v_{500}$ 、 $v_{200}$ ) 则正好相反，在夏半年与预报量相关性低而在冬半年

高。

这一结果也可以印证哈萨克斯坦明显的季节分明的气候特征, 冬季西伯利亚冷高压从哈萨克斯坦北部南侵, 造成大范围降温, 故南北方向的经向风场与平均气温相关性强, 而夏季气流自西向东从里海向哈萨克斯坦内地输送暖湿气流, 故此时纬向风场与平均气温相关性强。图 3b 为所选 11 站大尺度预报因子之间的相关系数平方的全年平均值。如图可见, 同一类大尺度预报因子的不同高度之间多有相关, 如  $u_{200}$  与  $u_{500}$  之间相关系数的平方达到 0.85; 不同类预报因子之间也有一定相关性, 如  $H_{200}$  与  $T_{500}$  之间相关系数的平方达到 0.85。

因此, 使用岭回归方法排除回归模型中的多重共线性是有必要的。在对岭回归模型参数  $\alpha$  的确定过程中, 经过对多站点多月份的试验, 本文认为在使用岭回归模型对哈萨克斯坦地区月平均气温进行统计降尺度时将  $\alpha$  设置为 10 较为合理, 既能起到对相关的预报因子限制的作用保证了预测结果的稳定性, 又不至于掩盖预报因子的贡献以至于丧失模型预测的准确性。以 Aktobe 站 1 月为例 (如图 4), 当  $\alpha$  过小时, 正则项起不到作用, 回归模型各项系数分散, 此时模型如普通最小二乘多元回归模型, 出现过拟合现象, 预测结果不稳定; 当  $\alpha$  过大时, 模型各项系数收敛到一处, 出现欠拟合现象, 预测结果不准确; 而当  $\alpha$  合理确定时, 平衡了模型的稳定性和准确性。

对岭回归统计降尺度模型在测试集上的预测结果进行分析, 图 5 展示了所选哈萨克斯坦 11 站 1990~2009 年月平均气温的季节周期的 20 年平均值和标准差, 以及作为对照的 GHCN 观测资料的平均值和标准差。从图 5 可以看出, 本文采用的统计降尺度方法整体上表现良好, 抓住了哈萨克斯坦地区各站的月平均气温变化趋势, 但在不同站点上依然存在一定程度上的偏差。在 Atyrau、Aralskoe、Uil、Zharkent 站上降尺度模型存在明显的高估, 而在 Atbasar、Irtysk、Petropavlovsk 站则表现出了一定程度上的低估。此外, 各站预测的偏差幅度在各个月份也不同: 多数站点如 Zharkent、Atyrau、Irtysk 等站偏差幅度冬季明显大于夏季, 少数站点如 Aralskoe 站夏季大于冬季。最终用所选 11 站各站平均值 (见图 5f) 来看, 由于高低相抵, 预测结果和观测资料对比无论平均值还是标准差都合得很好。

对误差的统计及其时间分布和空间分布统计图 (如图 6、图 7 所示), 也可以反映出以上结果。

从图 6 所示的各统计量在各季节上的变化特征可以看到, 夏季 (JJA) 模拟结果最好, RMSE 的范围为 0.55~1.0, 中位数为 0.71, MAE 结果类似, 而冬季 (DJF) 最差, 比夏季显著偏高, RMSE 中位数达 1.94, 秋冬两季介于前二者之间。 $R^2$  的季节分布并不明显, 四季均在 0.6~0.8 左右浮动。这说明了模型对夏季月份的月平均气温的预测要优于冬季, 其可能的原因是: 哈萨克斯坦由于地处欧亚大陆中心位置, 属于大陆性气候, 夏季受季风影响小, 所以夏季月平均气温的年际变率不大, 回归效果较好, 如图 3a 所示夏季各月温度与纬向风场 ( $u_{925}$ 、 $u_{500}$ 、 $u_{200}$ ) 的相关系数虽然比经向风场相关系数大, 但是依然在 0.2 以下; 而在冬季, 由于受西伯利亚冷高压爆发、寒潮侵袭的影响, 冬季月平均气温的年际变率本身就很大, 以至于经过回归后预测结果表现出偏差较大, 冬季各月温度与经向风场 ( $v_{925}$ 、 $v_{500}$ 、 $v_{200}$ ) 相关系数大 (如图 3a 所示), 有甚者如 1 月预报因子  $v_{925}$  相关系数平方达到 0.41, 表示寒潮从哈萨克斯坦北部的西伯利亚地区携北风南侵而影响天气状况, 会造成年际变率大, 回归结果不准确。

