Nonlinear post-processing of seasonal climate forecasts

J. Finnis, W. Hsieh, Z. Zeng, (University of British Columbia)

Introduction

- **Goal:** Develop a robust method of correcting systematic biases in GCM (CHFP) forecasts
- **Approach:** Regression-based post-processing
  - Link model output to true climate outcomes
- Several regression methods tested

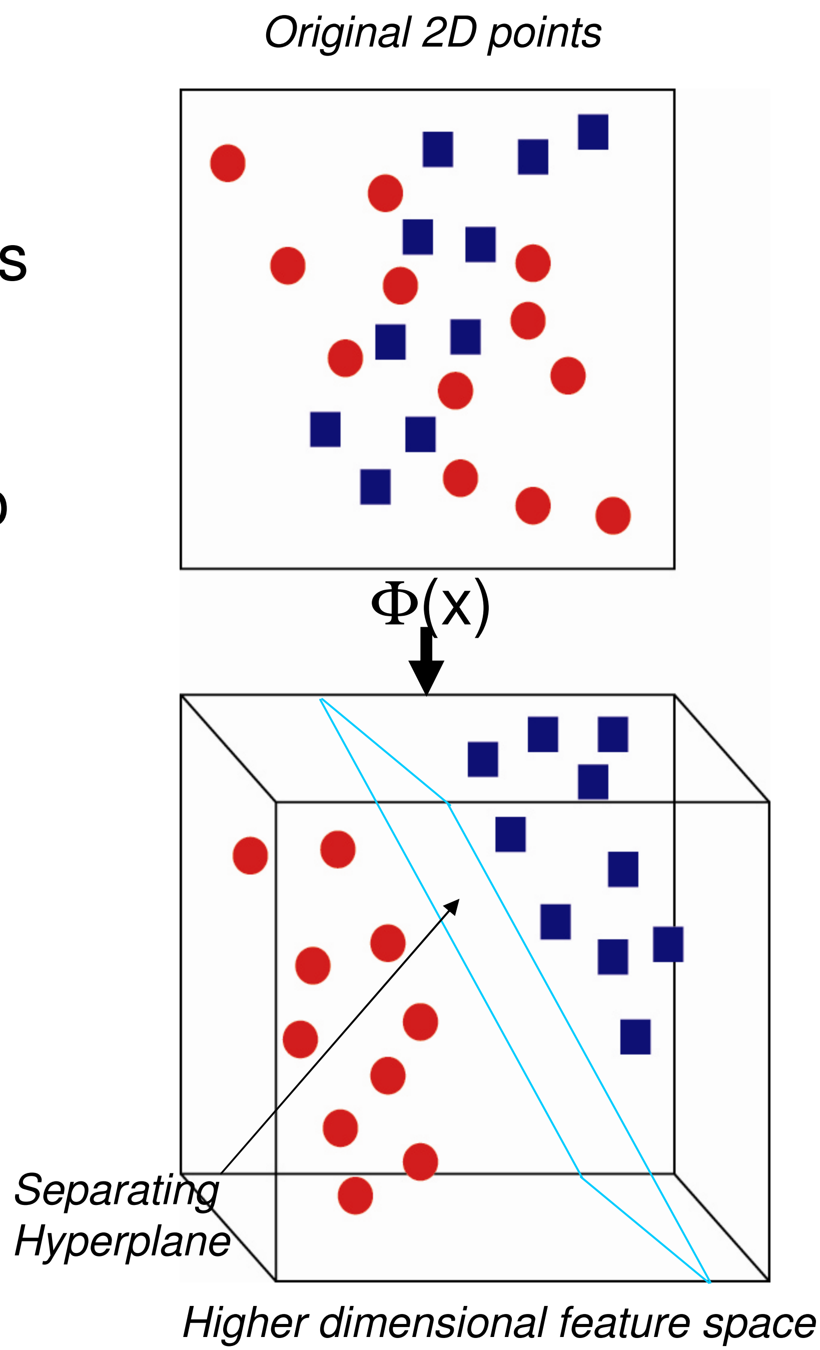
1. Linear regression
2. Support Vector Regression (SVR)
   - Machine learning approach
   - Capable of identifying nonlinear relationships
   - Use both linear & nonlinear flavours

