Seasonal Prediction of Winter Extreme Precipitation over Canada by Support Vector Regression

Zhen Zeng[1], William Hsieh[1], Amir Shabbar[2] and William Burrows[2]

[1]University of British Columbia, Vancouver, Canada

[2]Environment Canada

Abstract

For forecasting the maximum 5-day accumulated precipitation over the winter season at lead times of 3, 6, 9 and 12 months over Canada from 1950 to 2007, two nonlinear and two linear regression models were used, where the models were support vector regression (SVR) (nonlinear and linear versions), nonlinear Bayesian neural network (BNN) and multiple linear regression (MLR).

Both SVR models tended to forecast better than MLR & BNN, & the nonlinear SVR model tended to forecast slightly better than the linear SVR. The eastern Prairies region displayed the highest forecast skills & Arctic, the second highest. Nonlinearity was strongest over the eastern Prairies & weakest over the Arctic.

Introduction

Machine learning methods originate from artificial intelligence. The 1st wave of breakthrough came with neural network (NN) methods in the mid-1980s. The 2nd wave came with kernel methods in the mid-1990s.

NN methods are widely used for nonlinear regression, nonlinear classification, etc. However, they suffer from local minima during nonlinear optimization.

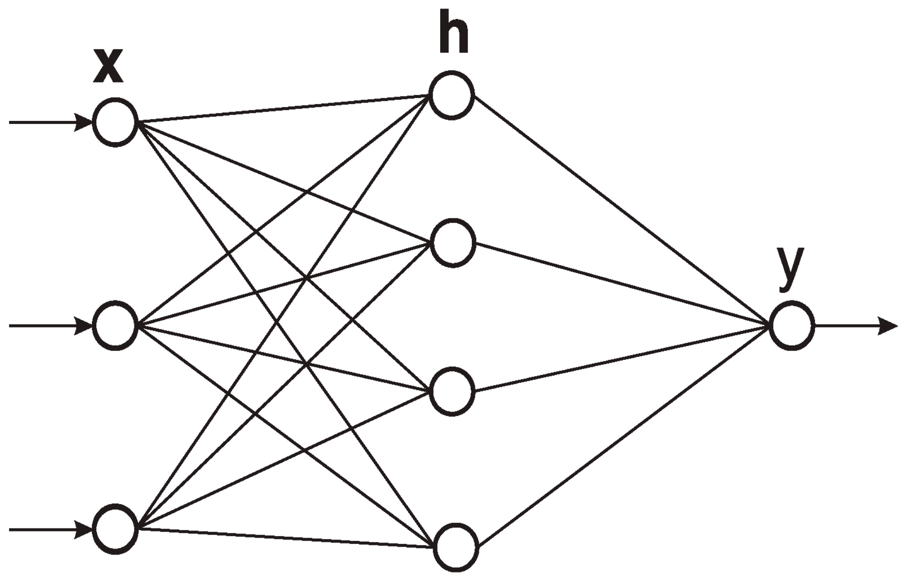
Kernel methods such as support vector regression improved on NN methods in 2 ways: (1) They do not have local minima during optimization, and (2) they use an error norm robust to outliers in the data (unlike the mean squared error norm used in NN).

Models

Linear regression (LR): $y = \sum_i a_i x_i + a_0$

Neural networks (NN):

Uses adaptive basis fns h_j



$$y = \sum_j a_j \, h_j(\mathbf{x}; \mathbf{w}) + a_0$$

$$= \sum_j a_j \overbrace{\tanh\left(\sum_i w_{ij} x_i + w_{0j}\right)} + a_0$$

Kernel methods

Non-adaptive basis fns.: $$y = \sum_j a_j \phi_j(\mathbf{x}) + a_0$$

Adv.: *linear* optimization, hence no local minima.
Disadv.: Very large no. of basis functions needed.

If optimization problem can involve only dot products like $\phi^\top(\mathbf{x}')\phi(\mathbf{x})$

and the dot product is given by a kernel function *K*: $$\phi^\top(\mathbf{x}')\phi(\mathbf{x}) = K(\mathbf{x}', \mathbf{x})$$

then there is no need to work with the high dimensional ϕ_j . It can be shown that

$$y = \sum_{k=1}^{n} \alpha_k K(\mathbf{x}_k, \mathbf{x}) + \alpha_0$$

Common kernel: Gaussian, i.e. radial basis function (RBF) kernel

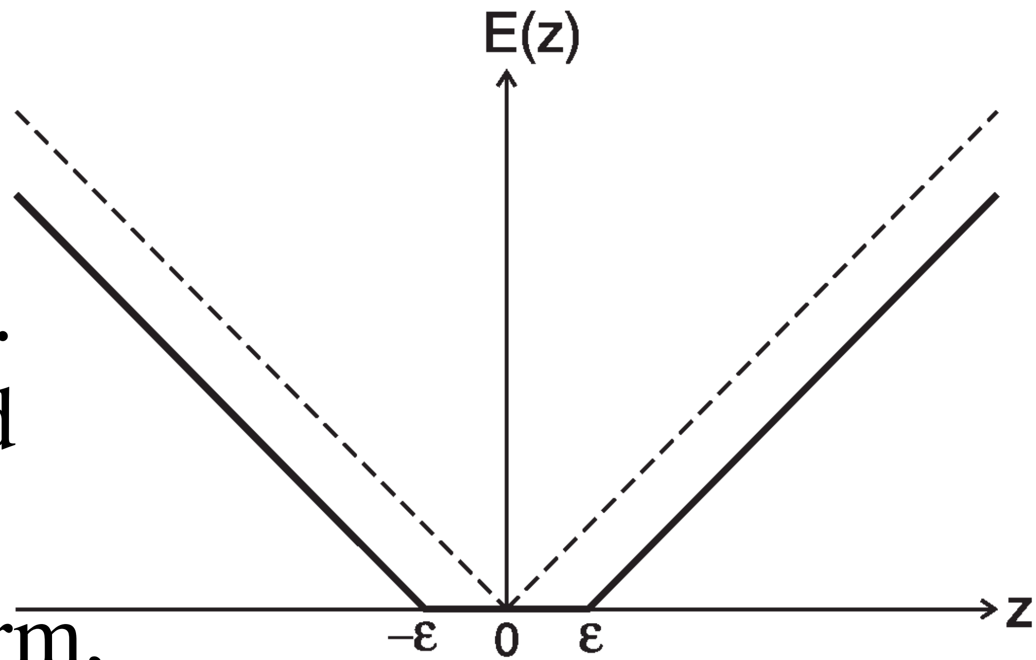
$$K(\mathbf{x}_k, \mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_k)\|^2}{2\sigma^2}\right)$$

