METHODS COMPARISON
FOR THE PRODUCTION OF
INTERPOLATED CLIMATE LAYERS FOR USE IN SPECIES MODELING:
Interpolation of maximum temperature in Oregon.

11-01-2012

Additional analyses

Benoit Parmentier

<u>What I did so far:</u>

New interpolation:

1. Ran Fusion with **all monthly stations** (193) using same models as presented in GAM1
2. Ran CAI with **all monthly stations** (193) using same models as presented in CAI1
3. Ran Fusion with **all monthly stations** (193)using same models as presented in GAM1 and constant sampling over 365 dates
4. Ran CAI with all monthly stations (191) using **simplified** models (called "CAI3), with into account **screening** of ELEV_SRTM and LST.
5. Ran CAI with all monthly stations (191) using same models as presented in CAI1, with into account **screening** of ELEV_SRTM and LST.

5. Ran Fusion with all monthly stations (1991) using same models as presented in GAM1 and **constant sampling** over 365 dates with **screening** of ELEV_SRTM and LST.
6. Running **Fusion** with all monthly stations (191) using **simplified models (called GAM4**), taking into account **screening** of ELEV_SRTM and LST.

<u>Method comparison:</u>

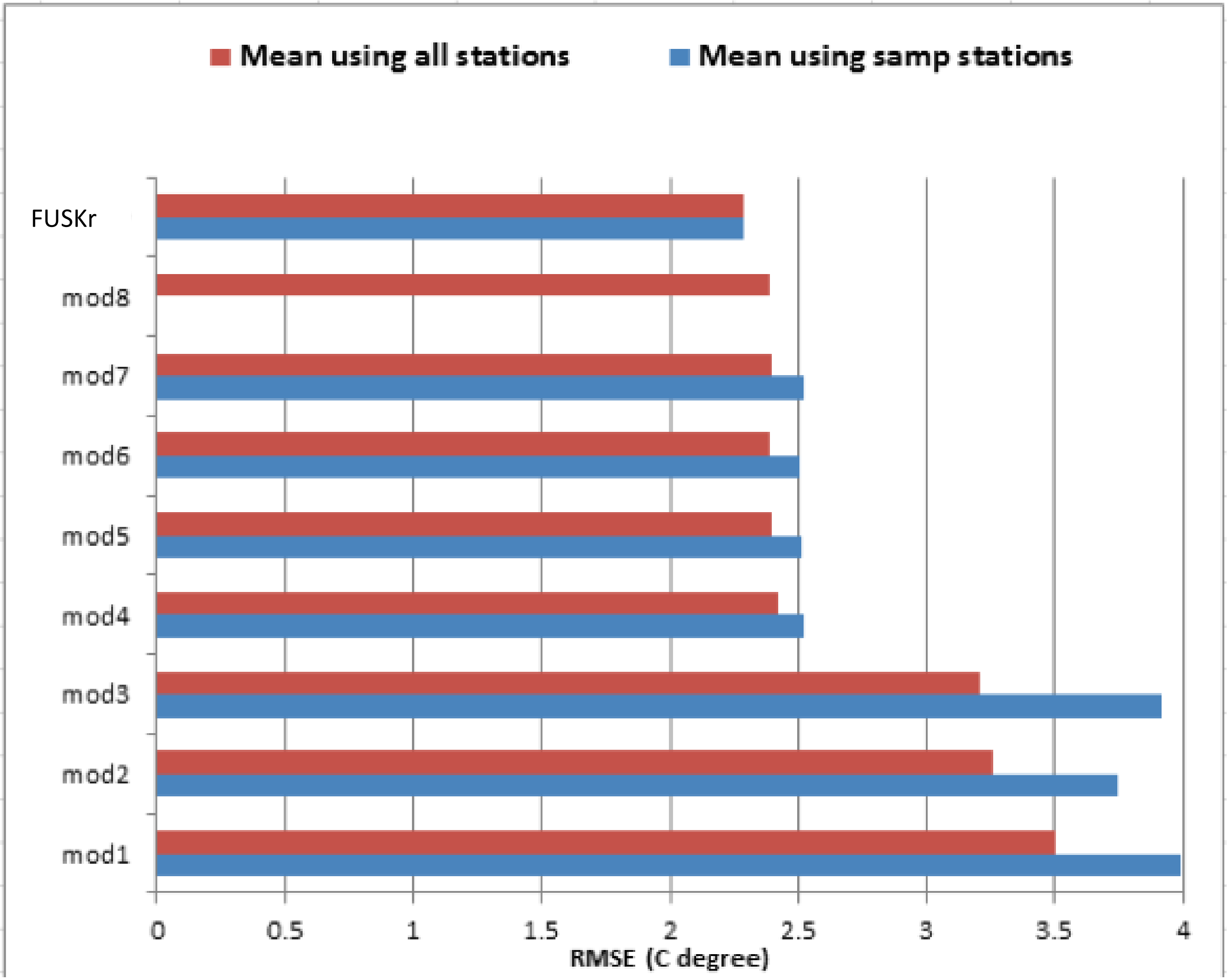
1. Comparing CAI2 and Fusion GAM1 using all monthly stations
2. Examining simplified model for CAI (called CAI3)
3. Examining specific residuals using constant sampling output from fusion GAM1
4. Examining LST, TMax and bias to see where extreme values occur.
5. Examining MAE and RMSE in term of season.

