WeeklyNote

2019.10.27

張慕琪

Received: 30 July 2018 | Revised: 2 January 2019 | Accepted: 4 January 2019 | Published on: 25 January 2019

DOI: 10.1002/joc.5990

International Journal of Climatology | RMetS

RESEARCH ARTICLE

A hierarchical analysis of the impact of methodological decisions on statistical downscaling of daily precipitation and air temperatures

Sara C. Pryor[1] | Justin T. Schoof[2]

本文既有对于温度设计的模型又有针对降水设计的模型，在此仅关注降水部分

Introduction

研究变量：Tmin, Tmax, PoP(Probability of precipitaion), Amount of precipitation on a wet day.

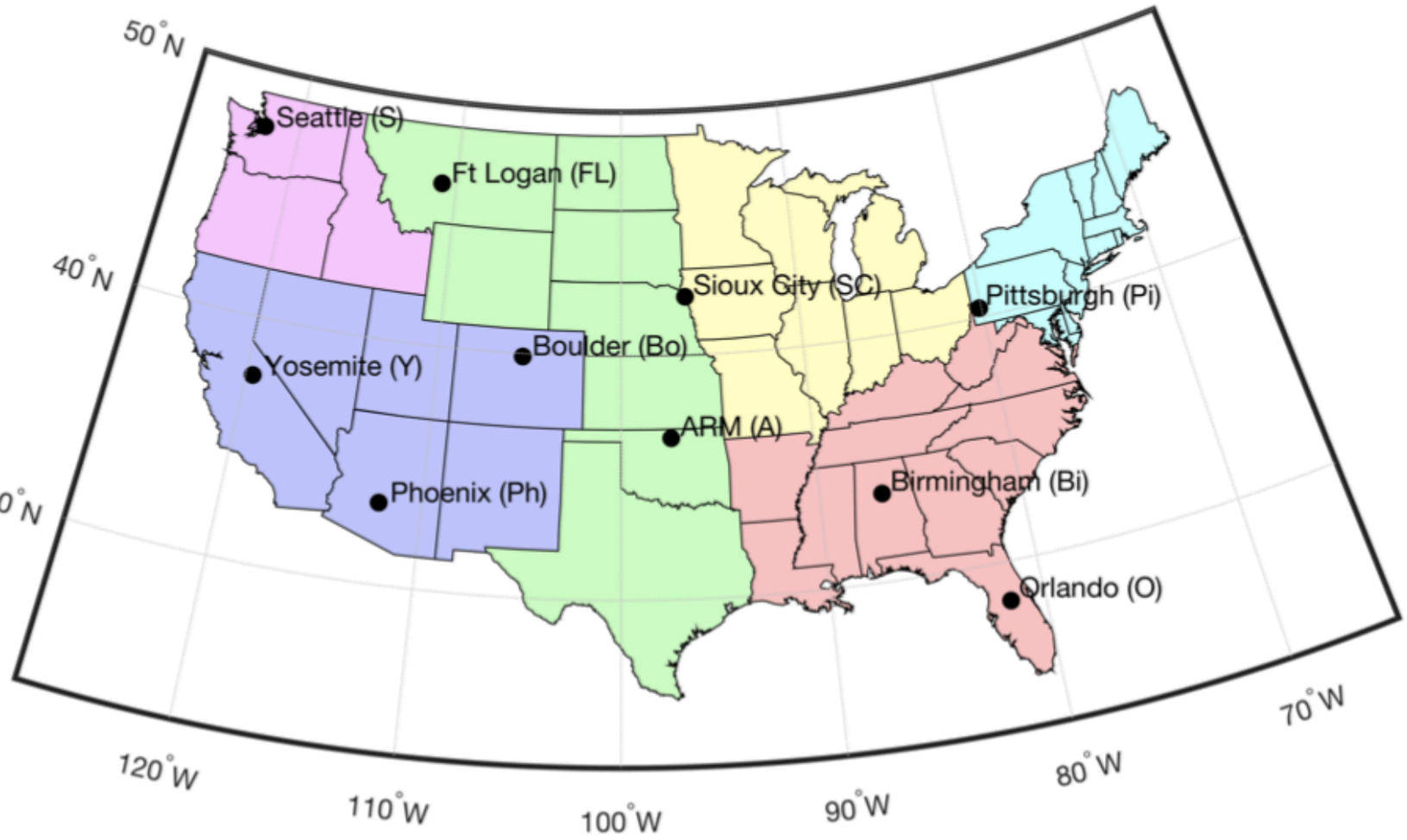
涉及模型：Generalized linear models[GLMs] and Artificial neural networks[ANNs].