Common kernel method:

Support vector machines for regression (SVR)

$z = y - y_{obs}$

Robust error norm $E(z)$. For comparison, dashed lines indicate the mean absolute error norm, which is also robust to outliers.



Data

Precipitation (prcp): 5-day total prcp for 118 stations over Canada (1950-2007). Predict max. 5-day prcp over the winter season (Dec.-Feb.).

Predictors:

Sea sfc.temperature (SST) (30°S-70°N) (NOAA ERSST3)

500 hPa geopotential ht. (Z500) (20°N-90°N) (NCEP)

Extended EOF (space-time PCA) applied to the SST anomalies and to the Z500 anomalies. 5 principal components (PC) retained for each field.

6 climate indices (CPC): Nino3.4 SST, NAO, PNA, PDO, SCA (Scandinavia pattern), EA (Eastern Atlantic).

Total: 5+5+6 = 16 predictors.

Cluster analysis of the prcp data yielded 6 regions:

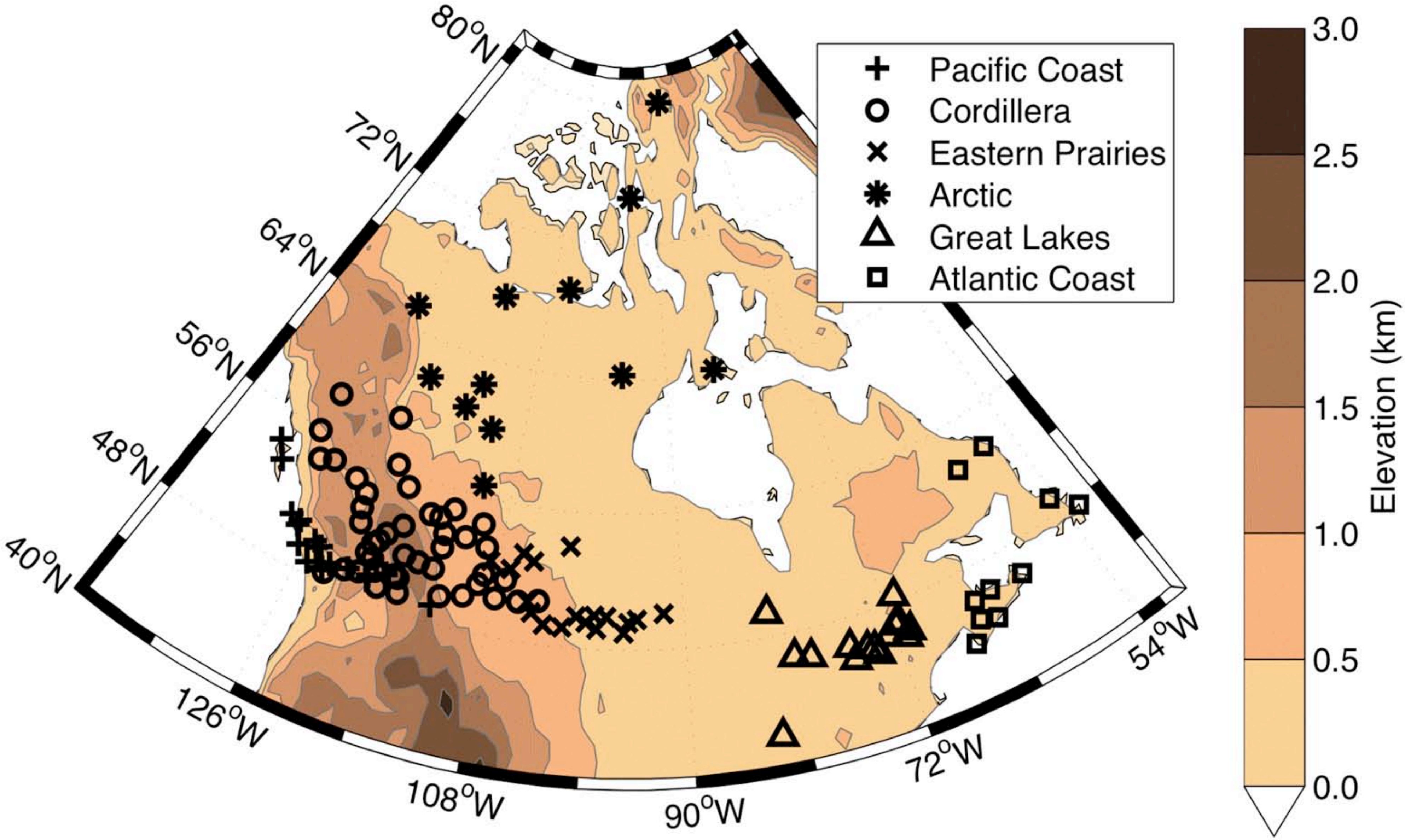


Fig.1

Forecast results

Two rounds of cross validation (CV1 and CV2) were used.