对各月平均气温做全年平均之后, 可从空间分布上看出, 描述误差的统计量 RMSE、MAE 在各站之间的分布较为均匀, 变化幅度不大, 分别为  $0.4^{\circ}\text{C}$  和  $0.2^{\circ}\text{C}$  左右。此外, 从  $R^2$  的空间分布看, 除 Atyrau、Zharkent 两个站点  $R^2$  较低之外, 各个站点的  $R^2$  均在 0.7 左右, 表现出各站模拟效果基本良好。Atyrau 站位于里海之滨, 而由于本文中预报因子区域选取为距离站点最近的 9 个 NCEP 格网, 覆盖了里海上空的大尺度预报因子, 此外, Atyrau 站海拔低于海平面之下, 为  $-22\text{ m}$ , 而本文所选择的大尺度预报因子最低层为 925 hPa, 所以该站所在的低于海平面的海拔高度与预报因子的高度的差距比其它站点大, 进而造成了各预报因子代表性的偏差和回归结果的不准确。同理,  $R^2$  最低的 Zharkent 站位于天山脚下, 拥有所有选取的站点中的最高海拔, 达到 645 m, 同时其预报因子区域包含了海拔 5 km 以上的天山山脉, 如本文采用的低层大气的温度预报因子  $T_{925}$ , 在一个  $2.5^{\circ}$  (纬度)  $\times 2.5^{\circ}$  (经度) 的 NCEP 格点中的海拔落差就高达数千米, 那么此时的预报因子所代表一个格点中平均的气象要素