建立一个更普适的框架，在其中可以自动且客观的选择特定站点的预测变量（因此站点之间可能会有所不同）；优化模型受到季节循环的影响。

Introduction

研究区域：涵盖美国不同气候区域的10个地区

数据来源：预测值(Tmin, Tmax, precipitation occurrence and amount on a wet day)来自Liveh data set(Livneh et al., 2013).



预测数据：Geopotential height at 500 hPa (Z500)，Air temperature at 700 hPa (T700)，Specific humidity at 700 hPa (Q700)，Air temperature at 500 hPa (T500)，Specific humidity at 500 hPa (Q500)，West–East (u-component) wind speed at 700 hPa (U700)，South–North (v-component) wind speed at 700 hPa (V700)

Precipitation Downscaling

RCMs的数据总是产生“经常下雨”的情况，且由于方法的系统偏差RCMs会低估降水时间的强度；

研究发现ANN对于预报PoP表现较差；

研究发现，方法的偏差是由潮湿天气定义的阈值决定的，因此本文设定三种阈值：>0 mm/day, >0.1 mm/day and >1 mm/day;

模型设计：三种阈值

输入变量3个/7个/使用PCA方法确定具体输入几个

GLM使用Poisson或gamma分布 VS. ANN

P.S. ANN有三层隐藏层

