On the Use of "Inflation" in Statistical Downscaling

HANS VON STORCH

Journal of Climate, 12/1999, 卷 12, 期 12

Abstract: The technique of "inflating" in downscaling, which makes the downscaled climate variable have the right variance, is based on the assumption that all local variability can be traced back to large-scale variability. For practical situations this assumption is not valid, and inflation is an inappropriate technique. Instead, additive, randomized approaches should be adopted.

1. Statistical downscaling

$$\underset{\text{predictand}}{\mathbf{y}} = \underset{\text{predictor}}{\mathcal{F}(\mathbf{x})} \qquad (1)$$

y—> small-scale, not adequately described in GCMs.
[weather variables or climatic statistics]
eg. daily temperatures, monthly precipitation amounts
x—> large-scale features, well resolved.
[characteristics of the circulation]
eg. temperature

绝大多数应用中，方程 $\mathcal{F}$ 都是线性的。

$$\mathcal{F}(\mathbf{x}) = \alpha\mathbf{x} + \epsilon \qquad (2)$$

在大多数应用中，上述公式会被简化为以下形式以应用。

$$\hat{\mathbf{y}} = \alpha\mathbf{x} \qquad (3)$$

the downscaled values $\hat{y}$ have smaller variance than the local values y.
From $\mathrm{var}(\hat{y}) = \alpha^2 \cdot \mathrm{var}(x)$ and $\mathrm{var}(y) = \alpha^2 \cdot \mathrm{var}(x) + \sigma^2$, follows $\mathrm{var}(\hat{y}) < \mathrm{var}(y)$.

2. Inflation and randomization

许多研究中, 时间序列 $\tilde{y}$ 需要满足$\mathrm{var}(\tilde{y}) = \mathrm{var}(y)$. 为了达到这一条件，这就需要对 $\hat{y}$ 进行放大：

$$\tilde{\mathbf{y}} = \beta\hat{\mathbf{y}} = \sqrt{\frac{\mathrm{var}(\mathbf{y})}{\mathrm{var}(\hat{\mathbf{y}})}} \cdot \hat{\mathbf{y}} \qquad (4)$$

然而，这种方法是毫无意义的。
首先, 它暗示着y中的所有变化都将与x中的变化相关这一假设，显然这不是进行该操作的初衷。
其次，扩大后的变量($\tilde{y}$)的均方误差比原始的估计变量($\hat{y}$)大:
假设$\mathrm{var}(x) = \mathrm{var}(y) = 1$, 那么$\mathrm{var}(\hat{y}) = \alpha^2$, $\beta=1/\alpha$.
那么$\mathrm{var}(\hat{y} - y) = \sigma^2$, 而$\mathrm{var}(\tilde{y} - y) = (1 - \alpha^2) + \sigma^2 > \sigma^2$.
既然如此，何不使用随机添加噪音的方法(randomization)代替上述这种扩大(inflation)方法，

$$\mathbf{y}^* = \hat{\mathbf{y}} + \text{noise} \qquad (5)$$

这也同样满足最初的原始公式（1）。

"Inflation"和"randomization"最大的区别在于，前者导致$\tilde{y}$的时空变化完全由x控制，而后者得到的y^*的时空变化只受大尺度x变量的部分控制。

方法（5）已经成功实现在对于山崩(landslides)的研究中[Buma and Dehn 1999; Dehn and Buma 1999].