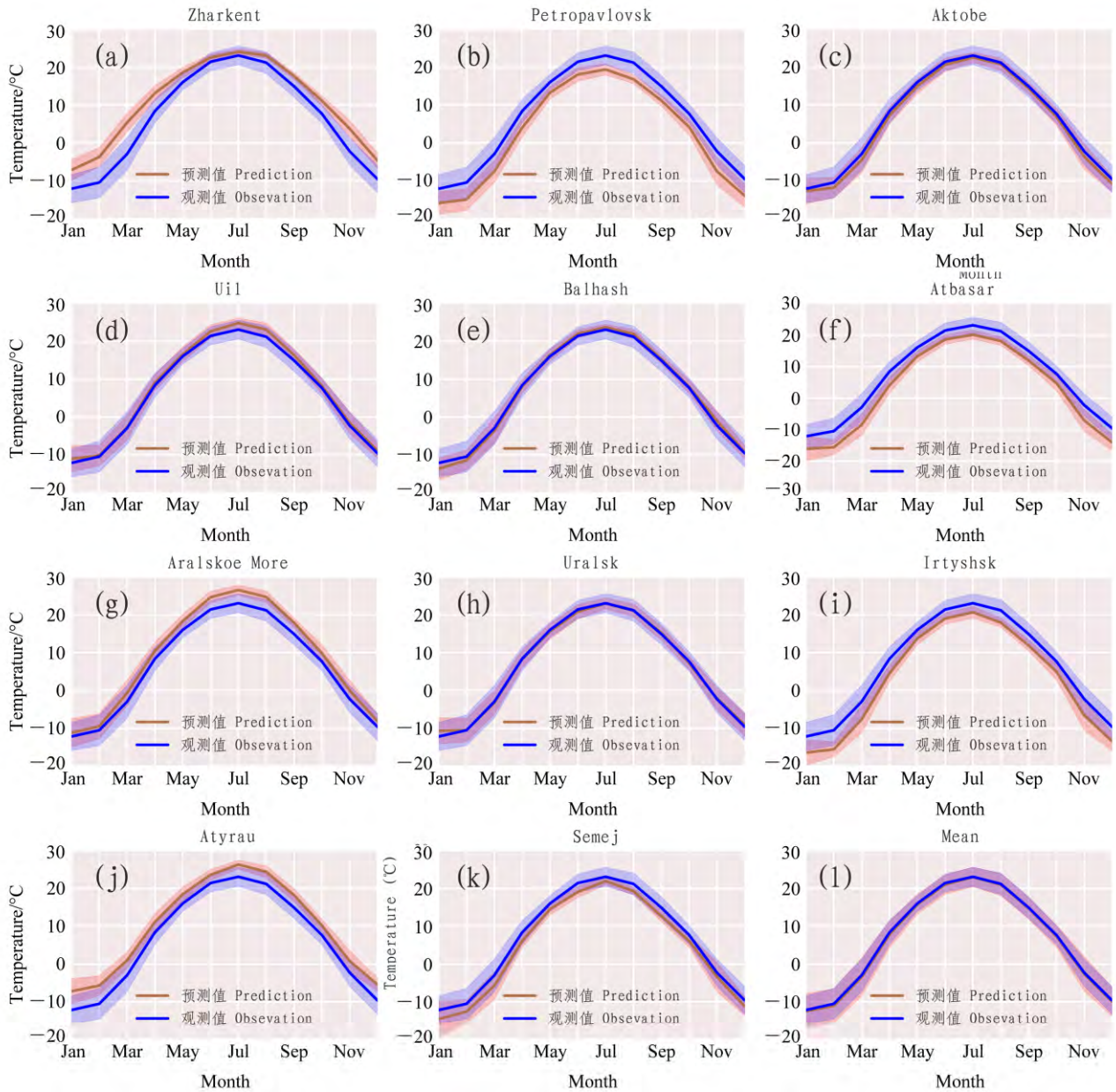


图5 1990~2009年各站及11站平均的多年平均气温季节变化（阴影表示标准差范围）

Fig. 5 Seasonal variations of mean temperature during 1990–2009 at the 11 stations in Kazakhstan and the seasonal variation of temperature averaged over all the stations (standard deviation is indicated with shading)

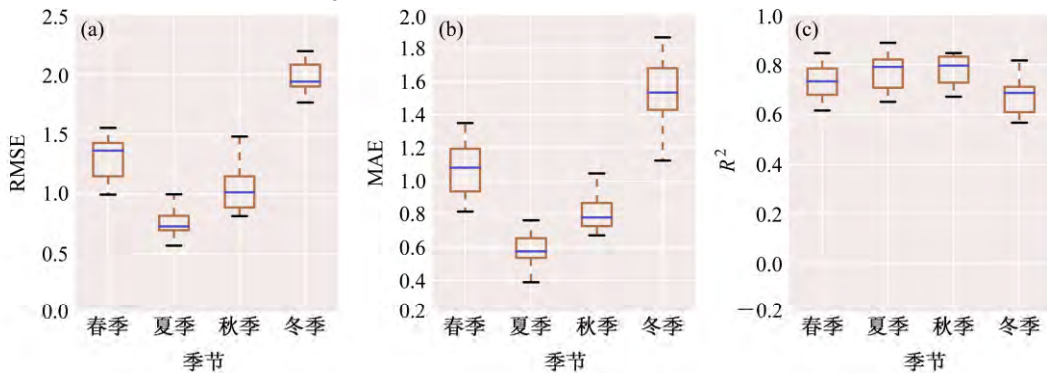


图6 模型评估统计量 (a) RMSE、(b) MAE、(c)  $R^2$  在春季 (MAM)、夏季 (JJA)、秋季 (SON)、冬季 (DJF) 的分布

Fig. 6 Box plots of performance statistics (a) RMSE, (b) MAE, and (c)  $R^2$  in the spring (MAM), summer (JJA), autumn (SON), and winter (DJF)