In CV1, the middle 3 years of a 5-year validation segment were reserved as independent data to test model forecasts, while for the training data, CV2 is used to determine the optimal hyperparameters and optimal number of predictor and predictand PCs to use.

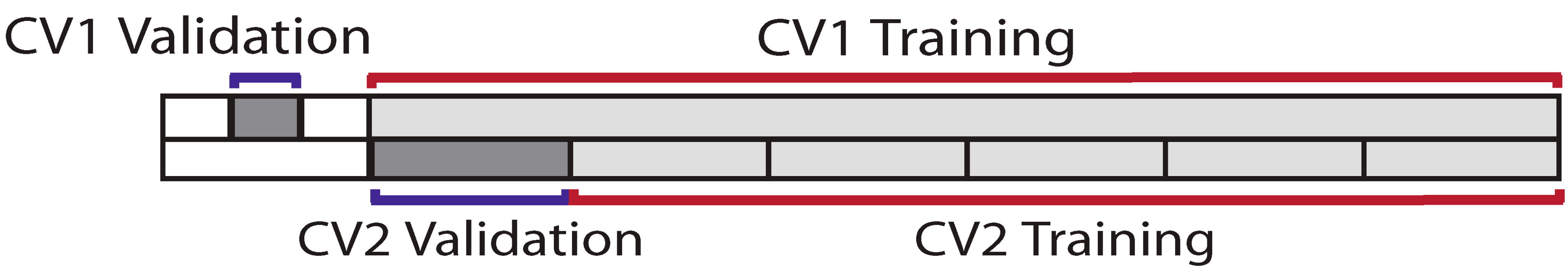


Fig.2

Forecast scores

(1) Correlation

(2) Willmott's Index of Agreement IOA

P = predicted, *O* = observed

overbar = mean,

perfect score = 1

$$\mathrm{IOA} = 1 - \frac{\sum_{i=1}^{N} (P_i - O_i)^2}{\sum_{i=1}^{N} \left(\left| P_i - \bar{O} \right| + \left| O_i - \bar{O} \right| \right)^2},$$

(3) Mean absolute error skill score (MAESS) (relative to climatology forecasts): perfect score = 1.

(4) Skill_v = Standard deviation of forecasts / Std. of observations

4 models compared: **MLR** (multiple linear regression), **SVR-L** (support vector regression with **Linear** kernel), **SVR-R** (SVR with **RBF** nonlinear kernel) and **BNN** (Bayesian NN) at forecast lead times of 3, 6, 9, 12 months.

Figures 3-8 show forecast scores for the 6 regions.
On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually as "o".

In general, SVR-R performs best. Both SVR-L and MLR are linear models, however, SVR-L uses a robust error norm and MLR, a non-robust one.

Results most impressive in Region 3 (eastern Prairies). Arctic (Region 4) appears to only have linear relations.