2020/3/3

Report

2020.3.8

張慕琪

Methods

实验设计：

检查——

1. historical阶段，全国所有点，ANN得到的tmax和tmin与GMFD相差较大的问题；

地区：China

时间：Train & validation

数据：CCSM, GMFD, ANN(before BC)

变量：

·Temperature:

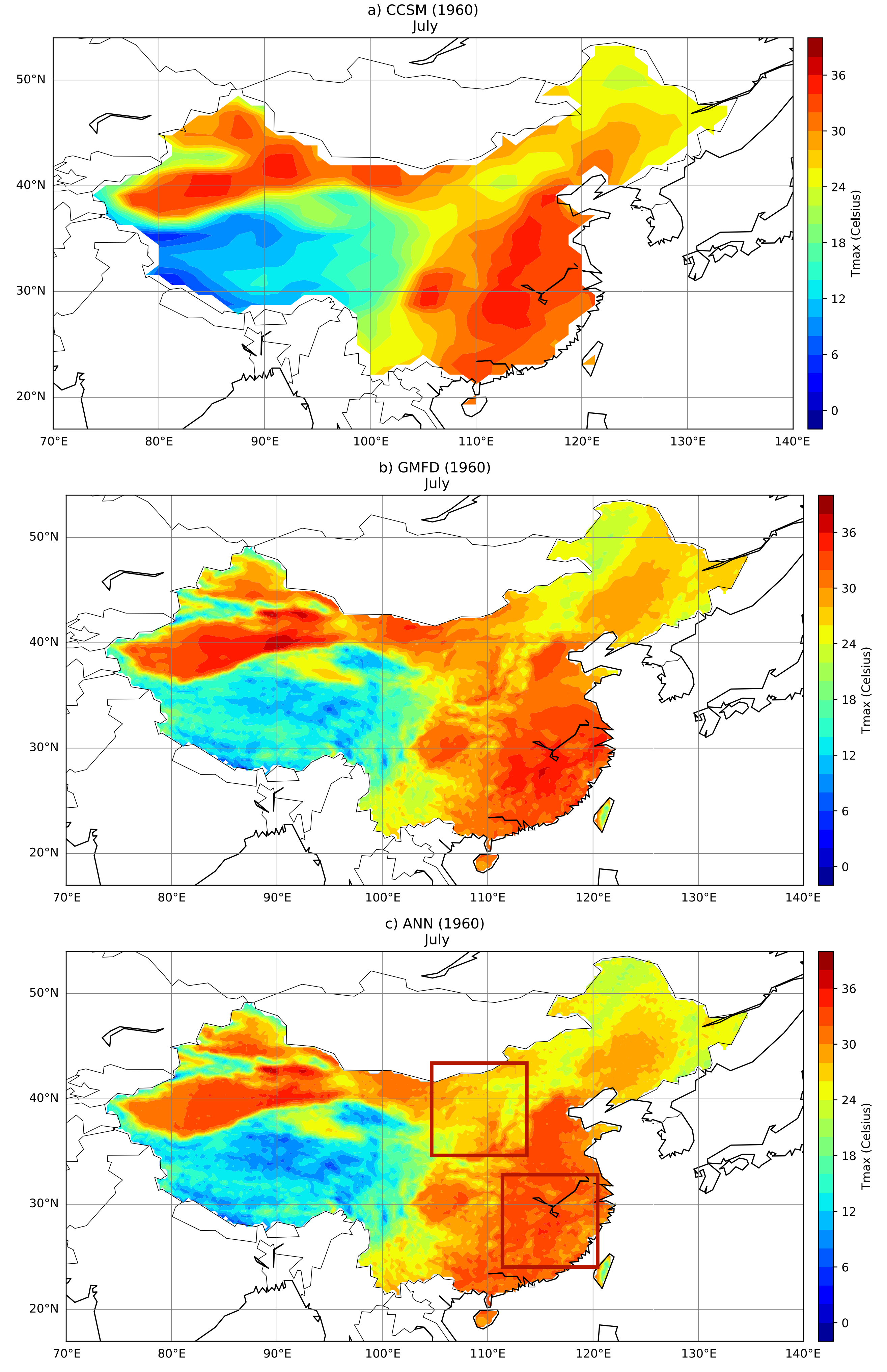
1. Maps: yr=1960, mon=1/7;
2. Histogram: yr=1960, mon=1/7;
3. r2_score: train & validation / tmax & tmin.