PART 1:

NEW INTERPOLATION USING ALL MONTHLY STATION FOR MONTHLY TIME STEPT

COMPARISON OF CAI AND FUSION WITH EARLIER RESULTS

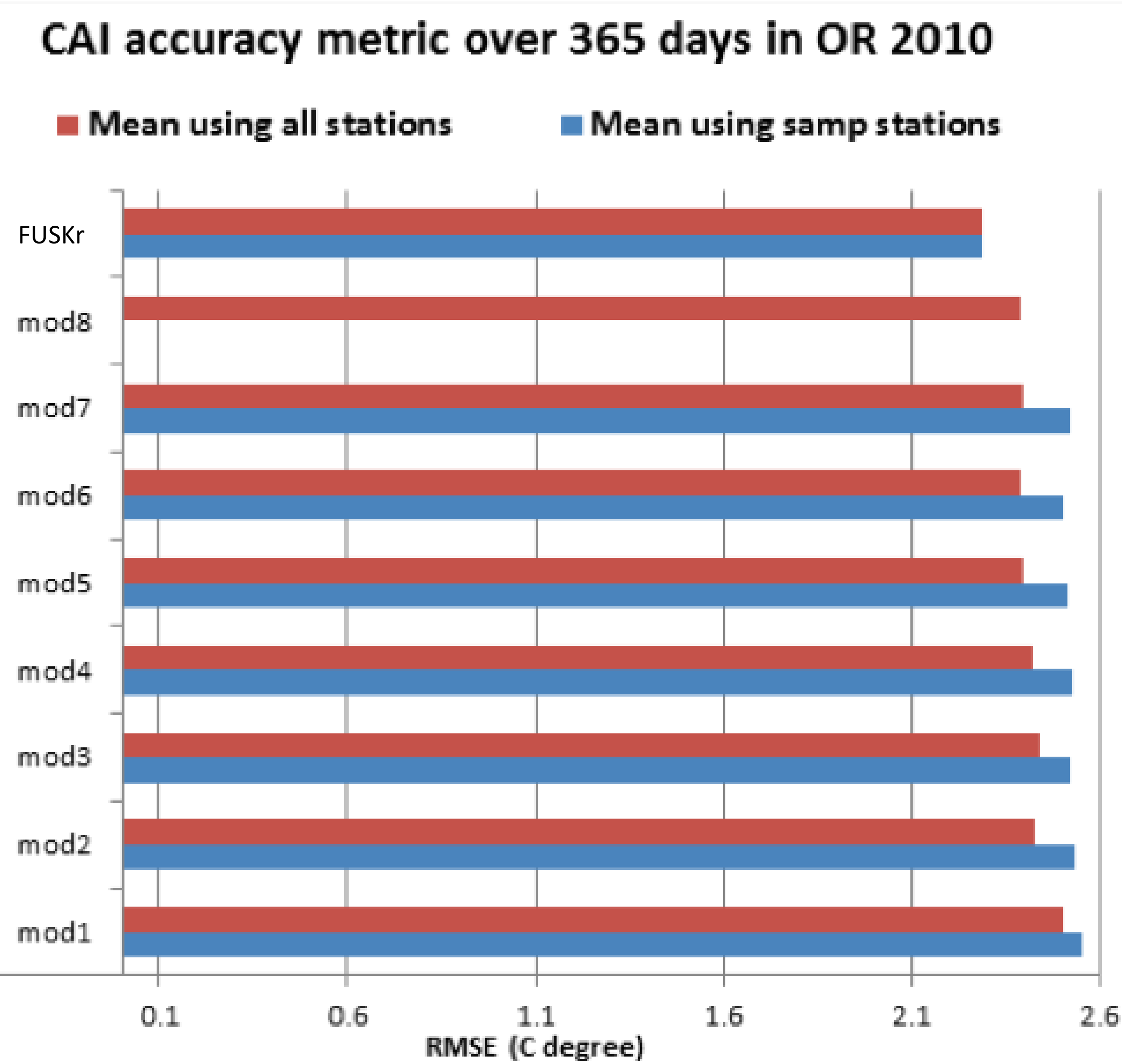
COMPARISON BETWEEN FUS WITH ALL STATIONS (red) AND DAILY STATIONS (SAMP in blue)