Results

OR = H/(1-H)
H = a/(a+c)
a: 事件发生且预报正确的个数;
c: 事件发生但未预报

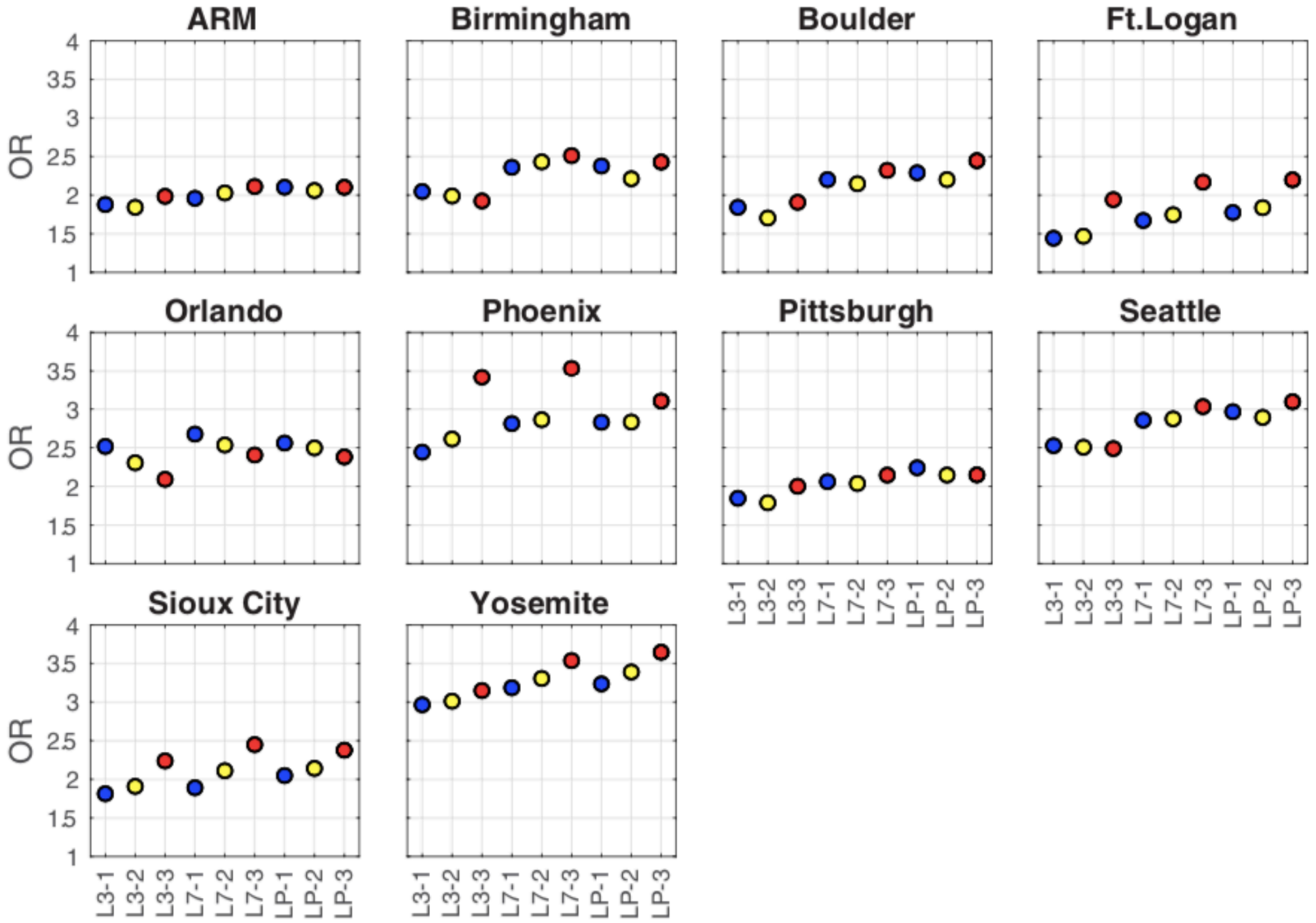


FIGURE 8 The ORs for the PoP at the 10 stations for the nine ESD models applied to the independent test data. Model abbreviations are as follows: L3 denotes logistic regression using grid cell values of Z500, T700 and Q700. L7 denotes stepwise logistic regression with L1 regularization wherein the predictors are Z500, T700, Q700, T500, Q500, U700, V700. LP indicates stepwise logistic regression with L1 regularization with PC scores as predictors. −1 indicates a wet-day threshold of >0 mm/day (red), −2 indicates a threshold of 0.1 mm/day (yellow) and 3 is used for a threshold of >1 mm/day (red) [Colour figure can be viewed at wileyonlinelibrary.com]

