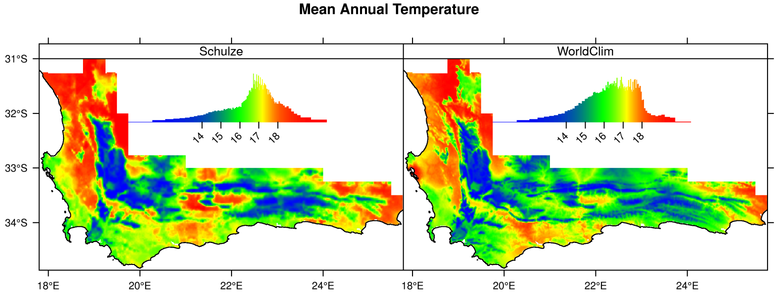
Quantifying uncertainty in daily weather interpolations: a Bayesian framework for developing climate surfaces

Adam M. Wilson

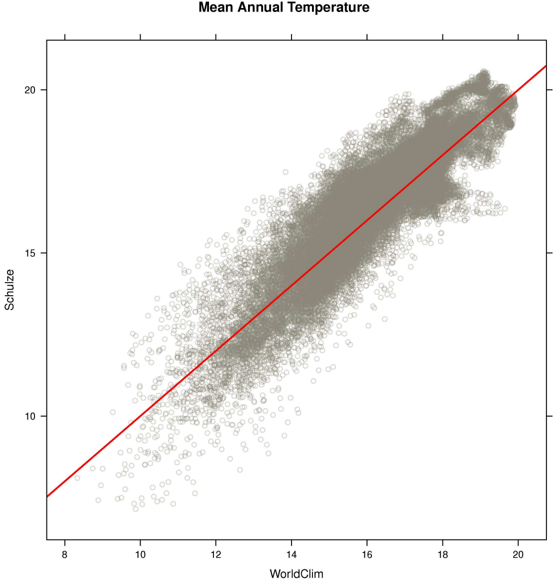
November 16, 2011

Climate data for South Africa: methods matter

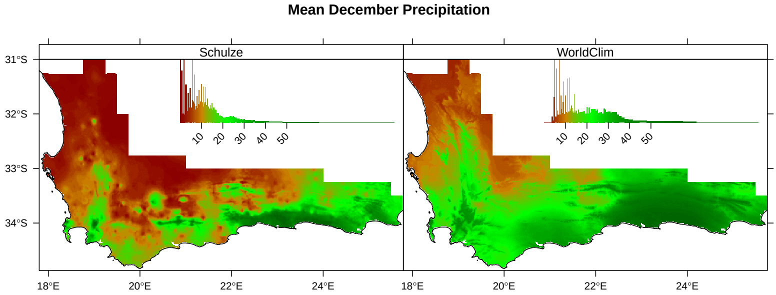


- Schulze, RE. (2008) *South African Atlas of Agrohydrology and Climatology*
- Hijmans, RJ., *et. al.* (2005) *International Journal of Climatology*, 25(15):1965–1978

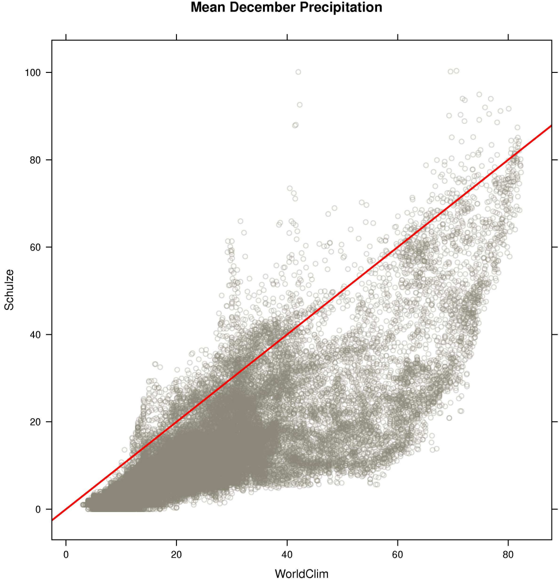
Climate data for South Africa: methods matter