- Slight improvement when using more stations monthly for the fusion method using GAM.
- Adding more station has more impact for models with fewer covariates (mod1, mod2 and mod2
- Note that Fusion with Kriging for bias modeling remains the "best".

Model 1 to model 8 use GAM with covariates to model the bias surface. Models are described In GAM1.

COMPARISON BETWEEN CAI WITH ALL STATIONS (red) AND DAILY STATIONS (SAMP in blue)

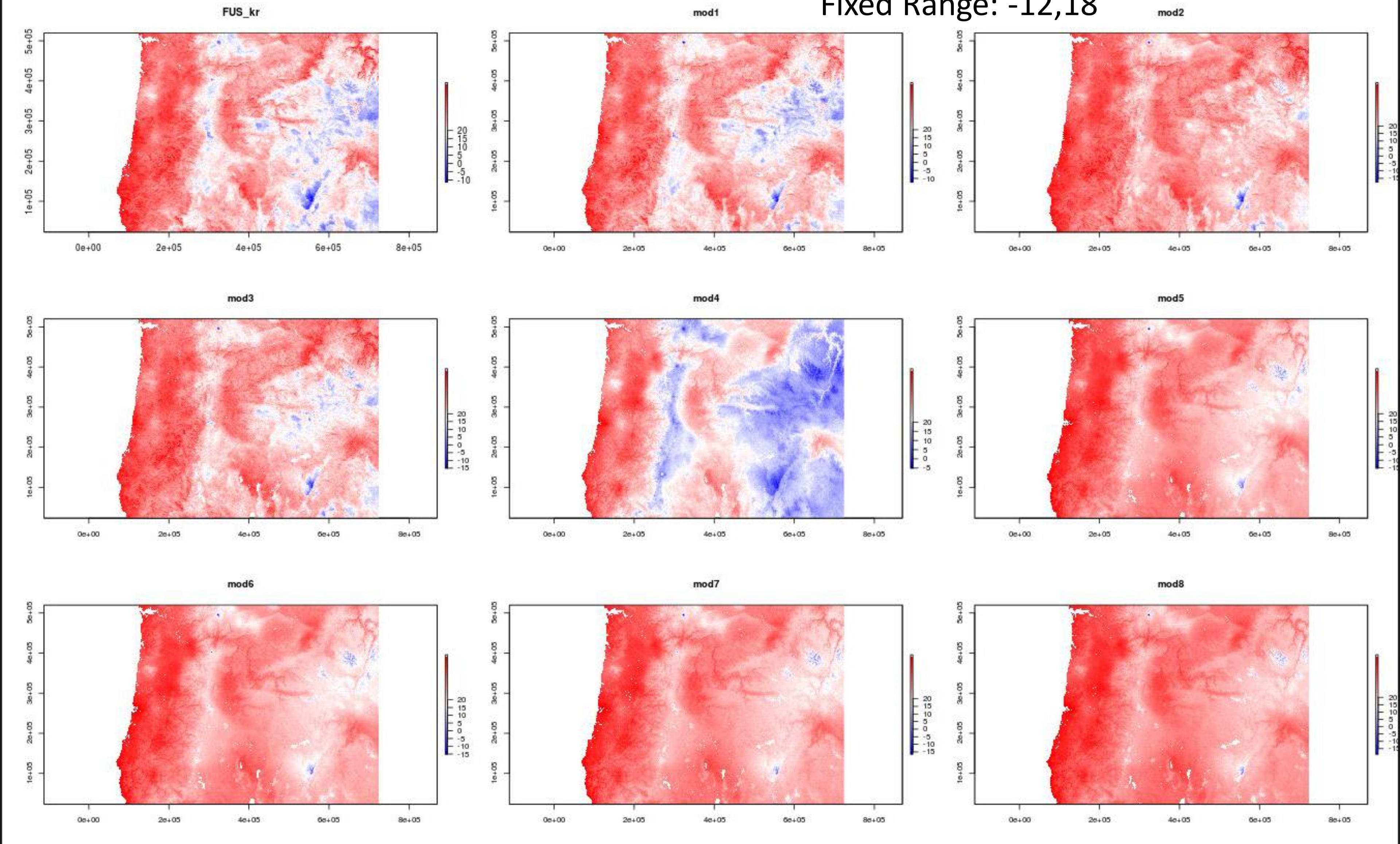


- Slight improvement when using more stations for CAI method.
- Note that Fusion with Kriged surface for bias remains the "best".
- Models perform similarly with a range of about 2.3 to 2.6 RMSE.

Model 1 to model 8 use GAM with covariates to model the bias surface. Models are described FUS1 and CAI1 (see previous slide).