Support Vector Machines

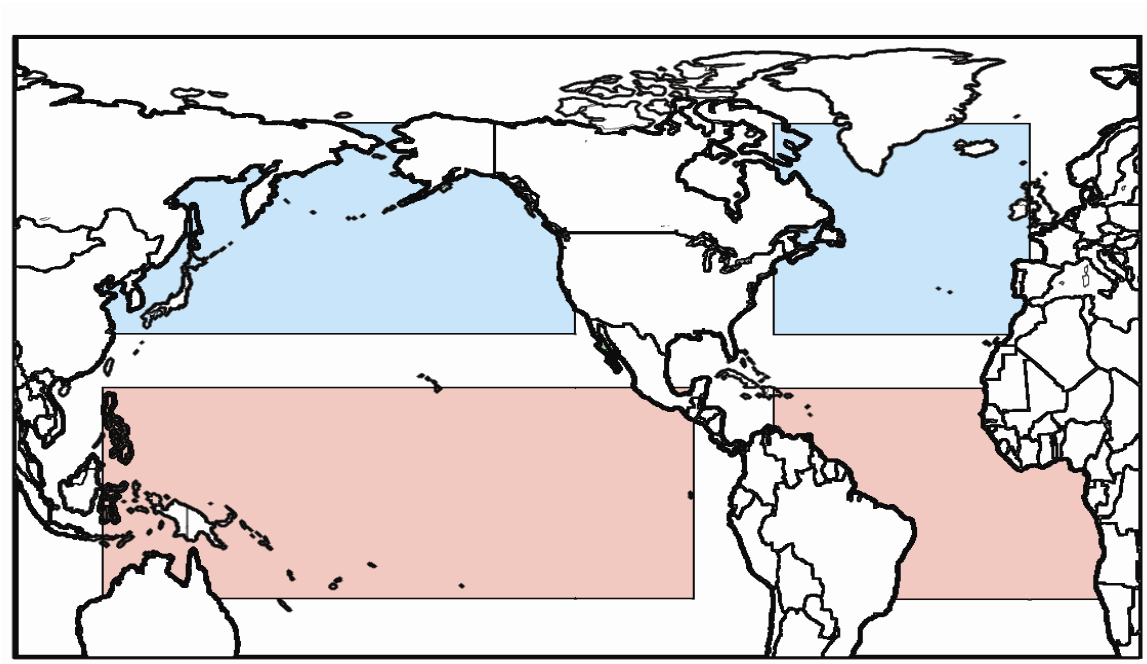
- Linearizes nonlinear relationships through feature space mapping ($\Phi(x)$)
- 'Kernel Trick': rewrite problem so we use only *inner products* of mapped points:

$$K(x', x) = \Phi^T(x')\Phi(x)$$

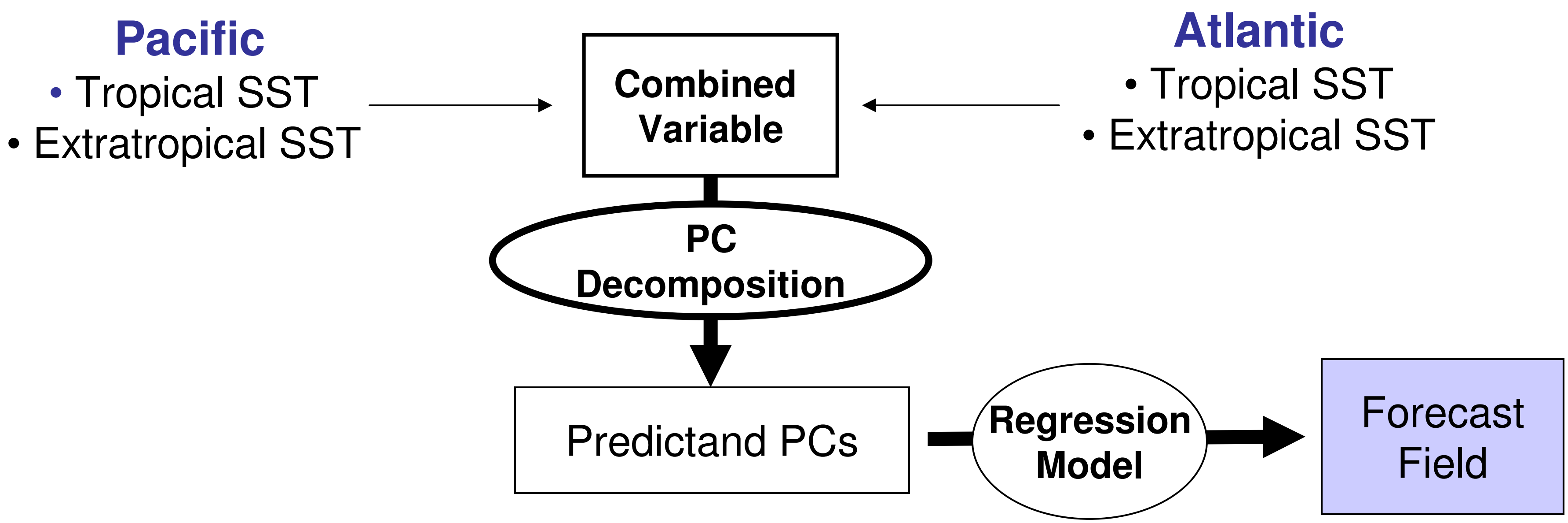
- A properly chosen kernel is solvable without performing the mapping $\Phi(x)$
- Feature space then have infinite dimension