Region 1: Pacific coast

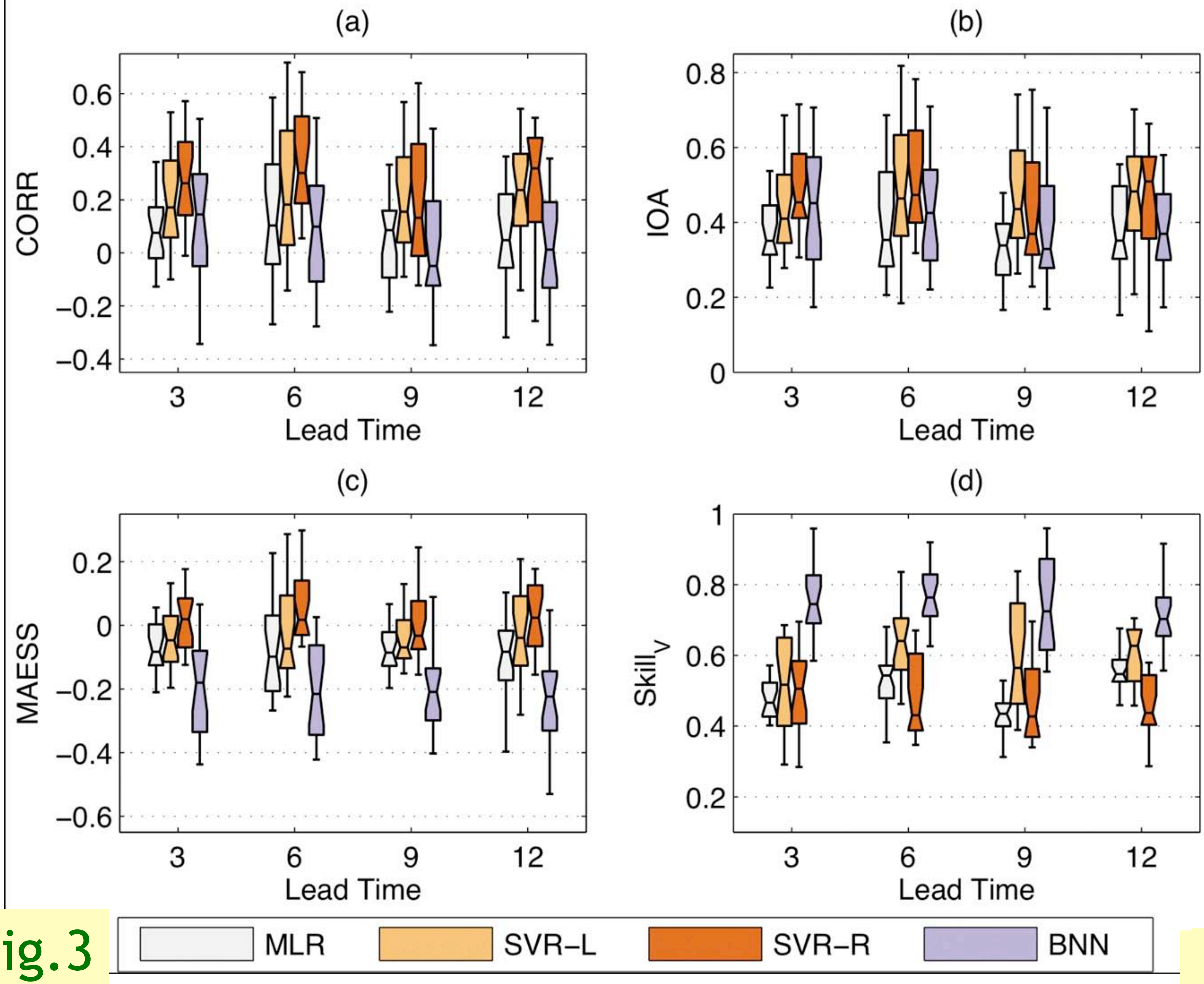


Fig.3

Region 2: Cordillera

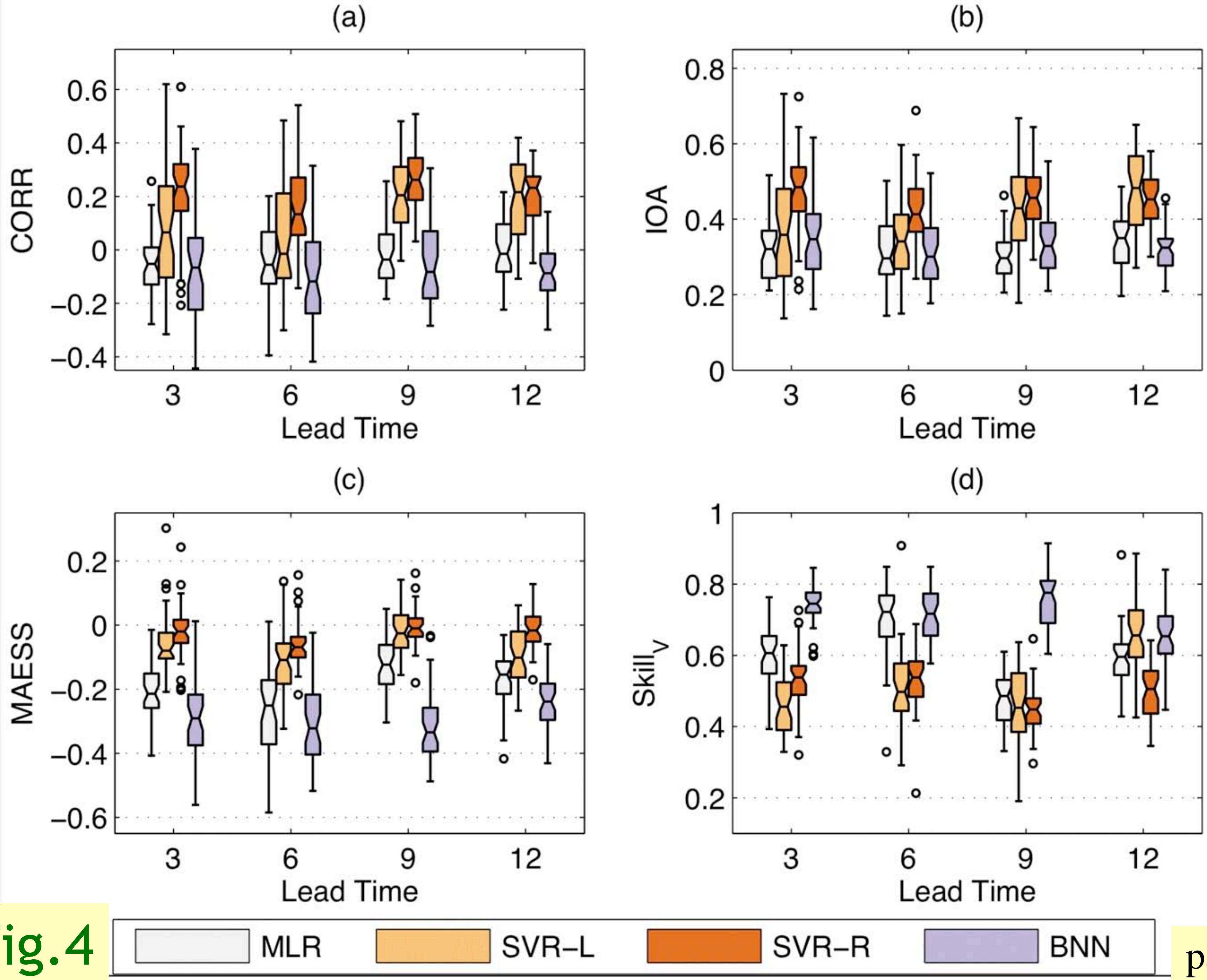


Fig.4

Region 3: Eastern Prairies