FUSION all: Adding more observation for monthly step FOR JAN 3 2010

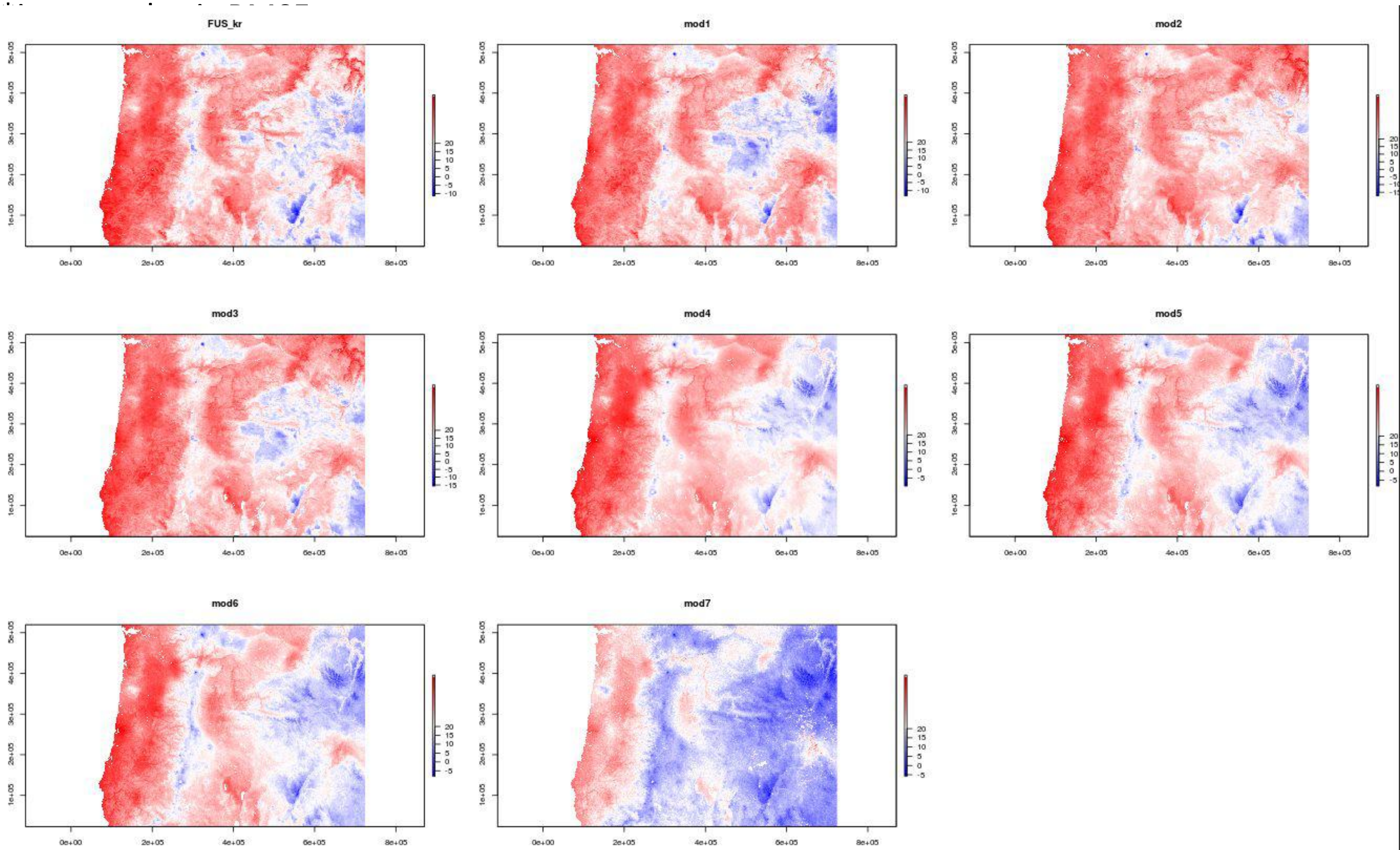
Fixed Range: -12,18



Average RMSE
For 365 dates:

	FUS_Kr	mod1	mod2	mod3	mod4	mod5	mod6	mod7	mod8
avg	2.29	3.50	3.26	3.20	2.42	2.40	2.39	2.39	2.39
sd	0.65	0.85	0.78	0.70	0.64	0.64	0.64	0.64	0.64

FUSION with sampled Daily observation for monthly step FOR JAN 3 2010



Average RMSE
For 365 dates:

	FUS_kr	mod1	mod2	mod3	mod4	mod5	mod6	mod7	mod8
avg	2.29	3.99	3.75	3.92	2.53	2.51	2.51	2.52	NA
sd	0.65	1.16	0.97	1.04	0.63	0.63	0.63	0.63	NA