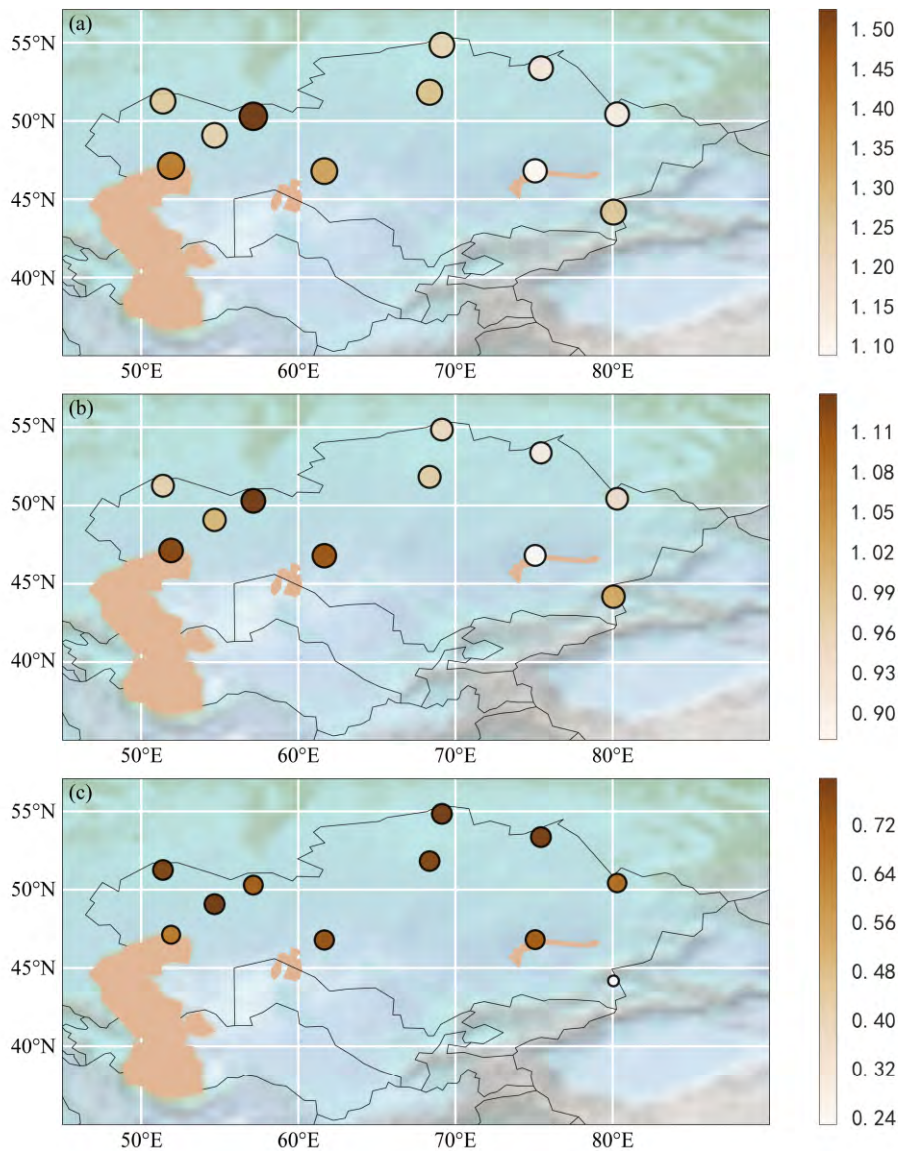


图7 1990~2009年各站 (a) RMSE、(b) MAE、(c)  $R^2$  的空间分布

Fig. 7 Distributions of (a) RMSE, (b) MAE, and (c)  $R^2$  during 1990–2009 at the 11 stations in Kazakhstan

就与这个复杂地形中的某一站点位置的气象要素差异很大, 所以这种空间上的复杂的地形由于尺度效应造成了降尺度模型预测结果的不准确。

#### 4 结论和讨论

哈萨克斯坦的自然环境及人类社会对气候变化敏感而脆弱, 本文探索和评估了使用岭回归方法对该地区月平均气温统计降尺度的能力。结果显示, 使用前 30 年数据和岭回归模型建立大尺度预报因子和观测资料的统计关系可以较好地预测后 20 年的月平均气温。具体来说: 预测能力在各月均