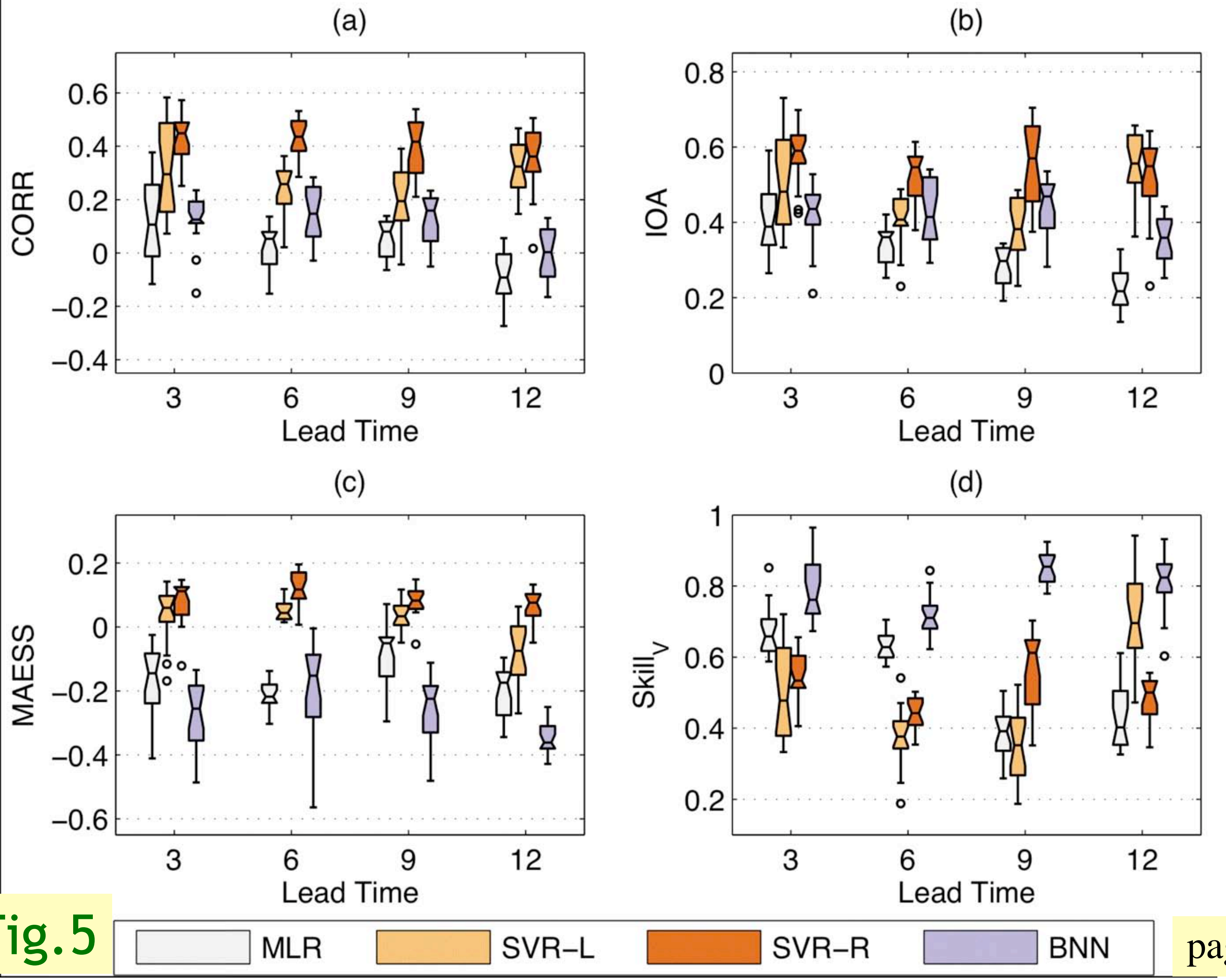


Fig.5

Region 4: Arctic

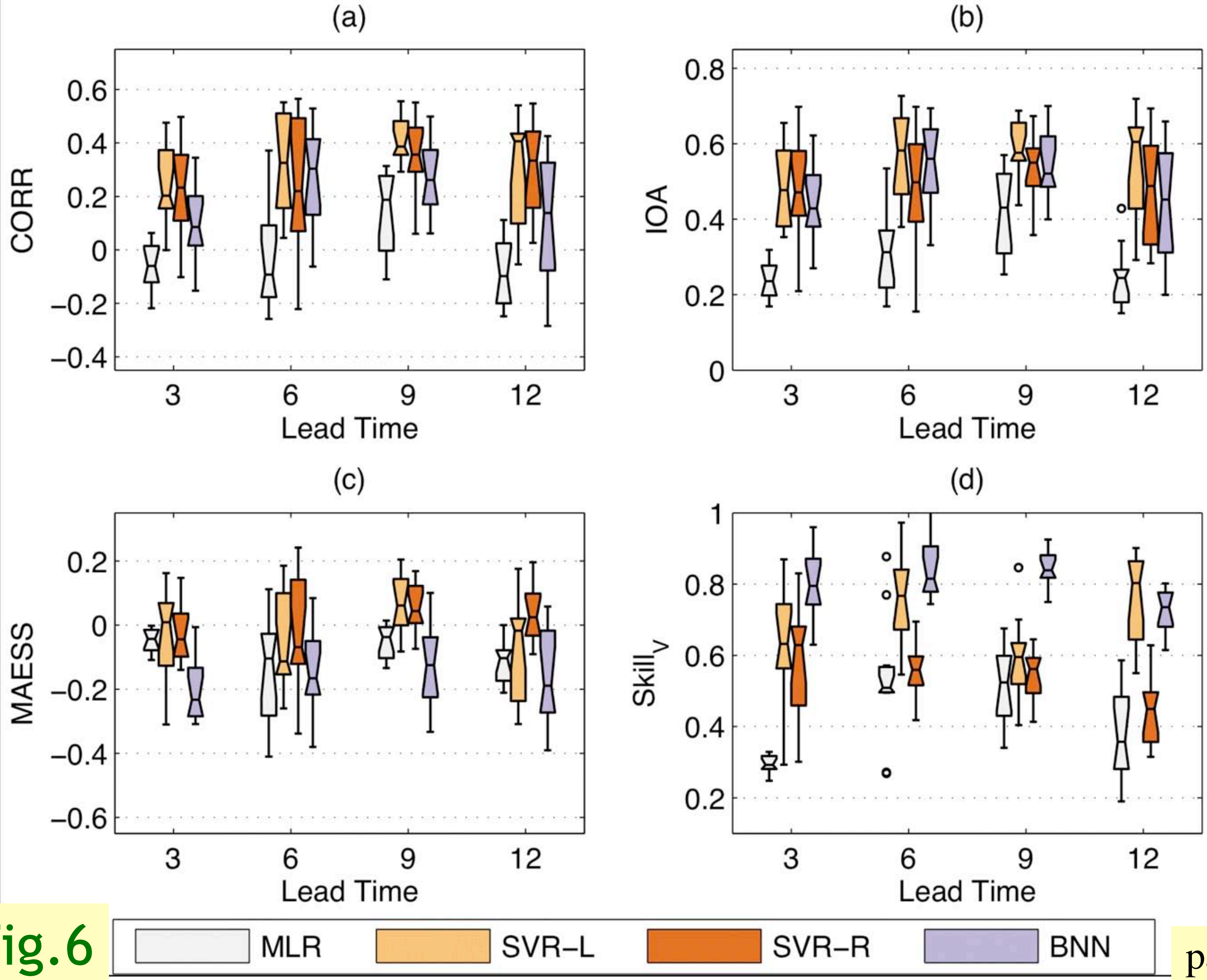


Fig.6

Region 5: Great Lakes

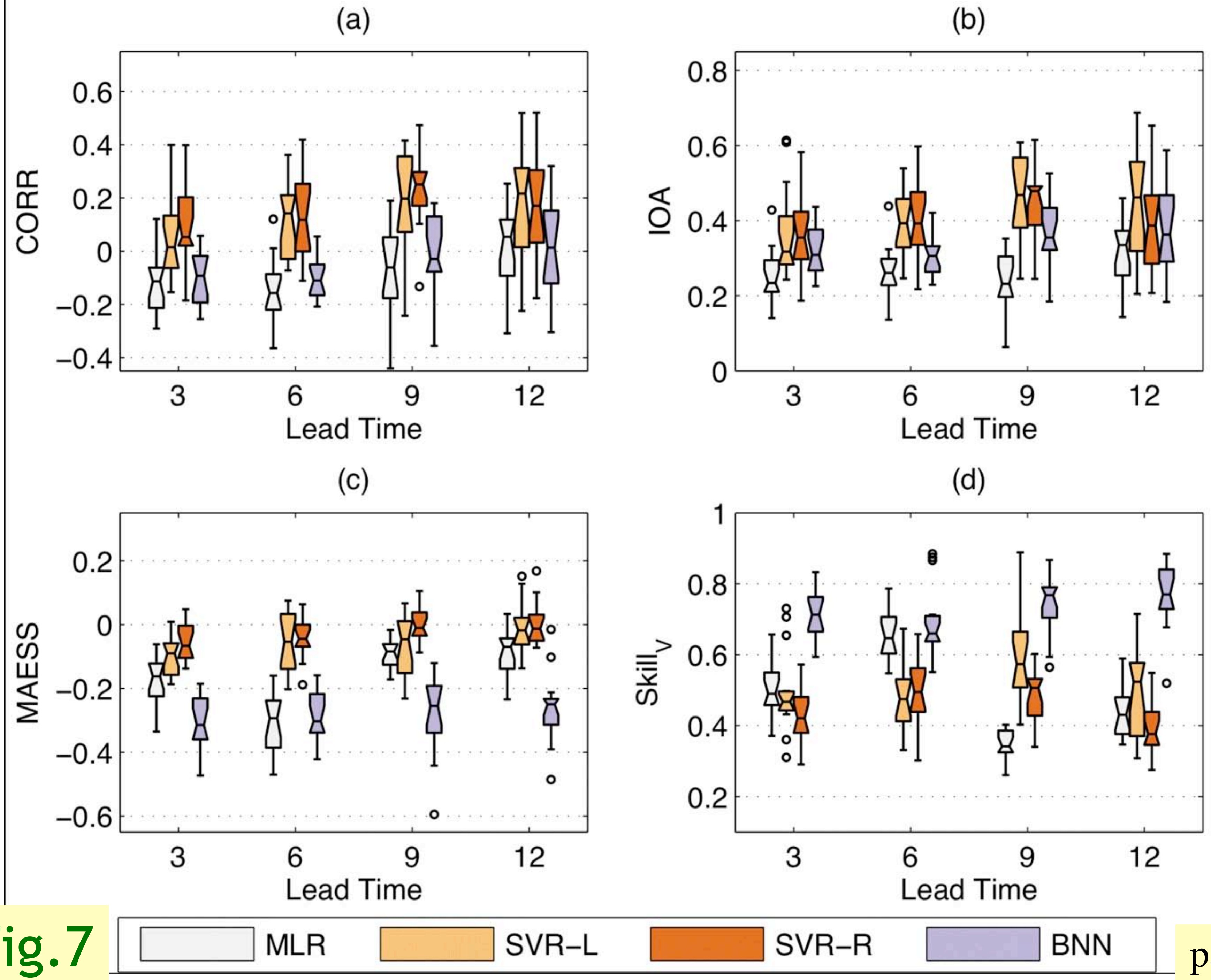