Climate data for South Africa: methods matter



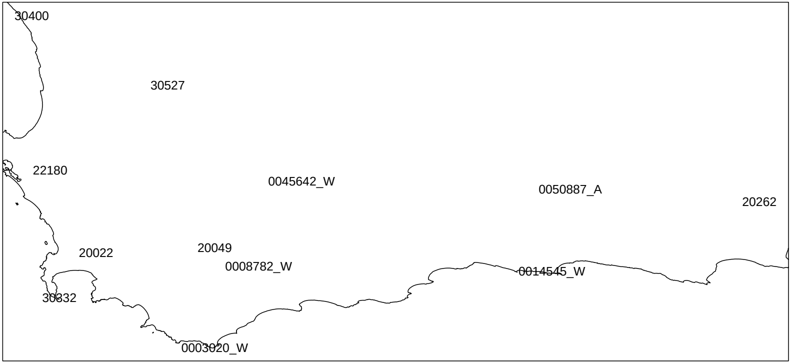
Climate data for South Africa: methods matter



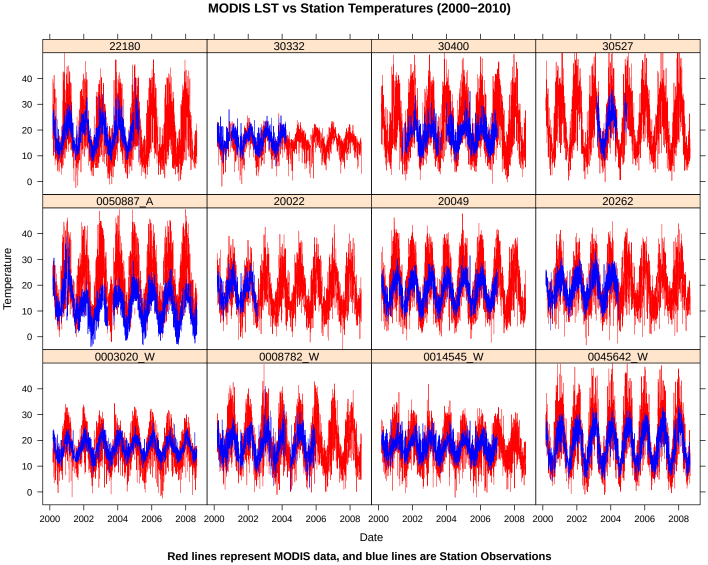
MODIS LST

A comparison of MODIS LST with selected station data from South Africa

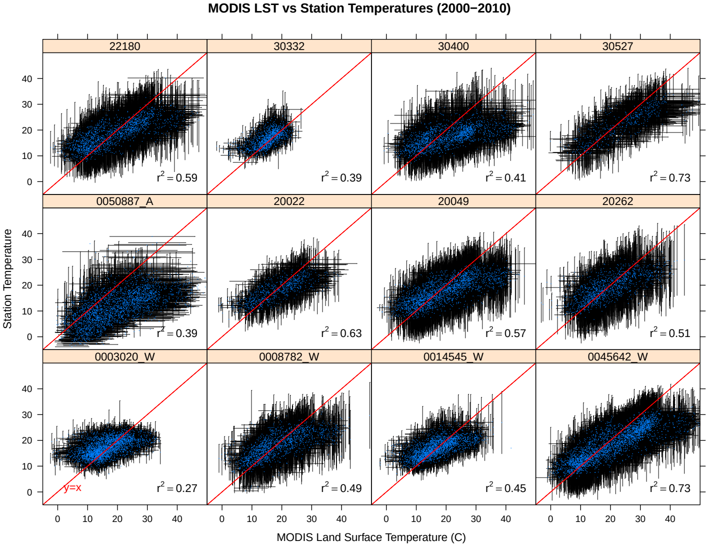
Station locations



MODIS LST

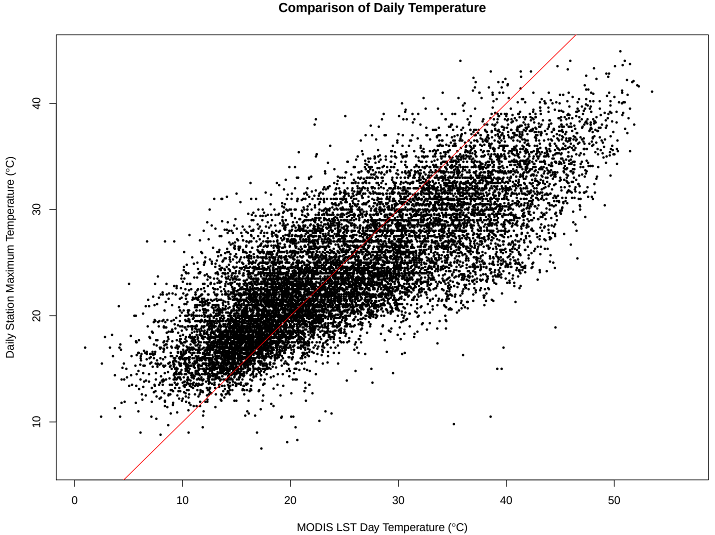

MODIS LST vs Station Temperatures (2000−2010)
22180
30332
30400
30527
0050887_A
20022
20049
20262
0003020_W
0008782_W
0014545_W
0045642_W
Temperature
0
10
20
30
40
2000
2002
2004
2006
2008
Date
Red lines represent MODIS data, and blue lines are Station Observations