Results

Maps

China

Tmax

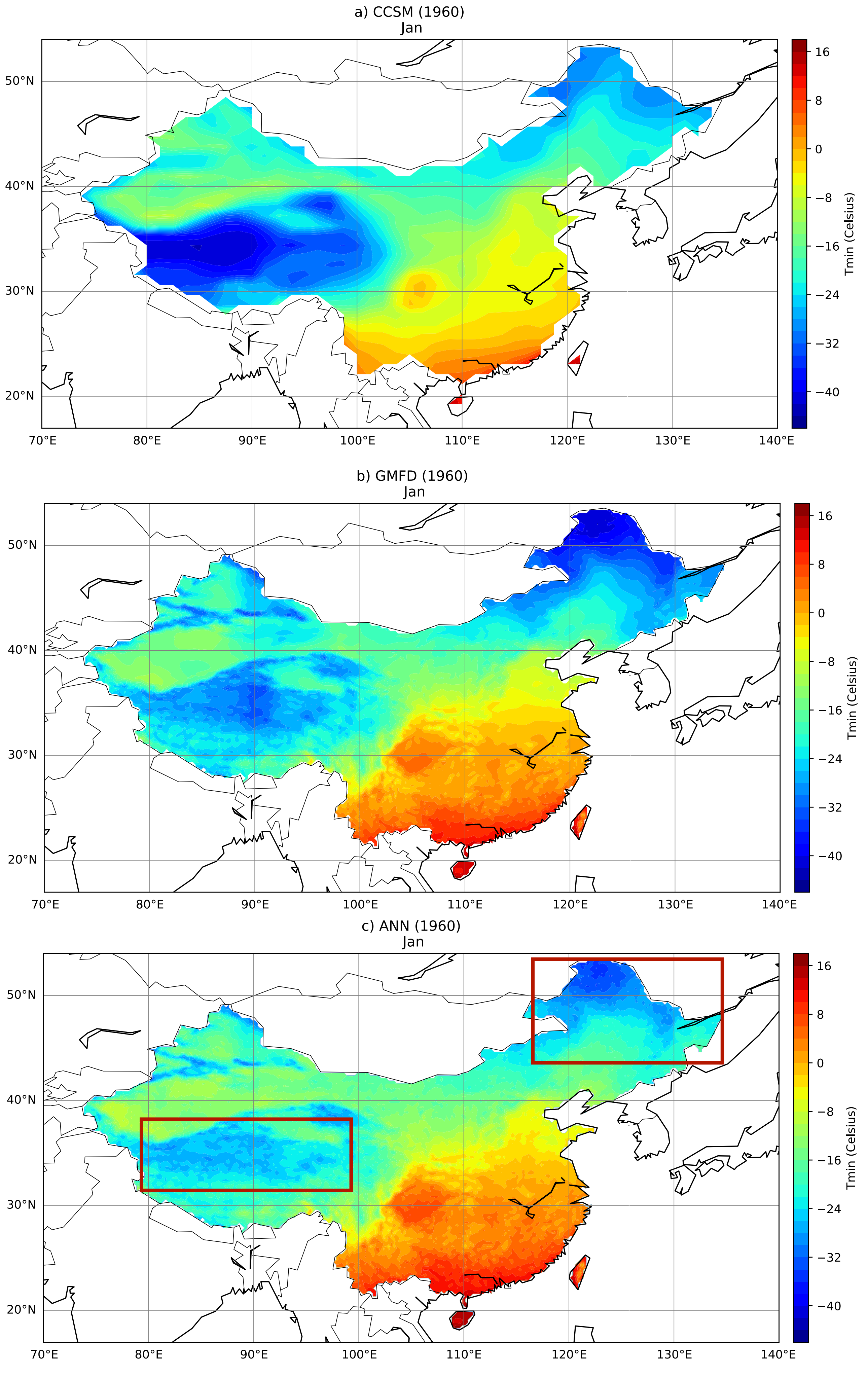


Results

Maps

China

Tmin

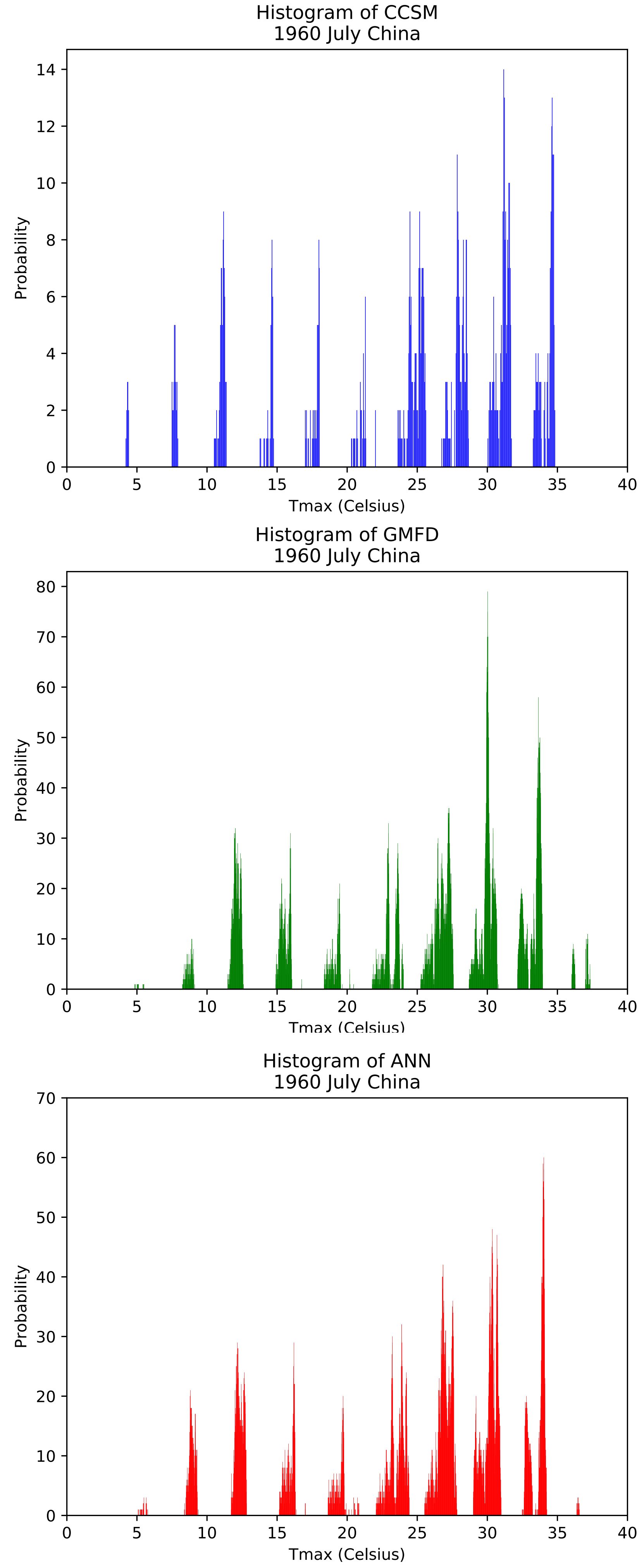


Results

Histogram

China

Tmax

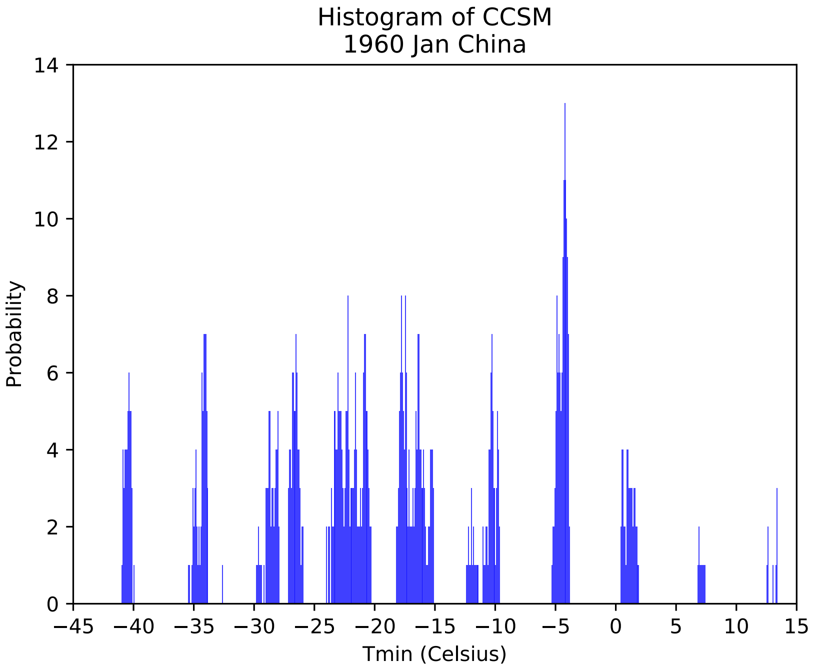