补充

ROC curve

Receiver operating characteristic curve

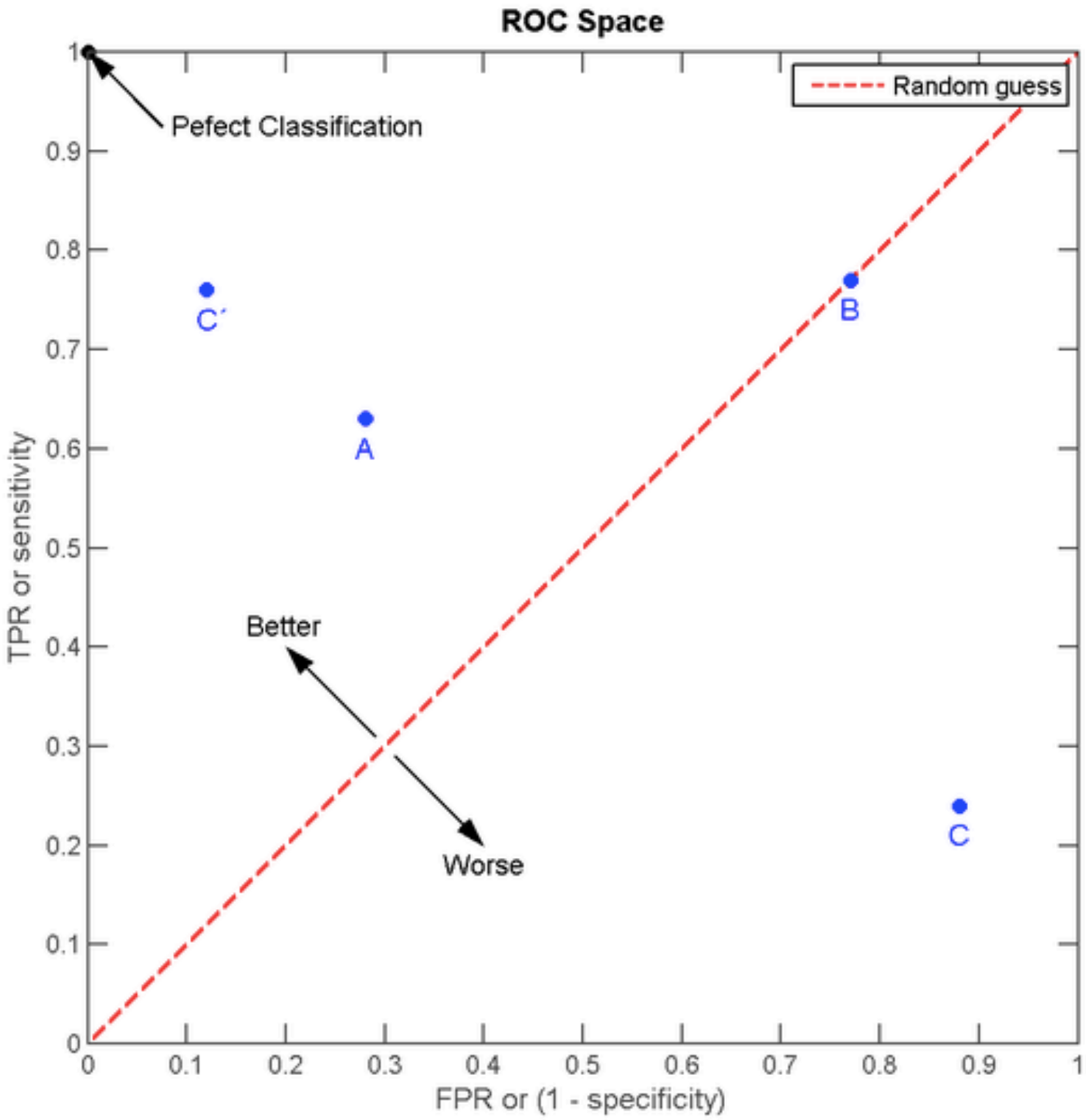
Let us look into four prediction results from 100 positive and 100 negative instances:

A	B	C	C′
TP=63 FP=28 91 FN=37 TN=72 109 100 100 200	TP=77 FP=77 154 FN=23 TN=23 46 100 100 200	TP=24 FP=88 112 FN=76 TN=12 88 100 100 200	TP=76 FP=12 88 FN=24 TN=88 112 100 100 200
TPR = 0.63	TPR = 0.77	TPR = 0.24	TPR = 0.76
FPR = 0.28	FPR = 0.77	FPR = 0.88	FPR = 0.12
PPV = 0.69	PPV = 0.50	PPV = 0.21	PPV = 0.86
F1 = 0.66	F1 = 0.61	F1 = 0.23	F1 = 0.81
ACC = 0.68	ACC = 0.50	ACC = 0.18	ACC = 0.82