MODIS LST

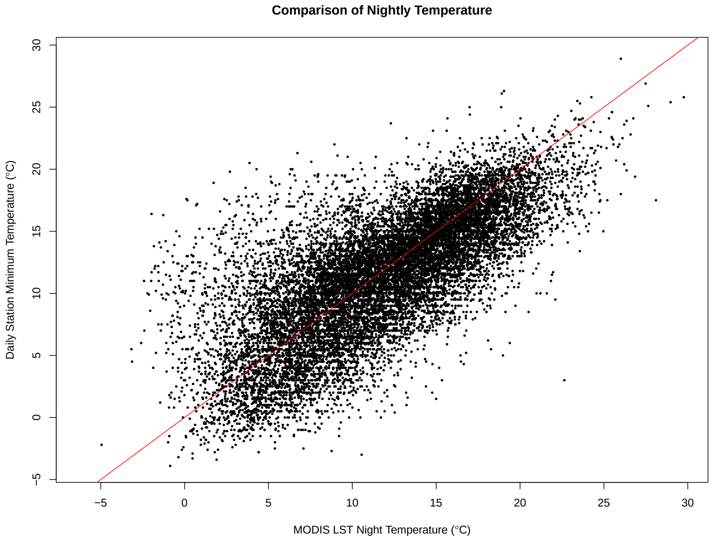


Grey lines represent range, blue points are mean temperatures, and red line is y=x

MODIS LST

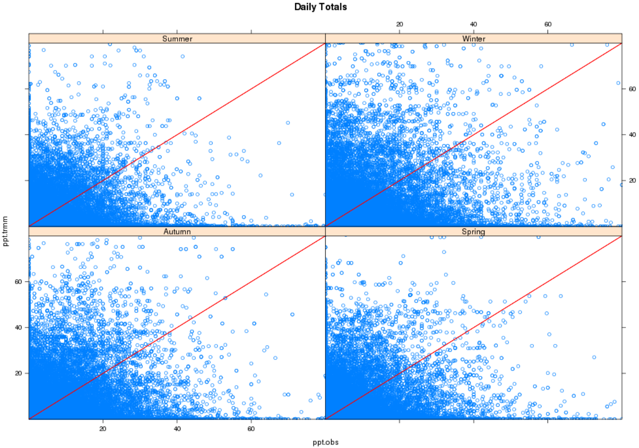

Comparison of Daily Temperature
Daily Station Maximum Temperature (°C)
10
20
30
40
0
10
20
30
40
50
MODIS LST Day Temperature (°C)

MODIS LST

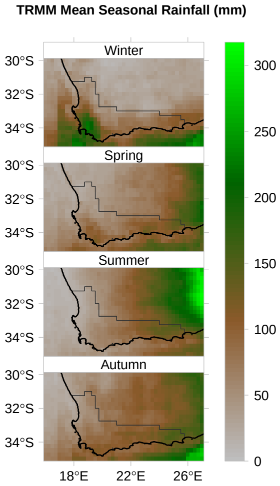

Comparison of Nightly Temperature
Daily Station Minimum Temperature (°C)
MODIS LST Night Temperature (°C)
−5
0
5
10
15
20
25
30

TRMM Precipitation

A comparison of TRMM daily precipitation with selected station data from South Africa



TRMM Precipitation

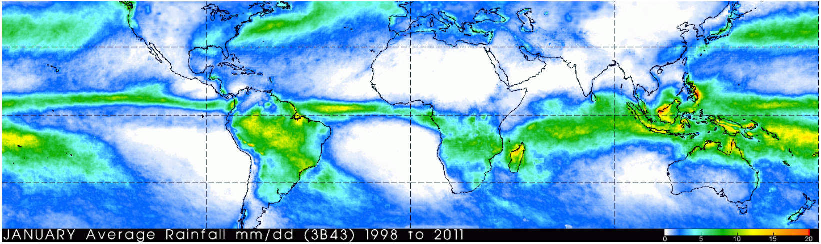


Seems to capture seasonal patterns much better than daily

MODIS and TRMM data

Long-term monthly means:

- probably more accurate than day-by-day (especially TRMM)
- better 'calibration' of satellite-station relationship
- reduce problem of clouds (though not everywhere)
- useful by itself (a better WorldClim)



Incorporating long-term means: Climate Aided Interpolation

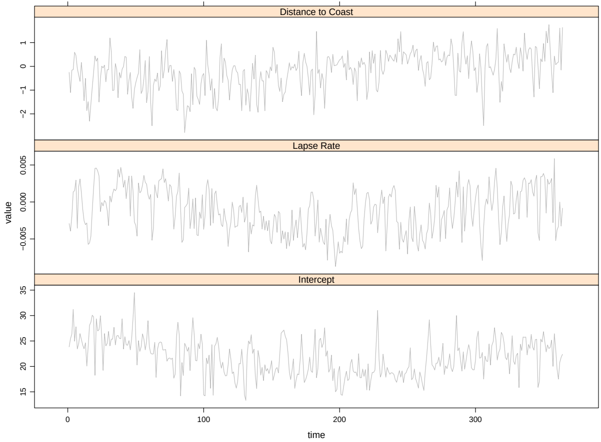
For each day:

1. Generate long term monthly means from station and satellite data
2. subtract long-term monthly mean from daily station observations
3. interpolate the anomolies
4. add anomaly surface back on to long-term means

Hunter & Meentemeyer, 2005; Willmott, & Robeson, 1995

Day-by-Day fitting

Fitted regression coefficients from day-by-day co-kriging on raw station temperatures over one year.



Incorporating long-term means: Climate Aided Interpolation

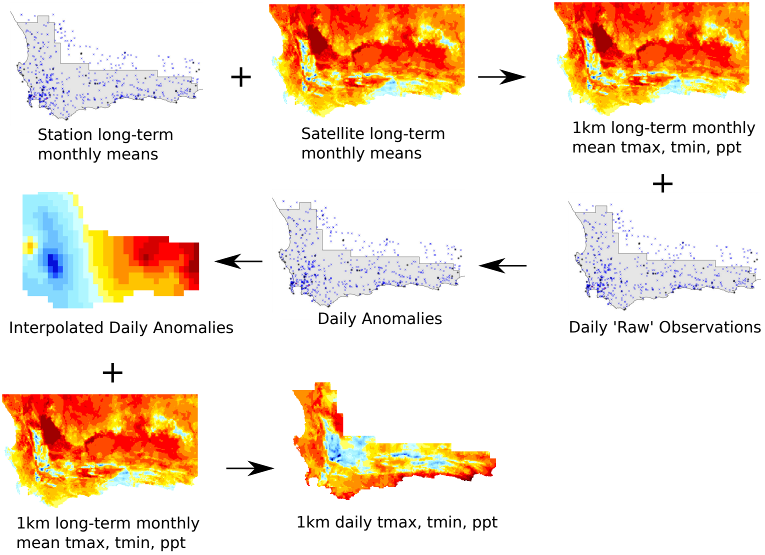
Advantages

- anomalies smoother and easier to interpolate
- don't need to estimate lapse rate, rain shadows, etc. each day
- altitudinal distribution of stations less problematic
- incorporate satellite data for entire period 1970-2010 (assuming stationarity)
- fewer problems with missing satellite data (clouds)

Disadvantages

- assume within-month spatial patterns (i.e. lapse rates) are constant (don't take direct advantage of daily satellite data)

Climate-aided Interpolation: The Workflow



Selecting Tile Size

Probably infeasable to interpolate a single day's values for globe, must break into tiles.

Factors to consider:

1. smaller is probably better for computation
2. larger is probably better for interpolation

Possible steps to select tile size

1. Compute semivariogram using moving window over globe to quantify spatial decay
2. Select smallest window above the range

The Workflow

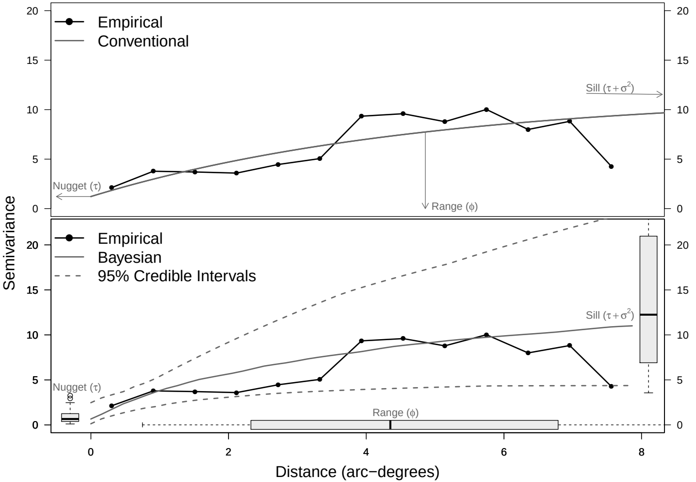
Generating daily climate anomalies

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

and temperature:

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

Climate-aided Bayesian Kriging



Semivariograms for maximum temperature on January 3, 2009

Climate-aided Bayesian Kriging

The full Likelihood:

$$L(\beta, \sigma^2, \phi | Y) \propto (\sigma^2)^{-\frac{n}{2}} |R_y(\phi)|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(y - X\beta)'(R_y(\phi))^{-1}(y - X\beta) \right\} \quad (3)$$

The posterior distribution:

$$pr(\beta, \sigma^2, \phi | y) = pr(\beta, \sigma^2 | y, \phi) pr(\phi | y) \quad (4)$$

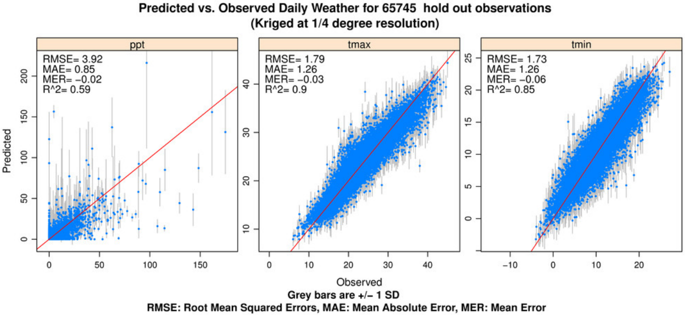
Day-by-day 'Bayesian krige' [1] using geoR package.

Climate-aided Bayesian Kriging

Computationally demanding. 20 years of interpolations requires:

- $>$1 year processor time
- $\sim$7TB of storage (though maybe not all at once)

Validation



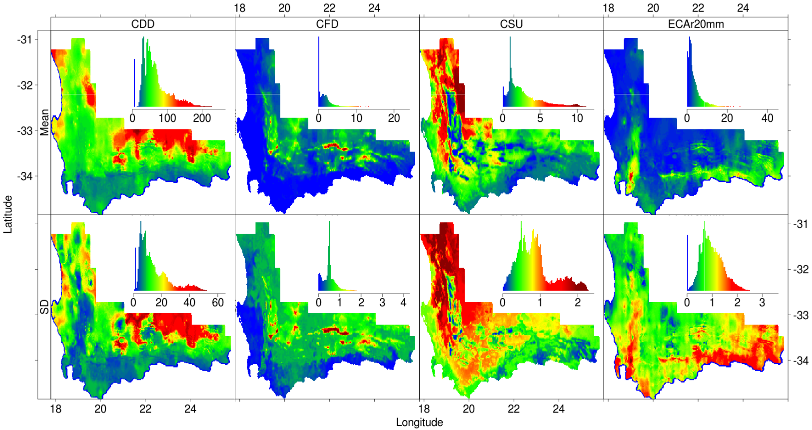
Successful prediction of dry days: 97.2% and wet days: 65.9%.

Climate metrics

Quantity	Description	Plant performance elements	Data	Functional form
MinT	Annual minimum temperature	Germination, growth	t_{min}	$\min(t_{min})$
MaxT	Annual maximum temperature	Germination, growth, Seedling mortality	t_{max}	$\max(t_{max})$
FD	Frost days	Seedling mortality	t_{min}	$\sum_{t \in \text{year}}(t_{min_t} < 0^o C)$
CFD	Longest consecutive period with frost	Seedling mortality	t_{min}	$\max(\text{consecutive}(t_{min} < 0^o C))$
GDD	Growing Degree Days	Growth	t_{max}	$\sum_{t \in \text{year}} \max(t_{min_t} - 10.0)$
CSU	Longest heat wave ($> 35^o C$)	Seedling mortality	t_{max}	$\max(\text{consecutive}(t_{max} > 35^o C))$
CDD	Annual maximum consecutive dry days	Growth, Seedling mortality	*ppt*	max(consecutive(*ppt* <2mm))
ECAr20mm	Very heavy precipitation days	Growth, Seedling mortality	*ppt*	Number of days with *ppt* >20mm
SDII	Simple daily precipitation intensity index	Growth, Seedling mortality	*ppt*	mean(*ppt*) where *ppt* >2mm