Fig.7

Region 6: Atlantic coast

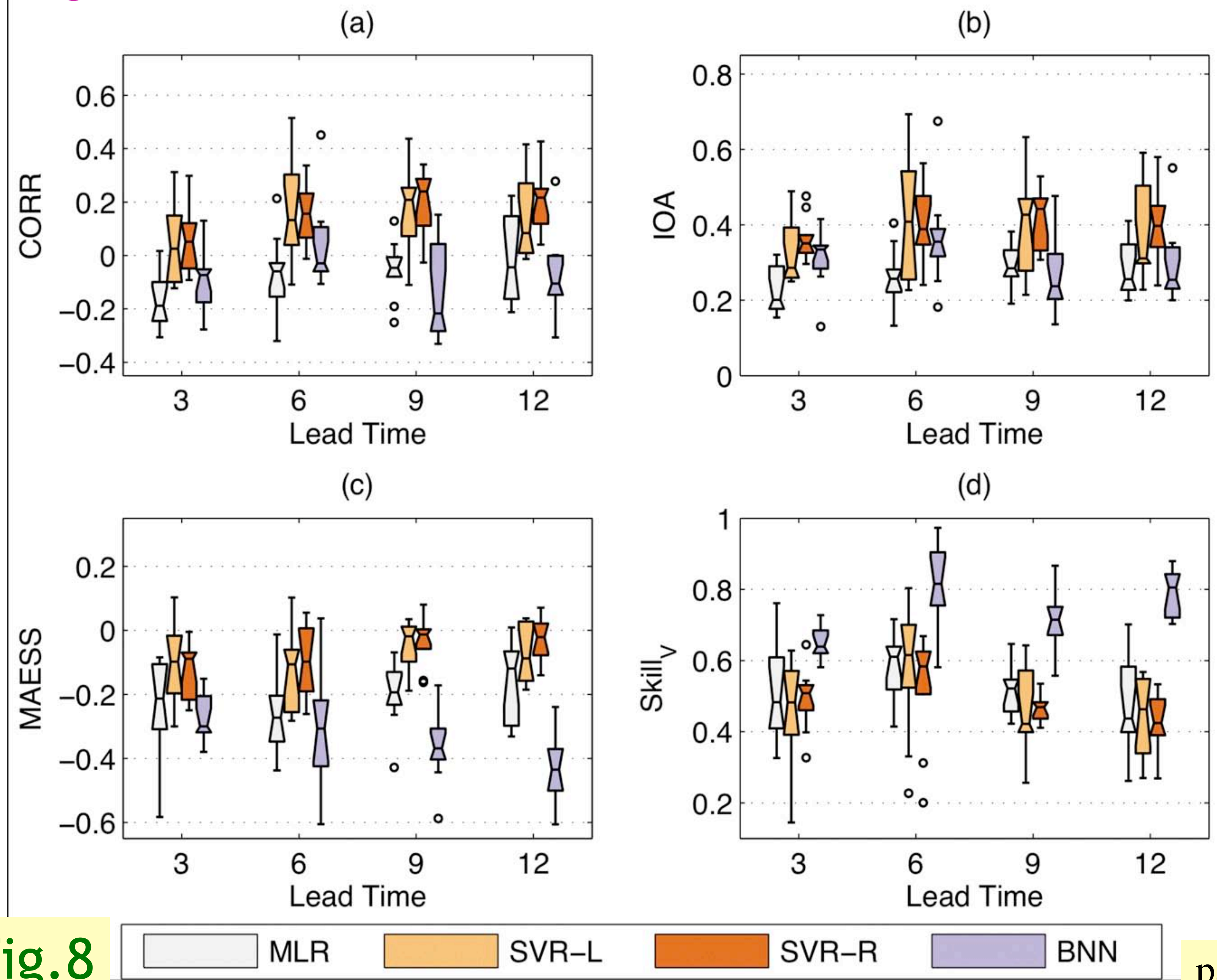


Fig.8

Fig.9 shows the geographic distribution of correlation skill for the SVR-R model at lead times of 3,6,9,12 months.

Fig.10 shows the difference between the correlation skill of the SVR-R and that of the MLR.

Correlation skills of SVR-R at 3,6,9 & 12 mo.