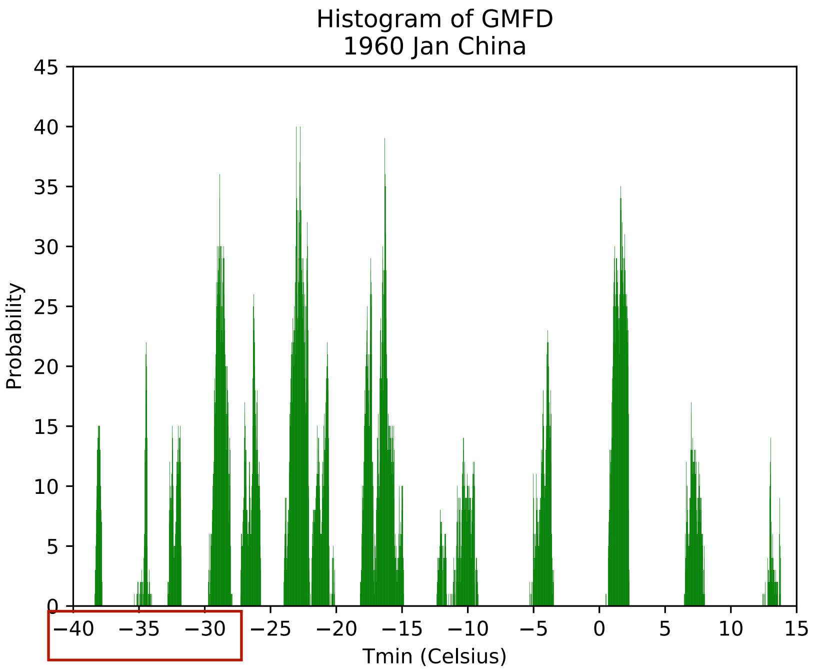
Results

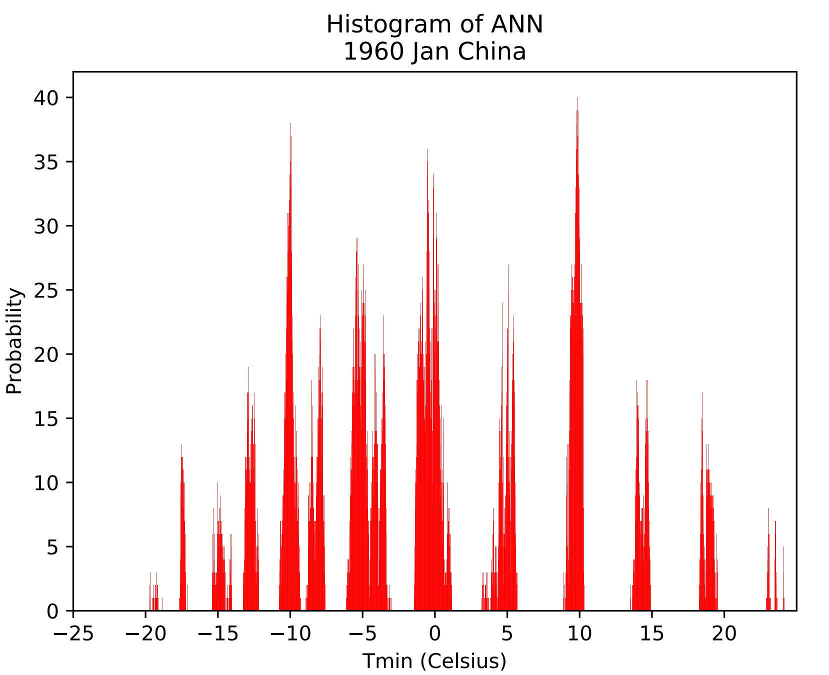
Histogram

China

Tmin







Results

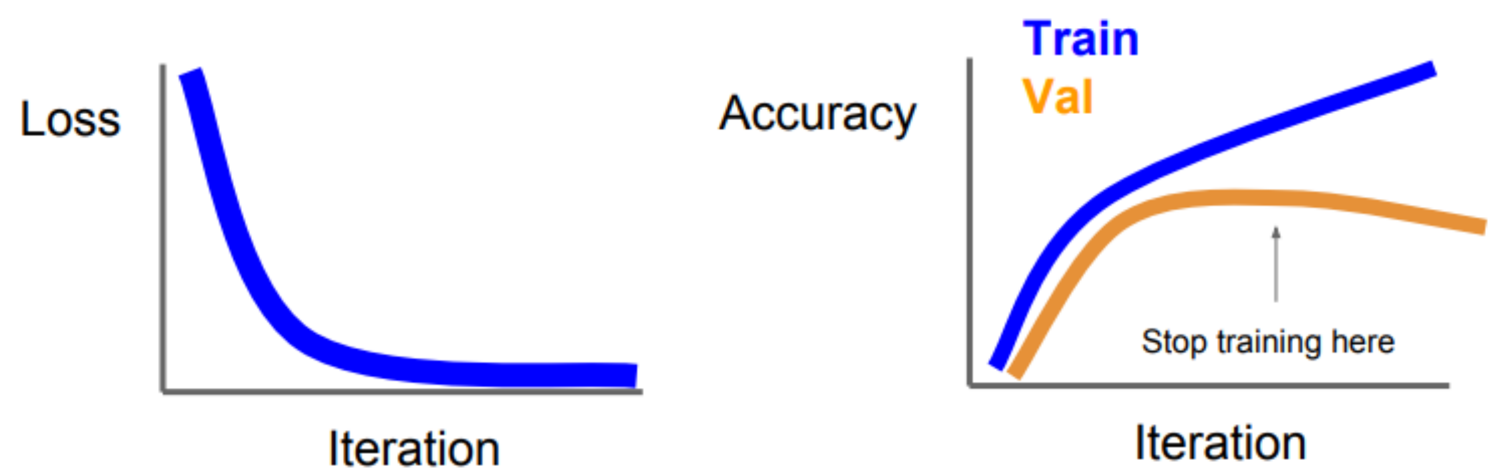
China

r2_score

并非过拟合；
针对Beijing的网络并非适用于全国所有地区。

```
train_tmax = 0.827
test_tmax = 0.828

train_tmin = 0.850
test_tmin = 0.851
```



Stop training the model when accuracy on the validation set decreases
Or train for a long time, but always keep track of the model snapshot that worked best on val

```
ppchina = 15275

train_tmax = 12624.834488990175
test_tmax =  12642.153059600247

train_tmin = 12986.240692143008
test_tmin =  12993.529028121588
```

谢谢