有不同程度的差异, 夏季各月模拟结果最好, RMSE 的范围为 0.55~1.0, 中位数为 0.71,  $R^2$  中位数接近 0.8, 冬季最差, RMSE 中位数达 1.94,  $R^2$  中位数为 0.7, 春秋两季介于前二者之间; 此外, 从误差的空间分布上看, 各站点的模拟效果略有不同但基本良好, 除个别站点外 (Atyrau、Zharkent),  $R^2$  均在 0.7 左右。

导致预测结果与对照的观测数据偏差的原因可能存在多种。首先, 为控制其它不确定性而仅讨论岭回归模型的模拟能力, 本文将预报因子的区域均设置为 9 个 NCEP 格点。然而, 根据 Wetterhall et al. (2006) 的研究结果, 预报因子区域的选择会对



PCA 的结果产生影响。因此, 一些针对主成分分析前预报因子区域的选定问题的创新方法 (Zhang and Yan, 2015), 可以在后续的哈萨克斯坦统计降尺度研究中加以验证。其次, 结果显示, 岭回归统计降尺度方法对具有复杂地形的区域的模拟能力存在较大缺陷。例如, 模型预测结果最差即  $R^2$  最低的站点为海拔最低的 Atyrau 站和海拔最高的 Zharkent 站, 并且此二者的大尺度预报因子区域中都包含了差异巨大的地表环境 (里海、天山)。那么后续研究中就可以针对复杂地形区域进行进一步的研究, 比如考虑对近年来涌现的基于机器学习的方法 (Goyal et al., 2012; Lu and Qin, 2014), 并在哈萨克斯坦这一区域加以验证。再次, 数据的选择上也存在着一定的不确定性。例如训练期长短对预报结果可能造成影响, 本文中的 30 年训练期是参考了该地区共 50 年可用的气象资料而决定的, 其它长度的训练期 (如 10 年、20 年) 会对结果造成什么影响? 再或是不同再分析数据集 (如 ERA40 等), 可能为结果带来哪些改变? 这些都是值得在进一步的研究中深入探讨的。

最后需要说明的是, 本文对预报因子仅进行了粗略的选择, 没有根据各个站点的特点或对预报因子的分析进行详细的筛选或剔除。岭回归模型因为比普通最小二乘多元线性回归多引入正则项, 避免了出现不重要的且相互相关的预报因子时模型发生过拟合现象而造成预测的不准确。本文的结果验证了这种在选择预报因子过程中不含人工干预的方法适合于如哈萨克斯坦这类研究较少、研究人员或许并不十分熟悉的地区进行统计降尺度。此外, 类似于哈萨克斯坦, 许多地区由于没有足够气象观测站点, 留给统计降尺度的挑战是没有足够的质量达到要求的资料来训练统计模型、建立统计关系。因此, 发展出气候观测资料相对匮乏的地区也能够使用的统计降尺度方法, 无疑对于应对气候变化问题是有前景、有意义的, 也将是下一步的研究重点。