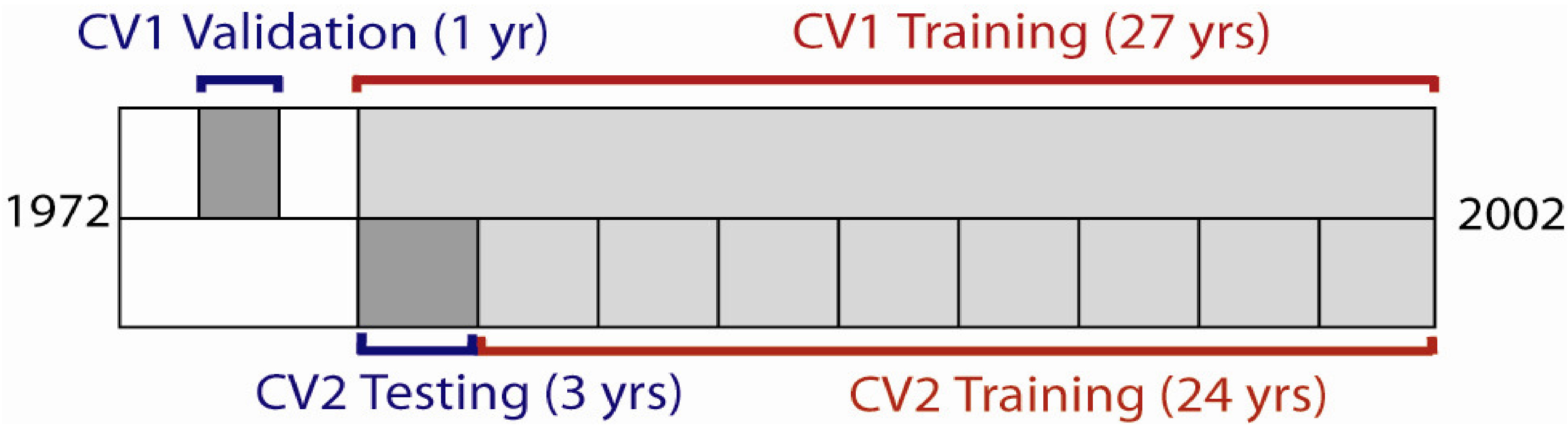
Variables & Domains



- Predictors from CHFP SST output
- 3 month 'seasons' (DJF, JFM, FMA, etc)
- Focus here on:
  - winter runs
  - 2m air temp forecasts over North America