COMPARISON BETWEEN CAI AND FUSION WITH ALL STATION MODELS USED IN CAI AND FUSION

FUS1 GAM

Model	Functional form
Fus_kr	Fusion with simple kriging
Mod1	LST_bias ~ f(lat) + f(lon) + f(ELEV_SRTM)
Mod2	LST_bias ~ f(lat,lon)+ f(ELEV_SRTM)
Mod3	LST_bias ~ f(lat) + s (lon) + s (ELEV_SRTM) + s (Northness)+ s (Eastness) + f(DISTOC)
Mod4	LST_bias ~ f(lat) + s (lon) + f(ELEV_SRTM) + f(Northness) + s (Eastness) + f(DISTOC) + f(LST)
Mod5	LST_bias ~ f(lat,lon) +f(ELEV_SRTM) + f(Northness,Eastness) + f(DISTOC) + f(LST)
Mod6	LST_bias ~ f(lat,lon) +f(ELEV_SRTM) + f(Northness,Eastness) + f(DISTOC) + f(LST)+f(LC1)
Mod7	LST_bias ~ f(lat,lon) +f(ELEV_SRTM) + f(Northness,Eastness) + f(DISTOC) + f(LST)+f(LC3)
Mod8	LST_bias ~ f(lat,lon) +f(ELEV_SRTM) + f(Northness,Eastness) + f(DISTOC) + f(LST) + f(LC1,LC3)

CAI1 with GAM

Model	Functional form
CAI_kr	CAI with simple kriging
Mod1	TMax ~ f(lat) + f(lon) + f(ELEV_SRTM)
Mod2	TMax~ f(lat,lon)+ f(ELEV_SRTM)
Mod3	TMax~ f(lat) + s (lon) + s (ELEV_SRTM) + s (Northness)+ s (Eastness) + f(DISTOC)
Mod4	TMax~ f(lat) + s (lon) + f(ELEV_SRTM) + f(Northness) + s (Eastness) + f(DISTOC) + f(LST)
Mod5	TMax~ f(lat,lon) +f(ELEV_SRTM) + f(Northness,Eastness) + f(DISTOC) + f(LST)
Mod6	TMax~ f(lat,lon) +f(ELEV_SRTM) + f(Northness,Eastness) + f(DISTOC) + f(LST)+f(LC1)
Mod7	TMax~ f(lat,lon) +f(ELEV_SRTM) + f(Northness,Eastness) + f(DISTOC) + f(LST)+f(LC3)
Mod8	TMax~ f(lat,lon) +f(ELEV_SRTM) + f(Northness,Eastness) + f(DISTOC) + f(LST) + f(LC1,LC3)

MODEL COMPARISON BETWEEN CAI AND FUSION

CAI and FUSION models give nearly same results when bias is modeled with LST in the covariates. This is due to the fact that BIAS= Tmax (monthly)-LST so there is linear dependence.

```
> head(mod7$model)
      y_var     lat       lon ELEV_SRTM  Northness  Eastness    DISTOC      LST LC3
1  5.705538 42.1761 -119.8961      1400 -0.90262747  0.4304226 358792.75 280.0207  51
19 6.037294 42.9694 -119.9933      1297  0.06831702 -0.9976637 350663.00 276.9052  16
28 8.464407 44.4044 -123.7533        78 -0.47911944  0.8777497  26887.09 277.9007   0
38 4.866667 44.8197 -120.7533       921  0.77743951 -0.6289577 257909.94 275.8346  51
54 5.692375 45.7211 -120.2064        80  0.15552745 -0.9878316 261590.95 276.8153   8
70 9.692625 42.2128 -122.7144       532 -0.98148218 -0.1915535 127372.23 279.1476   7
> head(mod7_f$model)
       y_var     lat       lon ELEV_SRTM  Northness  Eastness    DISTOC      LST LC3
1   1.155152 42.1761 -119.8961      1400 -0.90262747  0.4304226 358792.75 280.0207  51
19 -2.292112 42.9694 -119.9933      1297  0.06831702 -0.9976637 350663.00 276.9052  16
28 -3.723680 44.4044 -123.7533        78 -0.47911944  0.8777497  26887.09 277.9007   0
38 -2.192072 44.8197 -120.7533       921  0.77743951 -0.6289577 257909.94 275.8346  51
54 -2.037037 45.7211 -120.2064        80  0.15552745 -0.9878316 261590.95 276.8153   8
70 -3.705073 42.2128 -122.7144       532 -0.98148218 -0.1915535 127372.23 279.1476   7
> |
```