### 参考文献 (References)

- Benestad R E, Chen D L, Mezghani A, et al. 2015. On using principal components to represent stations in empirical-statistical downscaling [J]. *Tellus (Ser. A: Dynamic Meteorology and Oceanography)*, 67: 28326, doi: 10.3402/tellusa.v67.28326.
- Bobojonov I, Aw-Hassan A. 2014. Impacts of climate change on farm income security in Central Asia: An integrated modeling approach [J]. *Agriculture, Ecosystems and Environment*, 188: 245–255, doi: 10.1016/j.agee.2014.02.033.
- de Beurs K M, Henebry G M. 2004. Land surface phenology, climatic variation, and institutional change: Analyzing agricultural land cover change in Kazakhstan [J]. *Remote Sensing of Environment*, 89(4): 497–509, doi: 10.1016/j.rse.2003.11.006.
- De Sales F, Xue Y. 2013. Dynamic downscaling of 22-year CFS winter seasonal hindcasts with the UCLA-ETA regional climate model over the United States [J]. *Climate Dyn.*, 41 (2): 255–275, doi: 10.1007/s00382-012-1567-x.
- Eisfelder C, Klein I, Niklaus M, et al. 2014. Net primary productivity in Kazakhstan, its spatio-temporal patterns and relation to meteorological variables [J]. *Journal of Arid Environments*, 103: 17–30, doi: 10.1016/j.jaridenv.2013.12.005.
- 范丽军, 符淙斌, 陈德亮. 2007. 统计降尺度法对华北地区未来区域气温变化情景的预估 [J]. *大气科学*, 31 (5): 887–897. Fan Lijun, Fu Congbin, Chen Deliang. 2007. Estimation of local temperature change scenarios in North China using statistical downscaling method [J]. *Chinese Journal of Atmospheric Sciences (in Chinese)*, 31 (5): 887–897, doi: 10.3878/j.issn.1006-9895.2007.05.12.
- Flato G, Marotzke J, Abiodun B, et al. 2013. Evaluation of climate models [M]//*Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Stocker T F, Qin D, Plattner G K, et al., Eds. Cambridge, UK and New York, NY, USA: Cambridge University Press, 741–866.
- Fowler H J, Blenkinsop S, Tebaldi C. 2007. Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modeling [J]. *International Journal of Climatology*, 27 (12): 1547–1578, doi: 10.1002/joc.1556.
- Giorgi F. 2006. Climate change hot-spots [J]. *Geophys. Res. Lett.*, 33 (8): L8707, doi: 10.1029/2006GL025734.
- Goyal M K, Ojha C S P. 2011. Evaluation of linear regression methods as downscaling tools in temperature projections over the Pichola Lake basin in India [J]. *Hydrological Processes*, 25 (9): 1453–1465, doi: 10.1002/hyp.7911.
- Goyal M K, Burn D H, Ojha C S P. 2012. Evaluation of machine learning tools as a statistical downscaling tool: Temperatures projections for multi-stations for Thames River basin, Canada [J]. *Theor. Appl. Climatol.*, 108 (3–4): 519–534, doi: 10.1007/s00704-011-0546-1.
- Hanssen-Bauer I, Achberger C, Benestad R E, et al. 2005. Statistical downscaling of climate scenarios over Scandinavia [J]. *Climate Research*, 29 (3): 255–268, doi: 10.3354/cr029255.
- Hewitson B C, Crane R G. 1996. Climate downscaling: Techniques and application [J]. *Climate Research*, 7: 85–95, doi: 10.3354/cr007085.
- Hu Z Y, Zhang C, Hu Q, et al. 2014. Temperature changes in Central Asia from 1979 to 2011 based on multiple datasets [J]. *J. Climate*, 27 (3): 1143–1167, doi: 10.1175/JCLI-D-13-00064.1.
- Kalnay E, Kanamitsu M, Kistler R, et al. 1996. The NCEP/NCAR 40-year reanalysis project [J]. *Bull. Amer. Meteor. Soc.*, 77 (3): 437–471, doi: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.
- Kazmi D H, Li J P, Rasul G, et al. 2015. Statistical downscaling and future