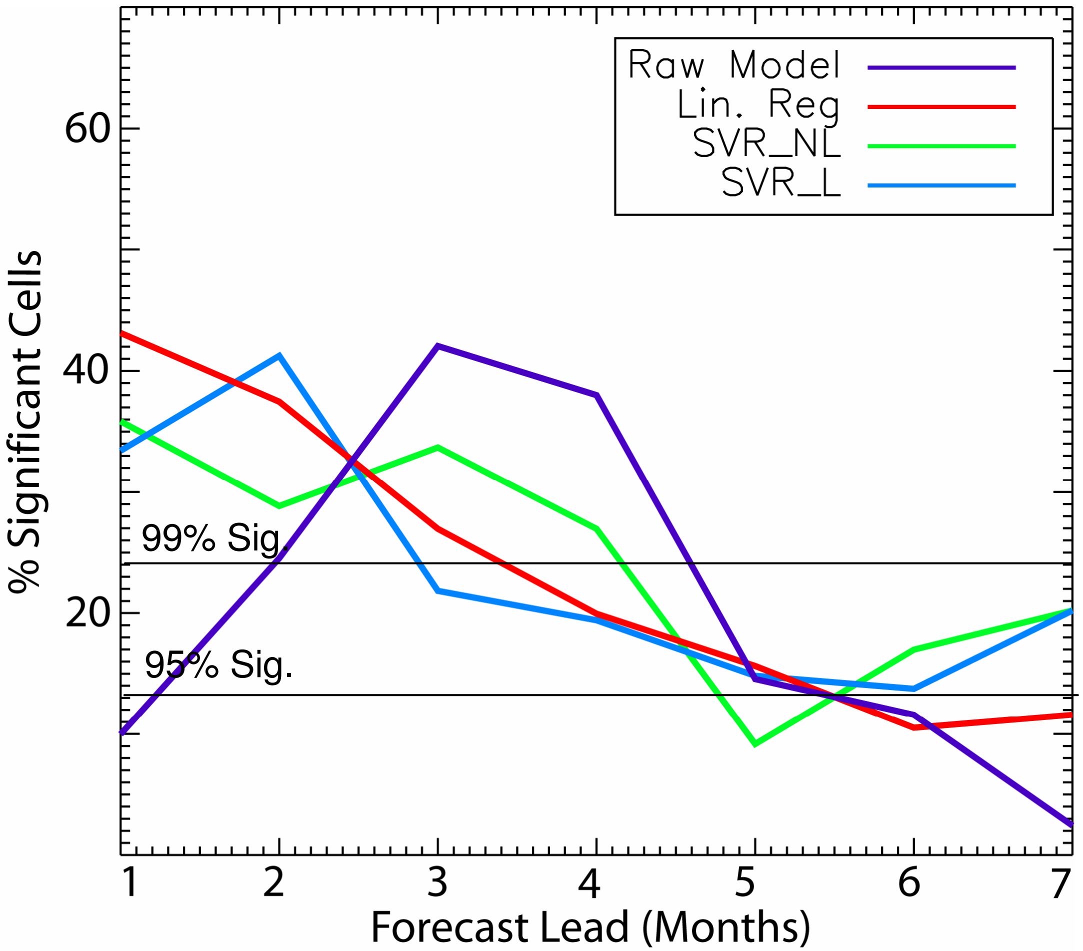


Method



- Methods tested using double cross-validation
  - CV1: Moving 3 year window, forecast center year
    - Removes skill due to autocorrelation
  - CV2: 9-fold x-validation on CV1 training data to select regression parameters
    - No information about forecast year used in parameter selection