Y_var= TMax

```
GCV score: 0.8195932
> mod7

Family: gaussian
Link function: identity

Formula:
y_var ~ s(lat, lon) + s(ELEV_SRTM) + s(Northness, Eastness) +
    s(DISTOC) + s(LST) + s(LC3)

Estimated degrees of freedom:
21.85  7.57  2.00  2.69  2.56  3.92  total = 41.59

GCV score: 0.8195932
> |
```

Y_var= LSTD_bias

```
GCV score: 0.8195932
> mod7_f

Family: gaussian
Link function: identity

Formula:
y_var ~ s(lat, lon) + s(ELEV_SRTM) + s(Northness, Eastness) +
    s(DISTOC) + s(LST) + s(LC3)

Estimated degrees of freedom:
21.85  7.57  2.00  2.69  2.56  3.92  total = 41.59

GCV score: 0.8195933
>
```

ALTERNATIVE SIMILAR MODELS FOR FUSION

- Adding LST on the left-hand side may not make sense in statistical sense but it does improve models in the Fusion+GAM predictions.

- I found that only model 1, model 2 and model 3 and model 9 were different than CAI when LST is used as covariate to model LST bias. I assume this must be due to the linear dependence.

- I added a few simple models that do not include LST in the covariates for the modelling of the LST_bias. I do not expect these models to perform better given results from model 1, model 2 and model 3.

Here are the models:

```
formula1 <- as.formula("y_var ~ s(ELEV_SRTM)", env=.GlobalEnv)
formula2 <- as.formula("y_var ~ s(lat,lon)", env=.GlobalEnv)
formula3 <- as.formula("y_var~ s(lat,lon,ELEV_SRTM)", env=.GlobalEnv)
formula4 <- as.formula("y_var~ s(lat) + s (lon) + s (ELEV_SRTM) + s(DISTOC)", env=.GlobalEnv)
formula5 <- as.formula("y_var~ s(lat,lon,ELEV_SRTM) + s(Northness) + s (Eastness) + s(DISTOC)", env=.GlobalEnv)
formula6 <- as.formula("y_var~ s(lat,lon) +s(ELEV_SRTM) + s(Northness,Eastness) + s(DISTOC)", env=.GlobalEnv)
```

→ **From the results, it is clear that Fusion+Kriging and CAI+Kriging are still the “best” based on RMSE values. CAI has however alternative models that might be selected because their spatial structure makes sense.**

PART 2: SIMPLIFY MODEL OF COVARIATES

CAI, modeling of monthly Tmax using ELEV_SRTM and LST as covariates.

CAI3: SIMPLIFIED MODELS

```
list_formulas[[1]] <- as.formula("y_var~ s(ELEV_SRTM)", env=.GlobalEnv)
list_formulas[[2]] <- as.formula("y_var~ s(LST)", env=.GlobalEnv)
list_formulas[[3]] <- as.formula("y_var~ s(LST) + s(ELEV_SRTM)", env=.GlobalEnv)
list_formulas[[4]] <- as.formula("y_var~ s(LST,ELEV_SRTM)", env=.GlobalEnv)
list_formulas[[5]] <- as.formula("y_var~ s(lat,lon,ELEV_SRTM)", env=.GlobalEnv)
```