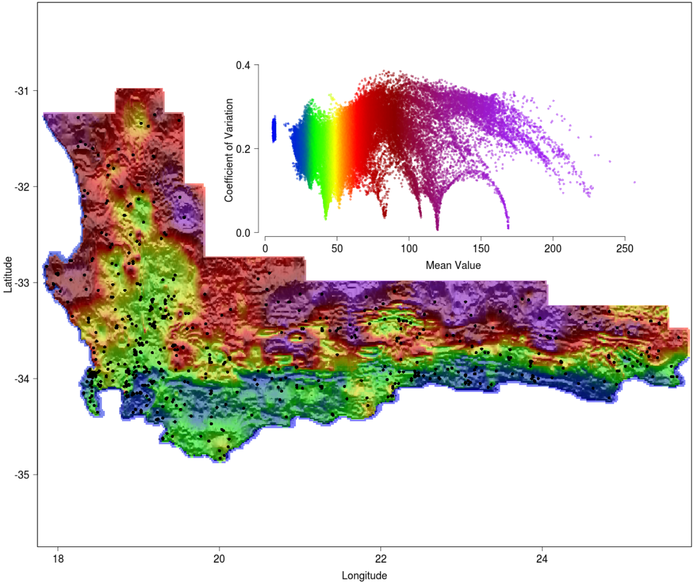
Climate metrics were calculated using 1,000 time series drawn from the posterior samples in each location to result in a posterior distribution that incorporates the uncertainty introduced by the interpolation. Climate metrics were calculated using CDO tools.

Summary of Climate Metrics

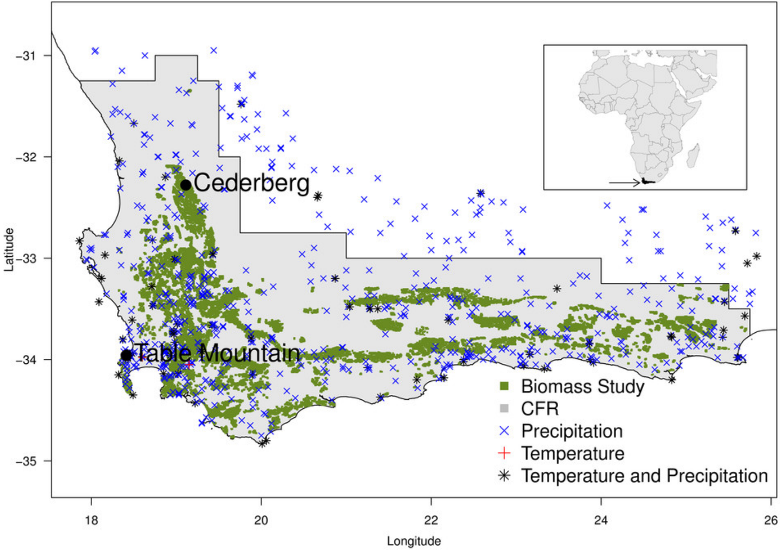


Mean (top row) and standard deviation (bottom row) of the posterior samples for four climate metrics.