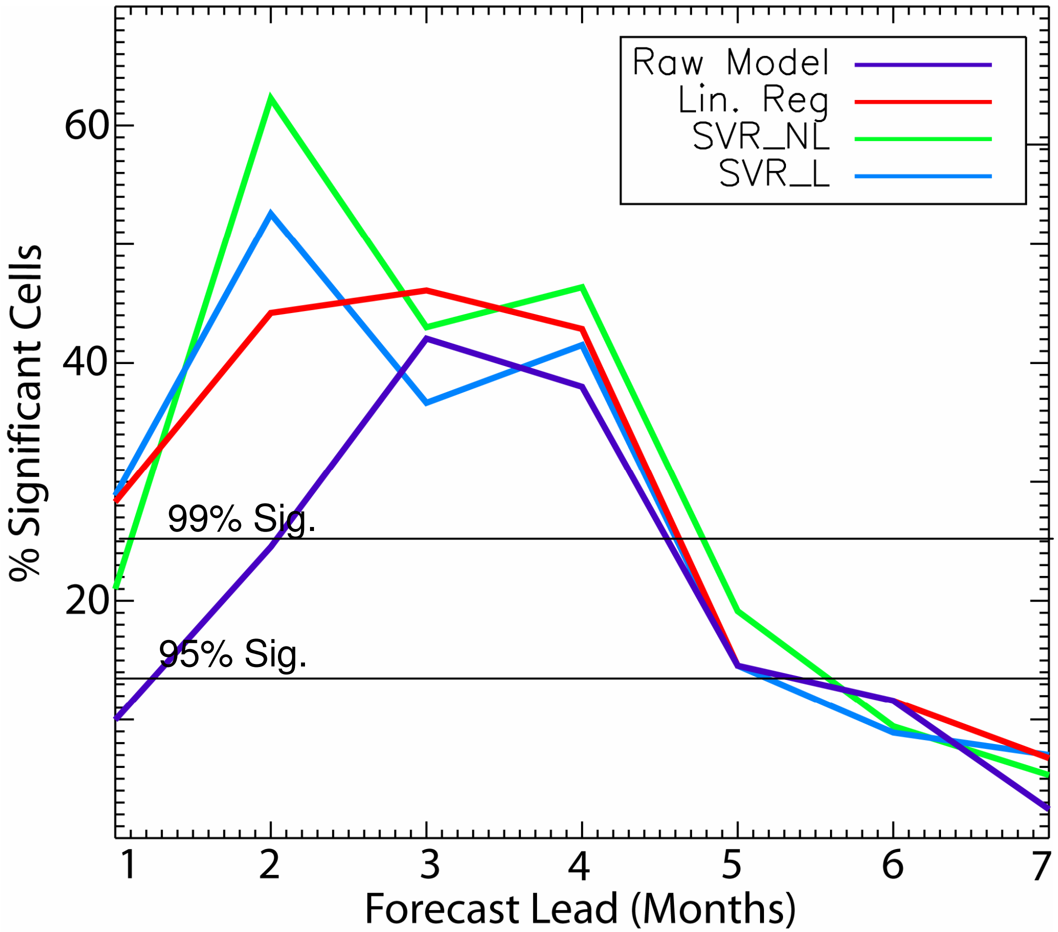
Direct Prediction: 2m Air Temp, North America



- Replace CHFP PCs with regression model