NASA/NCEAS/iPlant Update

Adam Wilson

April 17, 2012

Climate Aided Interpolation (a brief review)

Theoretical foundation:

1. *temperature anomalies are strongly correlated out to distances of the order of 1000 km* (Hansen and Lebedeff, 1987)
2. anomalies are *relatively free of the considerable topography-forced spatial variability* (Willmott & Robeson, 1995)
3. *spatial variability within the climatology accounts for most of the temporal between-station variability* (Willmott & Robeson, 1995)

Climate Aided Interpolation (a brief review)

1. Di Luzio, M., Johnson, G. L., Daly, C., Eischeid, J. K., & Arnold, J. G. (2008). Constructing Retrospective Gridded Daily Precipitation and Temperature Datasets for the Conterminous United States. *Journal of Applied Meteorology and Climatology*, 47(2), 475497.
2. Hunter, R. D., & Meentemeyer, R. K. (2005). Climatologically aided mapping of daily precipitation and temperature. *Journal of Applied Meteorology*, 44(10), 15011510.
3. Perry, M., Hollis, D., Perry, M., & Hollis, D. (2005). The generation of monthly gridded datasets for a range of climatic variables over the UK, The generation of monthly gridded datasets for a range of climatic variables over the UK. International Journal of Climatology, *International Journal of Climatology*, 25, 25(8, 8), 1041, 10411054, 1054. doi:10.1002/joc.1161, 10.1002/joc.1161
4. Willmott, C. J., & Robeson, S. M. (1995). Climatologically aided interpolation (CAI) of terrestrial air temperature. *International Journal of Climatology*, 15(2), 221229.

Climate Aided Interpolation (a brief review)

Generating daily climate anomalies

$$P_{\text{anomaly}} = \frac{P_{\text{daily}}}{P_{\text{monthlyclimate}}} \tag{1}$$

and temperature:

$$T_{\text{anomaly}} = T_{\text{monthlyclimate}} - T_{\text{daily}} \tag{2}$$

Goal: 1km Global Coverage over 1971-2010

A proposal: shift focus to two products

1. 1km Monthly (2000-2011 climatologies):
   topography + mean monthly station + mean monthly satellite

   Pros

   - Avoids problems of incomplete daily satellite data (clouds)
   - monthly means much smoother (easier) than daily
   - greatly simplifies processing (12 layers rather than 14,000)

   Cons

   - losing daily information from satellite data

Goal: 1km Global Coverage over 1971-2010

A proposal: shift focus to two products

1. 1km Monthly (2000-2011 climatologies):
   topography + mean monthly station + mean monthly satellite

   Pros

   - Avoids problems of incomplete daily satellite data (clouds)
   - monthly means much smoother (easier) than daily
   - greatly simplifies processing (12 layers rather than 14,000)

   Cons

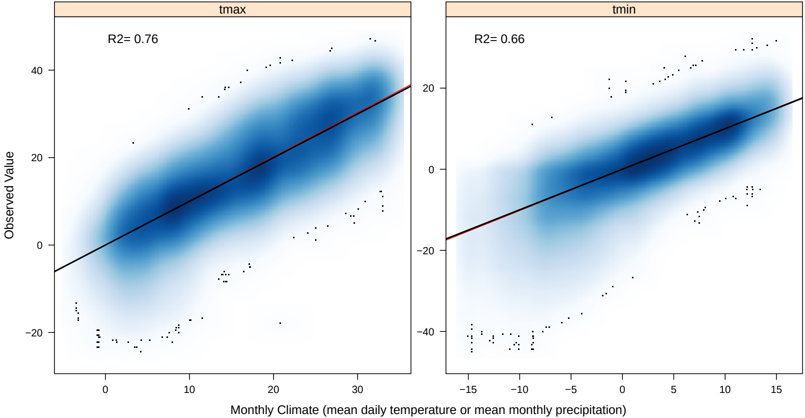
   - losing daily information from satellite data
2. Daily climate-aided interpolation for 1970-2011

   assume spatial stationarity!

CAI Comparison

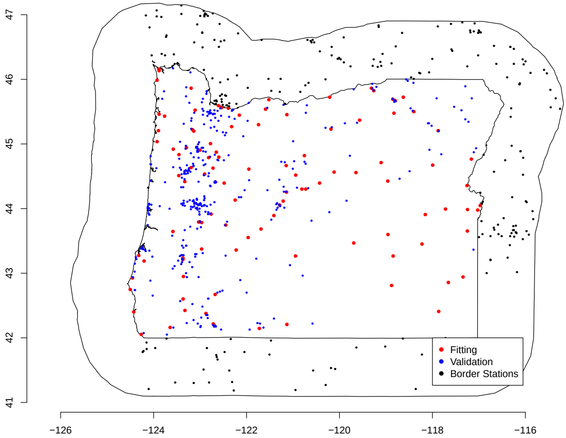
1. Oregon Stations
2. 70% holdout and 10 sample days identified by Benoit.
3. Bayesian 'krige' using spLM in spBayes R package
4. compare predictions using climate anomaly and 'raw' station values
5. Simple model with y∼intercept+lon+lat+elev
6. With anomalies, using no covariates fits about the same!