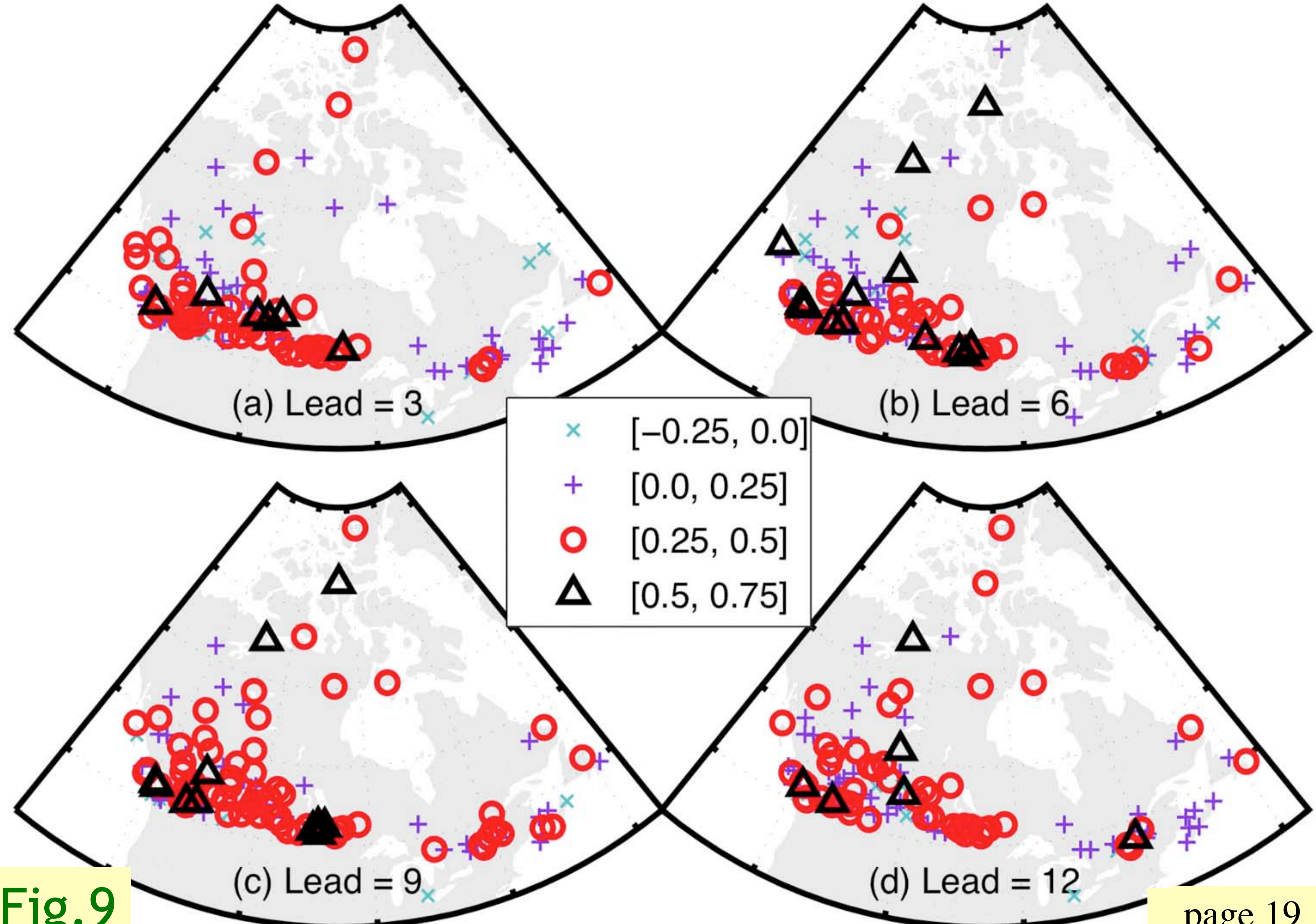


Fig.9

Correlation skills of SVR-R minus skills of MLR

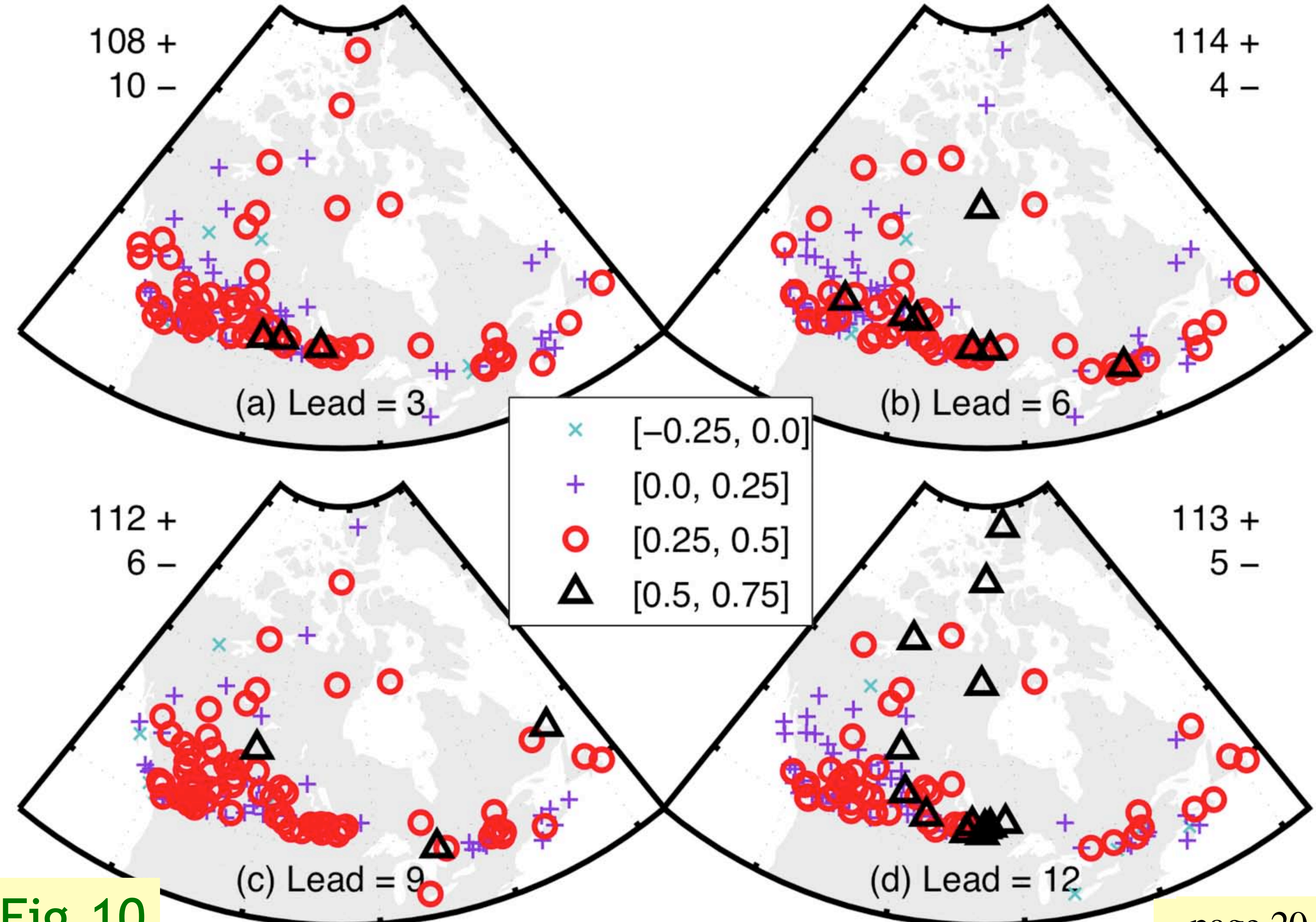


Fig.10

Conclusions

- In general, the robust SVR models tended to outperform the non-robust MLR and BNN models.
- The nonlinear SVR model tended to forecast slightly better than the linear SVR, except in the Arctic.
- The eastern Prairies region displayed the highest forecast skills & the Arctic region the second highest.
- Nonlinearity was strongest over the eastern Prairies & weakest over the Arctic.