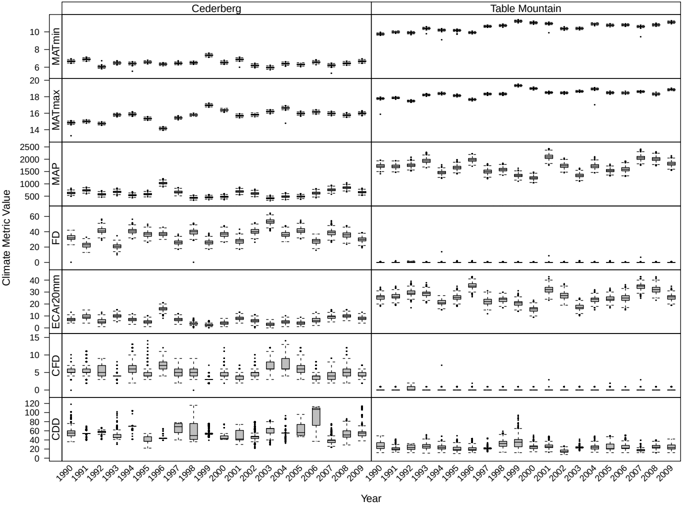
Consecutive Dry Days



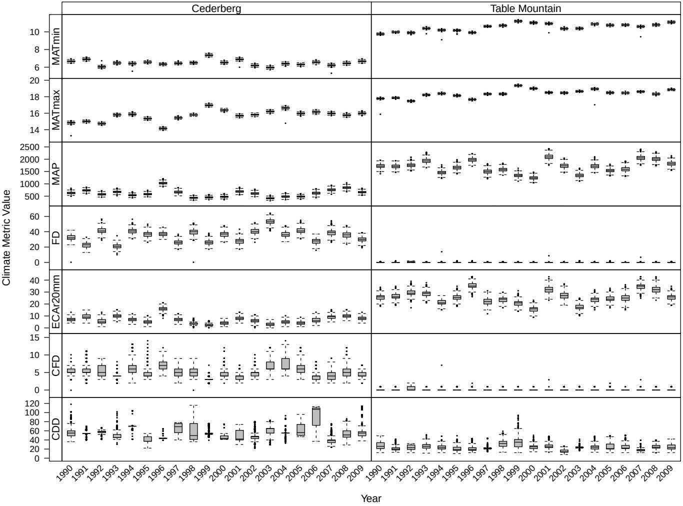
Comparison of two locations



Comparison of two locations



Comparison of two locations



We know what we don't know and we have more relevant metrics

Summary

Daily Bayesian interpolations provide:

- full accounting for uncertainty
- Posterior distribution for any $f(t_{max}, t_{min}, p_{tot})$ for any location

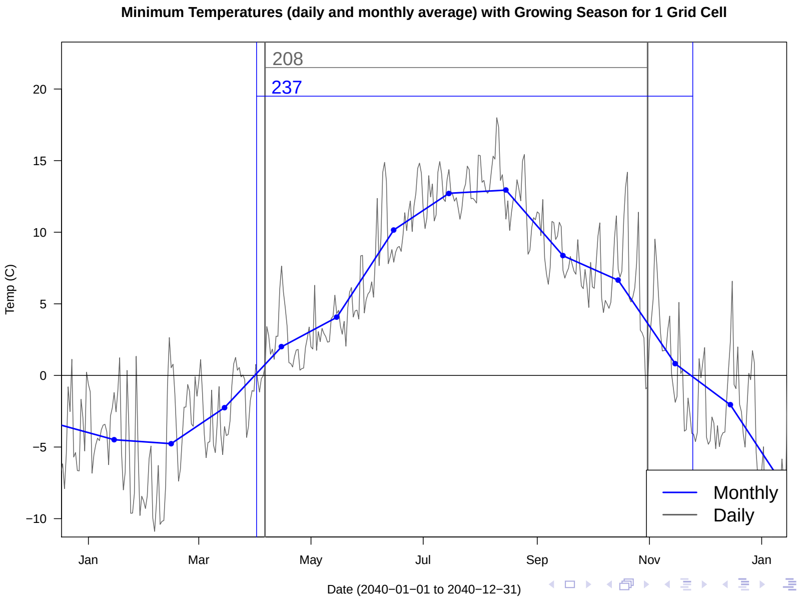
These distributions can propagate the uncertainty through:

- species distribution models
- ecosystem function models
- demographic models

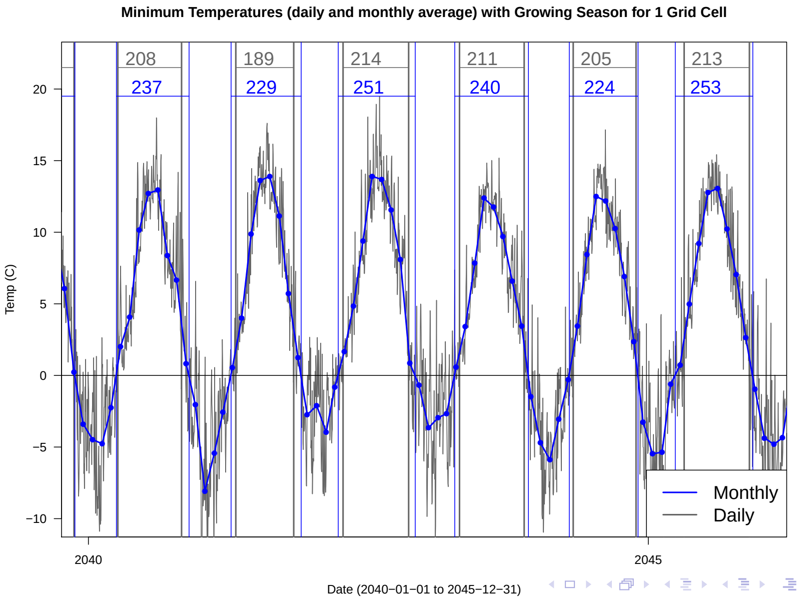
We can now quantify the effects of uncertainty in climate surfaces!