CAI Comparison



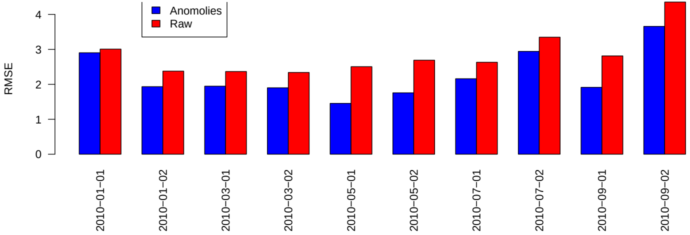
PRISM climate vs. daily values

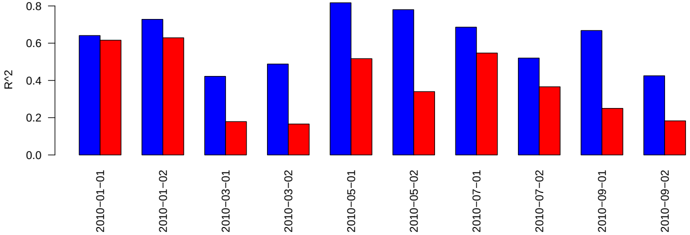
CAI Comparison



Note: we should include stations outside ROI to avoid edge effects.

CAI Comparison





Comparison of Methods

1. Interpolating climate anomalies improves predictive accuracy
2. Bayes overall mean RMSE 2.25^oC (compared with $\approx$2.39 for 'best case' other models).
3. Full accounting for predictive uncertainty
4. Time consuming ($>>$2 hours / surface)

Interpolation: Multivariate Response

Incorporate tmax, tmin, and (perhaps) ppt into a single model?

1. Borrow strength across variables (increase n)
2. Increases model complexity (and probably run-time)
3. Improves fit???

Develop using spBayes package in R `http://cran.r-project.org/web/packages/spBayes/index.html`

MODIS Cloud Product (MOD05_L2)?

Level 2 (swath) data at 1km resolution
2000-2012 Oregon data incudes 3,422 files and ≈70GB data

Layers:

1. Cloud Optical Thickness (0 to 100)
2. Cloud Optical Thickness Uncertainty (0 to 200%)
3. Cloud Effective Radius: particle size (0 to 90 μm)
4. Cloud Effective Radius Uncertainty(0 to 200%)
5. Cloud Water Path (0 to 9000 g/m^2)
6. Cloud Water Path Uncertainty (0 to 9000 g/m^2)
7. Cloud Phase Optical Properties (0=fill, 1=clear, 2=liquid water cloud, 3=ice cloud, 4=undetermined phase cloud)
8. Cloud Multi Layer Flag (0=fill, 1=single layer, 2 through 8=increasing confidence of multilayer clouds)

Partner with TOPS/NEX to get this processed to grid level monthly mean data?

Next Steps

First the climate surfaces:

1. MODIS LST (day/max and night/min) monthly means
2. MODIS Cloud? TRMM Cloud?
3. Use interpolator of choice (spline, krig, GWR)to predict monthly means
4. Stationarity: Show past (1970-2000) station data has similar relationship to climate as current (2000-2010)

Then interpolate anomalies:

1. GWR? Splines? Kriging? Bayesian Kriging?
2. goal changes from explanation (via covariates) to description of pattern

Another Idea

Use existing (continuously updated) gridded anomaly surface to 'add' to our high resolution climate.

Gridded daily GHCN data (2.5^o latitude by 3.75^o, 1950-present): `http://www.metoffice.gov.uk/hadobs/hadghcnd/`.
Too coarse to be useful??

Unified Geostatistical Modeling for Data Fusion and Spatial Heteroskedasticity with R Package `ramps`[1]

Allows fusion of data obtained at dierent resolutions (areal and point-referenced) and spatial heteroskedasticity.

1. Joint modeling of data from multiple sources (point-source, areal, or both).
2. Non-spatial random effects as in a linear mixed model.
3. Multiple variances for each variation source (measurement error, spatial, and random effects).
4. Prediction at measured or unmeasured sites.

Would be useful if we decide to incorporate $1/4^o$ TRMM data

Unified Geostatistical Modeling for Data Fusion and Spatial Heteroskedasticity with R Package ramps

$$Y = X\beta + W\gamma + KZ + \epsilon \tag{3}$$
$$\gamma \sim N(0, \Sigma_\gamma) \tag{4}$$
$$Z \sim N(0, \Sigma_Z) \tag{5}$$
$$\epsilon \sim N(0, \Sigma_\epsilon) \tag{6}$$

where X, W, and K are design matrices for fixed effects β $(p \times 1)$, non-spatial random effects γ $(q \times 1)$, and spatial random effects Z $(S \times 1)$, respectively.