output (first 5 only)

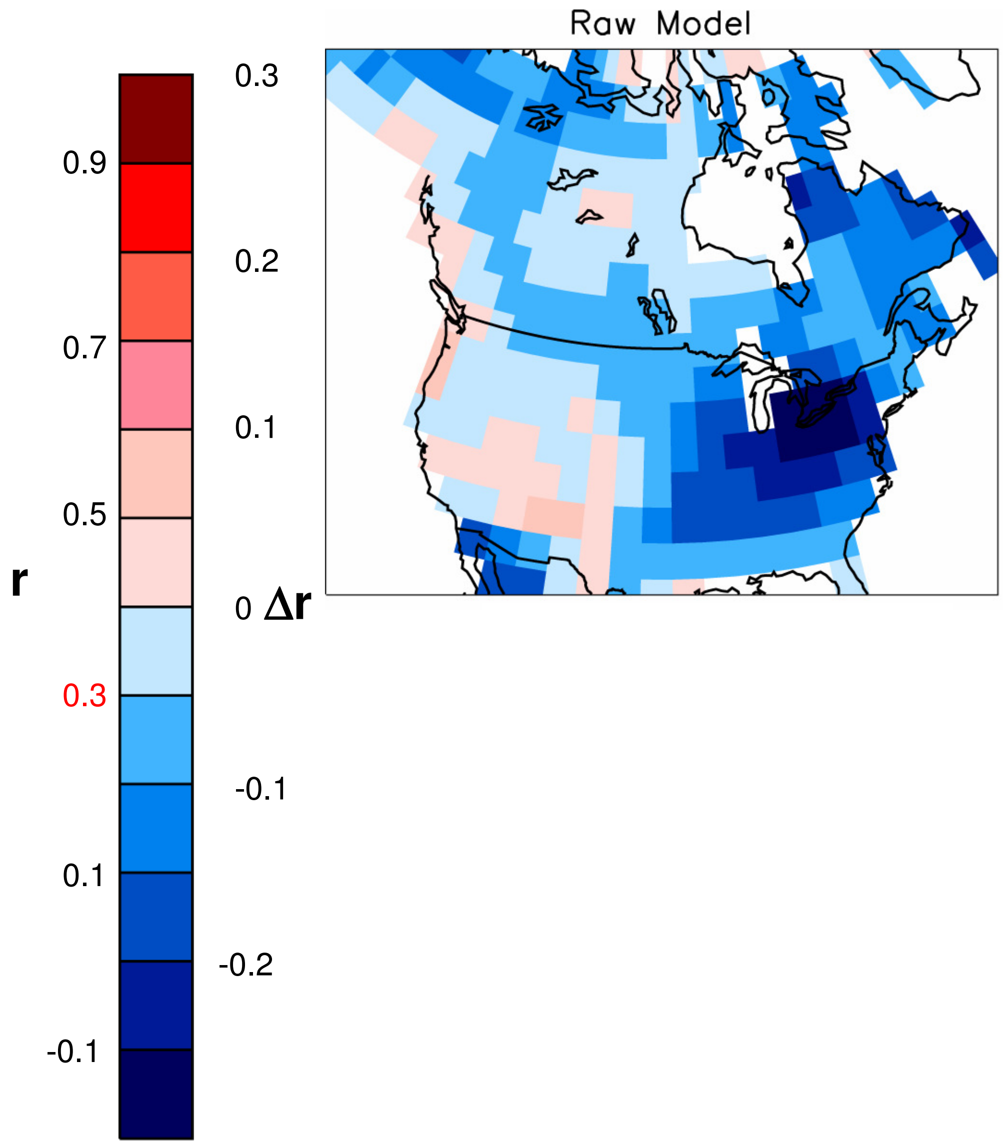
- Post-processing beats model at short and very long leads only