补充

ROC curve

Receiver operating characteristic curve



Results

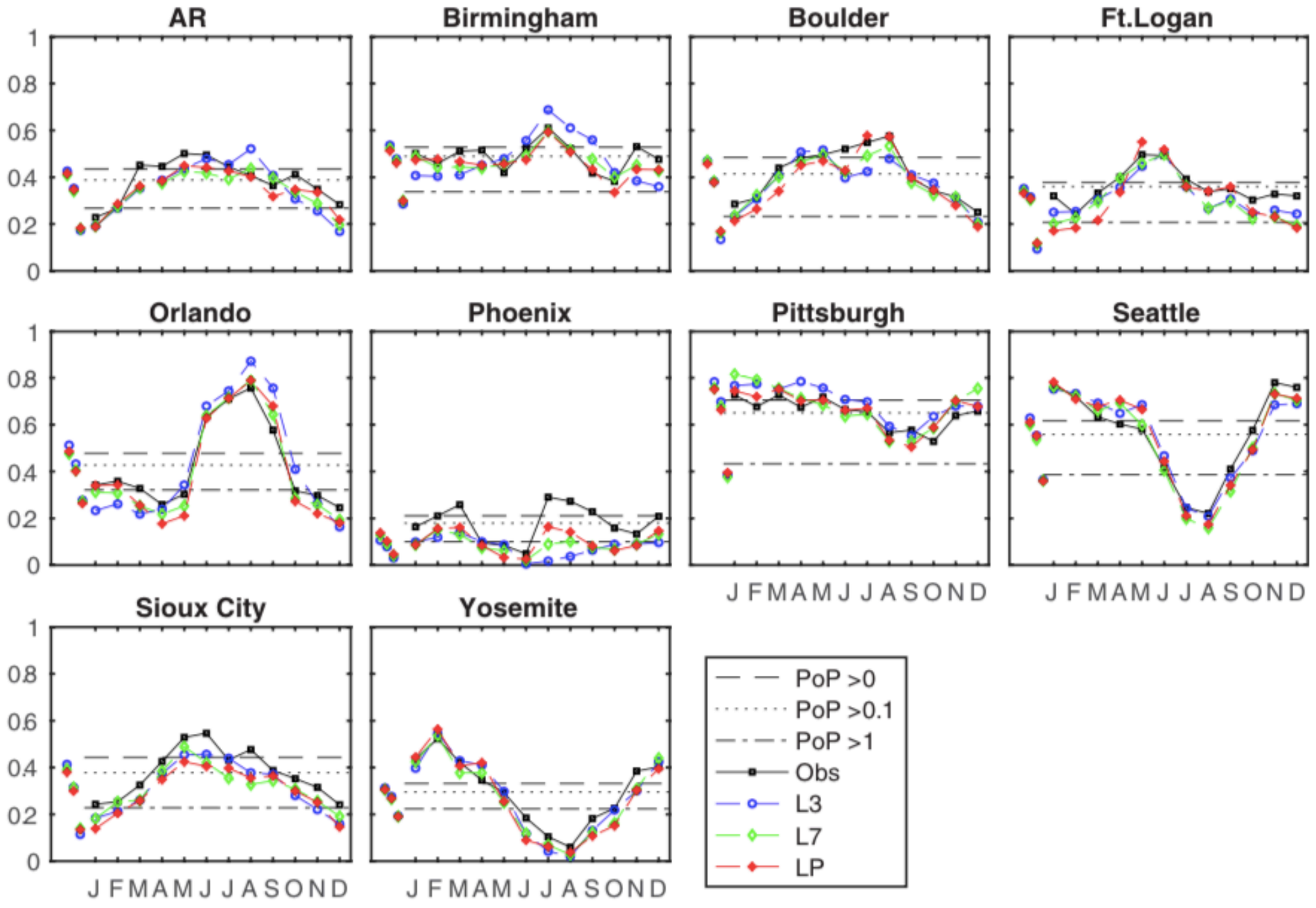


FIGURE 9 PoP in the test data from the 10 stations. The annual average PoP derived from the observations for the three thresholds is shown by the dashed horizontal lines, while equivalent estimates from the three different ESD models are indicated by the symbols close to *y*-axis. The monthly mean PoP from the observations and each of the models for a threshold for a wet day of >0.1 mm/day are shown by the lines with symbols on them. Model abbreviations are as in Figure 8 [Colour figure can be viewed at wileyonlinelibrary.com]