Thanks!

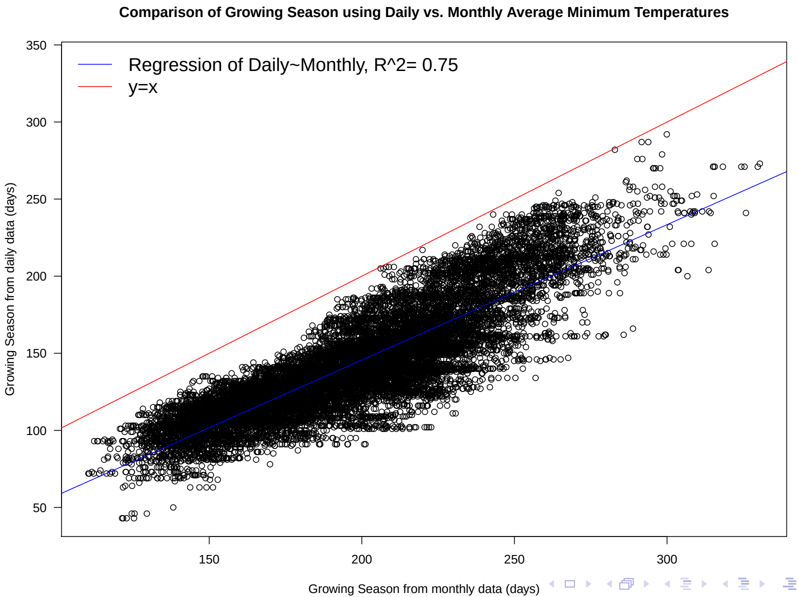
The need for daily data



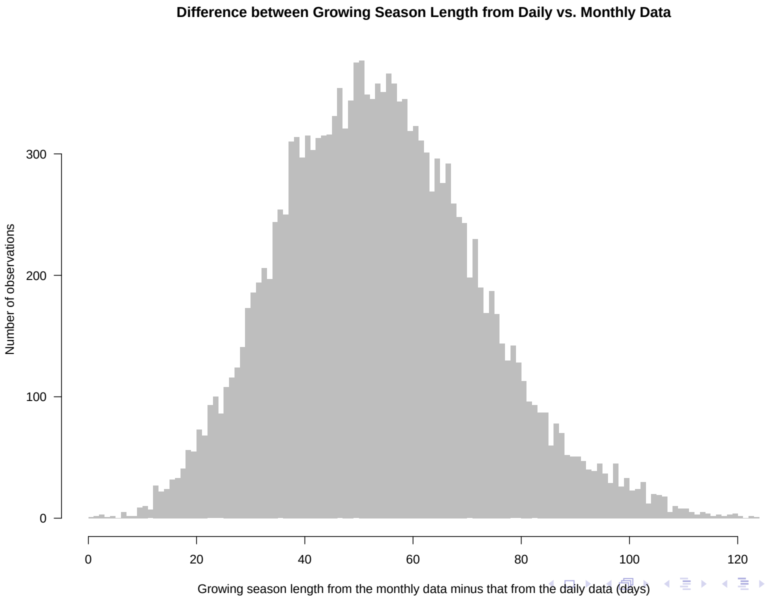
The need for daily data



The need for daily data



The need for daily data



Climate Metrics

Quantity	Description	Plant performance elements	Data	Functional form
MinT	"Chill" or annual minimum temperature	Germination, growth	Tt	$\text{min}_{\text{year}}(\text{Tmin}_t)$
FD	Frost days	Seedling mortality	T_{min}	Number of days during which Tmin $< 0^oC$
HDD	Heating Degree Days	Growth	T_{max}	$\sum_{t \in \text{year}} \max(T_t - 10.0)$
DLen	Annual maximum consecutive days with precipitation $<$ threshold (1mm)	Growth, Seedling mortality	ppt	max(consecutive(ppt $<$ 1mm))

Table: Climate metrics calculated from the daily data

Next... Climate projections

- Use CMIP3 or CMP5 GCM output
- Calculate anomalies (future daily - current monthly means)
- Apply to current high resolution climates
- Calculate metrics of interest

Then, maybe, I'll be able to think about ecology again...