- scenario generation of temperatures for Pakistan region [J]. *Theor. Appl. Climatol.*, 120 (1-2): 341-350, doi: 10.1007/s00704-014-1176-1.
- Laprise R. 2008. Regional climate modeling [J]. *J. Comput. Phys.*, 227 (7): 3641-3666, doi: 10.1016/j.jcp.2006.10.024.
- 李同昇, 龙冬平. 2014. 中亚国家地缘位置与中国地缘战略的若干思考 [J]. *地理科学进展*, 33 (3): 303-314. Li Tongsheng, Long Dongping. 2014. Geopolitical position of Central Asian countries and geopolitical strategy of China [J]. *Progress in Geography (in Chinese)*, 33 (3): 303-314.
- Lioubimtseva E, Henebry G M. 2009. Climate and environmental change in arid Central Asia: Impacts, vulnerability, and adaptations [J]. *Journal of Arid Environments*, 73 (11): 963-977, doi: 10.1016/j.jaridenv.2009.04.022.
- 龙爱华, 邓铭江, 李湘权, 等. 2010. 哈萨克斯坦水资源及其开发利用 [J]. *地球科学进展*, 25 (12): 1357-1366. Long Aihua, Deng Mingjiang, Li Xiangquan, et al. 2010. An analysis of water resources development in Kazakhstan and water problems [J]. *Advance in Earth Science (in Chinese)*, 25 (12): 1357-1366.
- Lu Y, Qin X S. 2014. A coupled  $K$ -nearest neighbour and Bayesian neural network model for daily rainfall downscaling [J]. *International Journal of Climatology*, 34 (11): 3221-3236, doi: 10.1002/joc.3906.
- Malsy M, Aus Der Beek T, Flörke M. 2015. Evaluation of large-scale precipitation data sets for water resources modelling in Central Asia [J]. *Environmental Earth Sciences*, 73 (2): 787-799, doi: 10.1007/s12665-014-3107-y.
- Mannig B, Müller M, Starke E, et al. 2013. Dynamical downscaling of climate change in Central Asia [J]. *Global and Planetary Change*, 110: 26-39, doi: 10.1016/j.gloplacha.2013.05.008.
- Orskaugo E, Scheel I, Frigessi A, et al. 2011. Evaluation of a dynamic downscaling of precipitation over the Norwegian mainland [J]. *Tellus (Ser. A: Dynamic Meteorology and Oceanography)*, 63 (4): 746-756, doi: 10.1111/j.1600-0870.2011.00525.x.
- Pearson K. 1896. Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia [J]. *Philosophical Transactions of the Royal Society (Ser. A)*, 187: 253-318, doi: 10.1098/rsta.1896.0007.
- Qi J G, Bobushev T S, Kulmatov R, et al. 2012. Addressing global change challenges for Central Asian socio-ecosystems [J]. *Frontiers of Earth Science*, 6 (2): 115-121, doi: 10.1007/s11707-012-0320-4.
- Sommer R, Glazirina M, Yuldashev T, et al. 2013. Impact of climate change on wheat productivity in Central Asia [J]. *Agriculture, Ecosystems and Environment*, 178: 78-99, doi: 10.1016/j.agee.2013.06.011.
- 王树舟, 于恩涛. 2014. 基于 MIROC/WRF 嵌套模式的中国气候变化预估 [J]. *气候与环境研究*, 19 (01): 11-22. Wang Suzhou, Yu Entao. 2014. Projected climate changes over China based on the nested MIROC/WRF model [J]. *Climatic and Environmental Research (in Chinese)*, 19 (1): 11-22.
- Wetterhall F, Bárdossy A, Chen D L, et al. 2006. Daily precipitation-downscaling techniques in three Chinese regions [J]. *Water Resour. Res.*, 42(11): W11423, doi:10.1029/2005WR004573.
- Wetterhall F, Bárdossy A, Chen D, et al. 2009. Statistical downscaling of daily precipitation over Sweden using GCM output [J]. *Theor. Appl. Climatol.*, 96 (1-2): 95-103, doi: 10.1007/s00704-008-0038-0.
- Wilby R L, Dawson C W, Barrow E M. 2002. SDSM — A decision support tool for the assessment of regional climate change impacts [J]. *Environmental Modelling and Software*, 17 (2): 145-157, doi: 10.1016/S1364-8152(01)00060-3.
- Wilby R L, Charles S P, Zorita E, et al. 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods [C]// Supporting material of the Intergovernmental Panel on Climate Change (IPCC), prepared on behalf of Task Group on Data and Scenario Support for Impacts and Climate Analysis (TGICA). Retrieved from the Intergovernmental Panel on Climate Change website: [http://www.ipcc-data.org/guidelines/dgm\\_no2\\_v1\\_09\\_2004.pdf](http://www.ipcc-data.org/guidelines/dgm_no2_v1_09_2004.pdf) [2016-4-28].
- Wright C K, de Beurs K M, Akhmedieva Z K, et al. 2009. Reanalysis data underestimate significant changes in growing season weather in Kazakhstan [J]. *Environmental Research Letters*, 4 (4): 45020, doi: 10.1088/1748-9326/4/4/045020.
- Zhang X L, Yan X D. 2015. A new statistical precipitation downscaling method with Bayesian model averaging: A case study in China [J]. *Climate Dyn.*, 45 (9-10): 2541-2555, doi: 10.1007/s00382-015-2491-7.