Results

Poisson分布比Gamma分布较好一些，但两者都没有ANN表现好

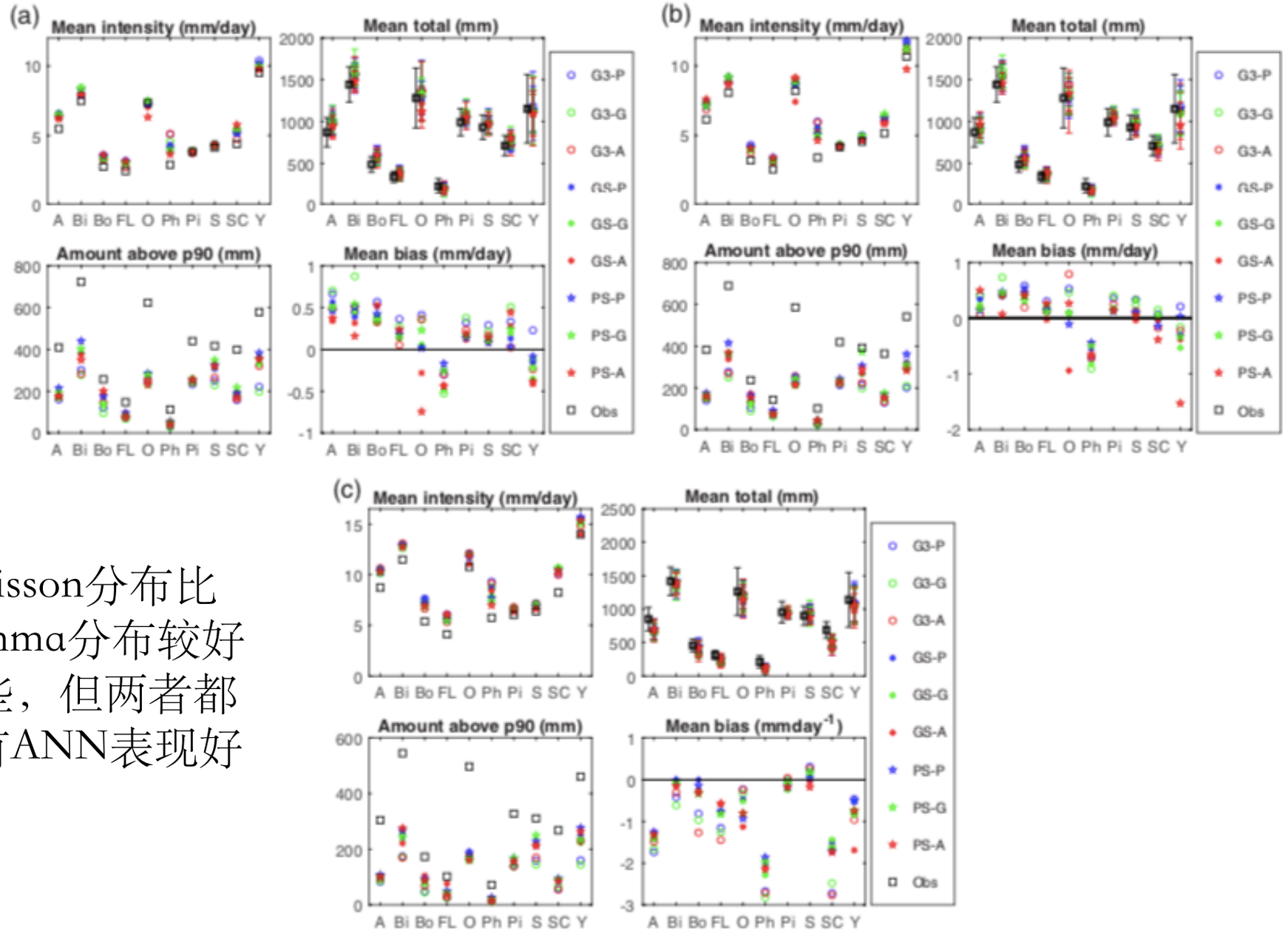


FIGURE 11 Results for the 10 stations of four aspects of the precipitation climate; mean intensity on a wet day, MB on a wet day, mean annual total (and interannual variability) and the amount (in mm) of precipitation above the 90th percentile value. Note in the panel showing annual total precipitation and interannual variability therein, the observed values are slightly displaced on the horizontal axis to aid legibility. Frame (a) shows results for a precipitation threshold of 0 mm/day, (b) for 0.1 mm/day and (c) for 1 mm/day for a wet day. Nine transfer functions (ESD models) are shown the first two letters denote the predictors; G3 = 3 predictors are used (grid cell values of Z500, T700 and Q700), GS = 7 gird cell predictors, and PS denotes models built using the PC scores as predictors. The final letter denotes the transfer function form; P is Poisson for the link function in the logistic regression, G is for gamma for the link function in the logistic regression, and A is ANN [Colour figure can be viewed at wileyonlinelibrary.com]

Conclusions

1. 无论哪种回归模型对于PoP的预测总是不尽人意；
2. 本文使用的数据划分是奇偶年特征，有人曾提出如果使用干旱年进行训练，非干旱年进行测试的新方法，但这种方法更多的是侧重于捕捉气候的内部变化而非辐射平衡导致的气候变化；
3. 非线性模型（ANN）要比GLM更具有技巧性，表现更好。

谢谢