Alternate Approach: Predict Forecast Error

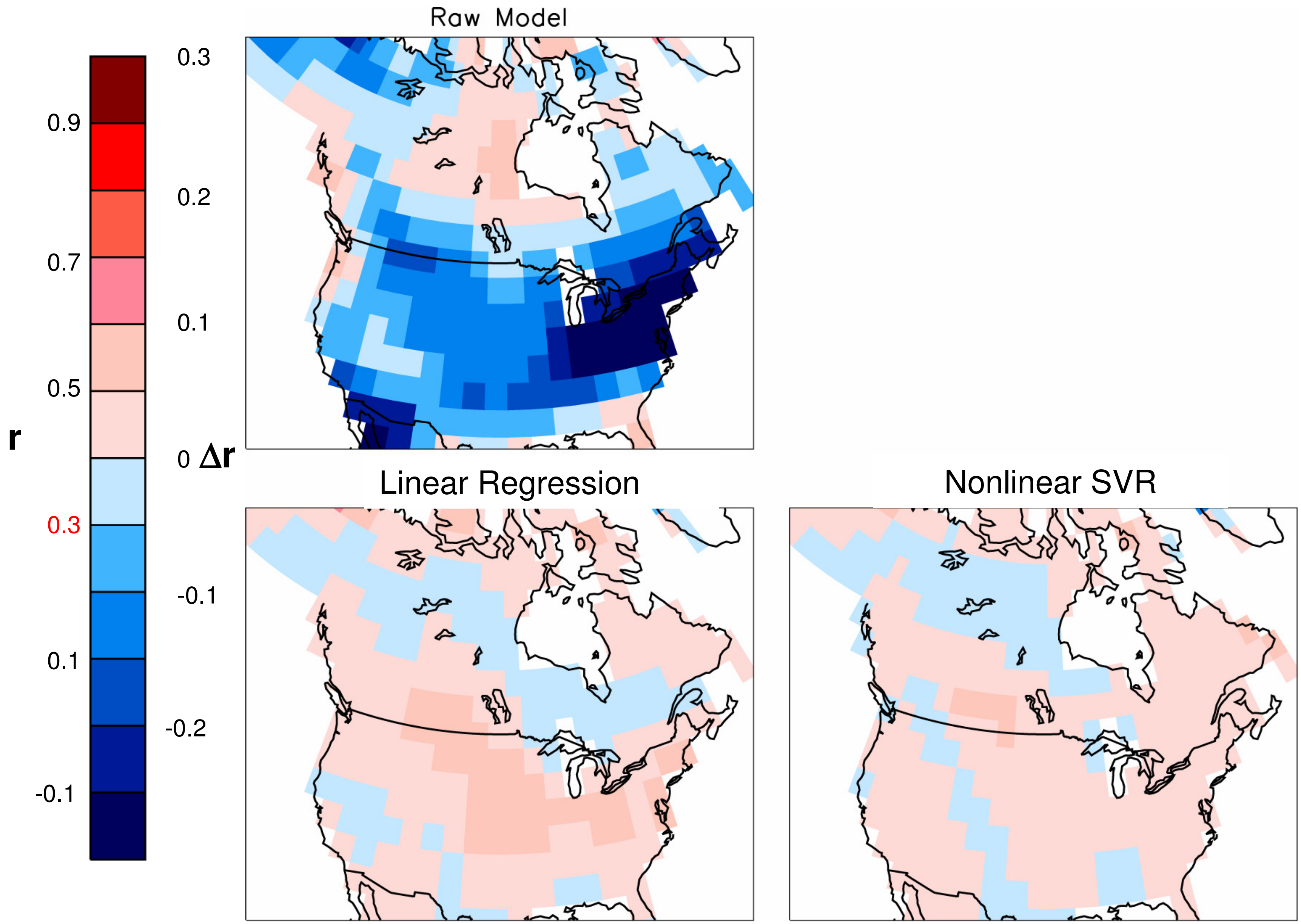


- Regression model predicts error in SAT PCs (model – observed)
- If prediction shows skill (r > 0.35), predicted error is subtracted from CHFP forecast
- Correct only first 5 PCs
- Improvement is more consistent

Grid Cell Correlations: 2 month Lead



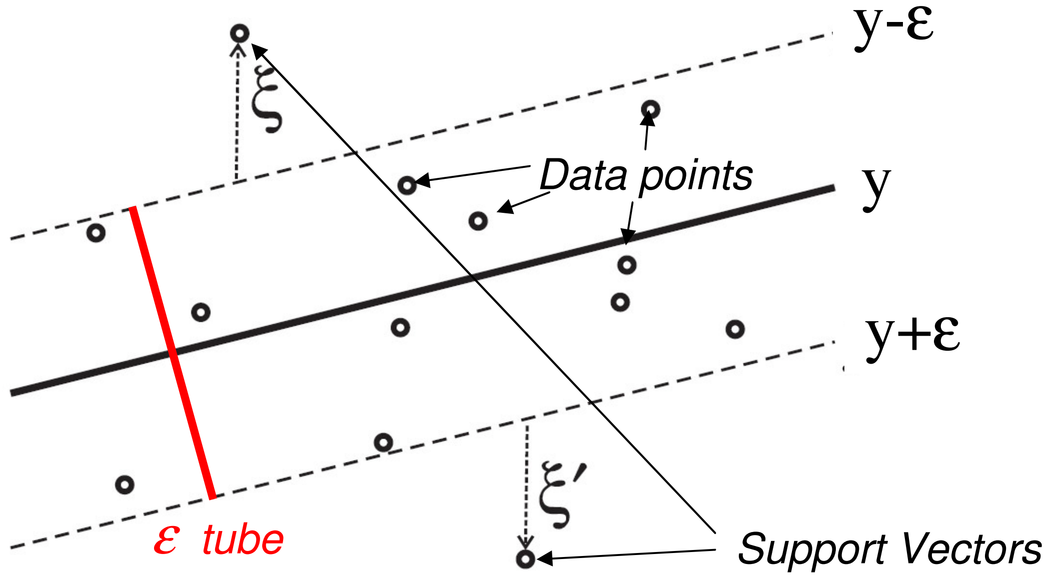
Grid Cell Correlations: 4 month Lead



Summary

- Regression-based post-processing can improve GCM forecasts at lead times of 1-4 months
- Predicting error gives the most consistent results
- Nonlinear SVR shows some promise relative to linear regression …
    - …although results vary between lead times
    - Too few points to identify nonlinear relationships?
    - Adding noise from individual ensemble members may improve skill

Support Vector Regression (SVR)



- Only certain training points define y = f(x)
    - Located a distance >= ε away from y
    - Referred to as *support vectors*
- 2 or more user selected hyperparameters

    a) ε-insensitive distance (which points to keep)

    b) Inverse Cost Parameter, *C* (penalizes complexity)

    c) Kernel shape parameters (e.g. σ for Gaussian kernel)