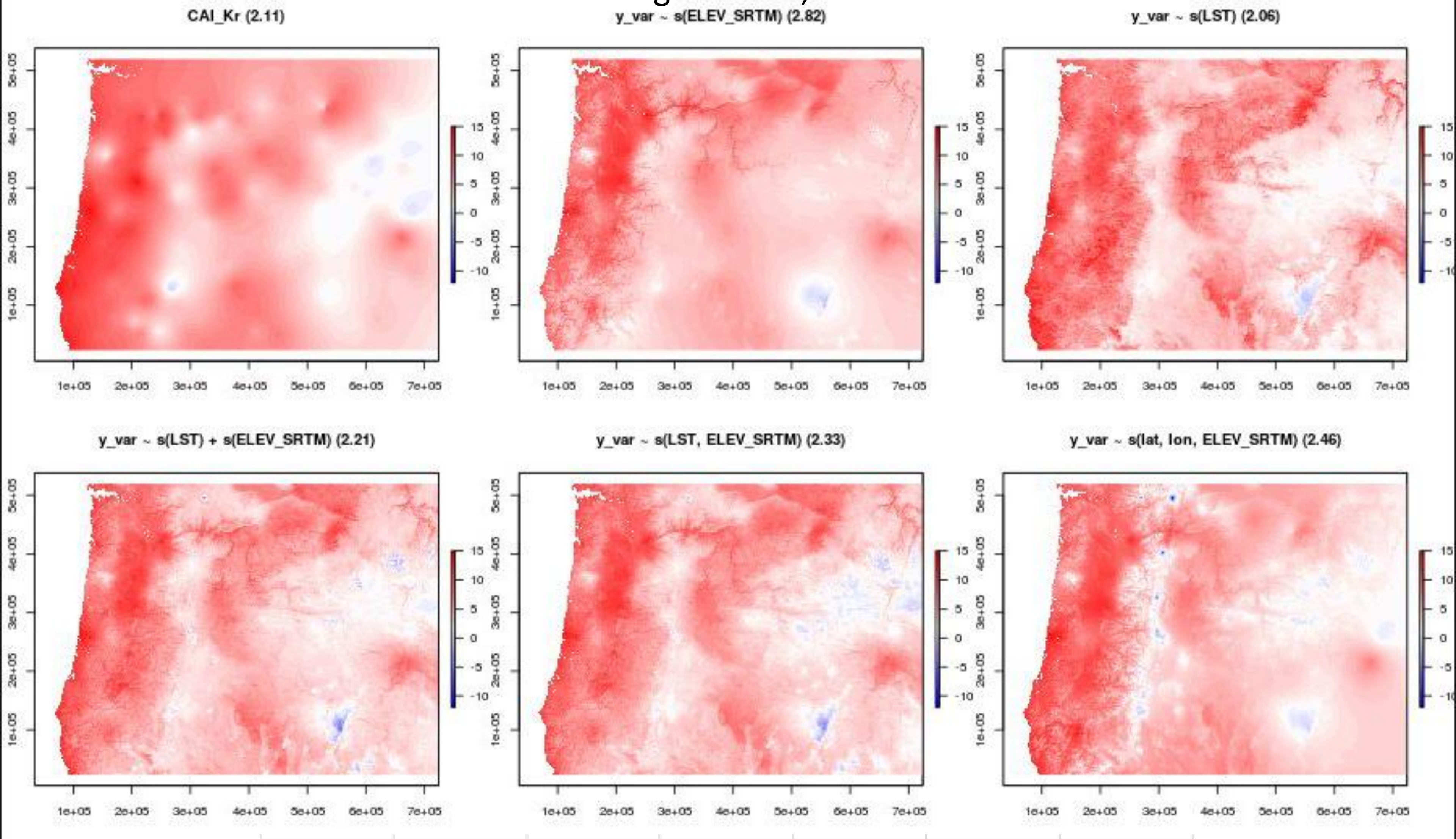
These models were added following the meeting on Wednesday 10/24/2102. The goal is to see how well simple models of Tmax (monthly) with a few covariates (LST and ELEV_SRTM) perform.

Note that we screened out LST and ELEV SRTM values. See part III for more details.

CAI3: SIMPLIFIED MODELS-MAPS FOR JAN 3 2010

*in parenthesis:RMSE

Range: -11.93,14.82



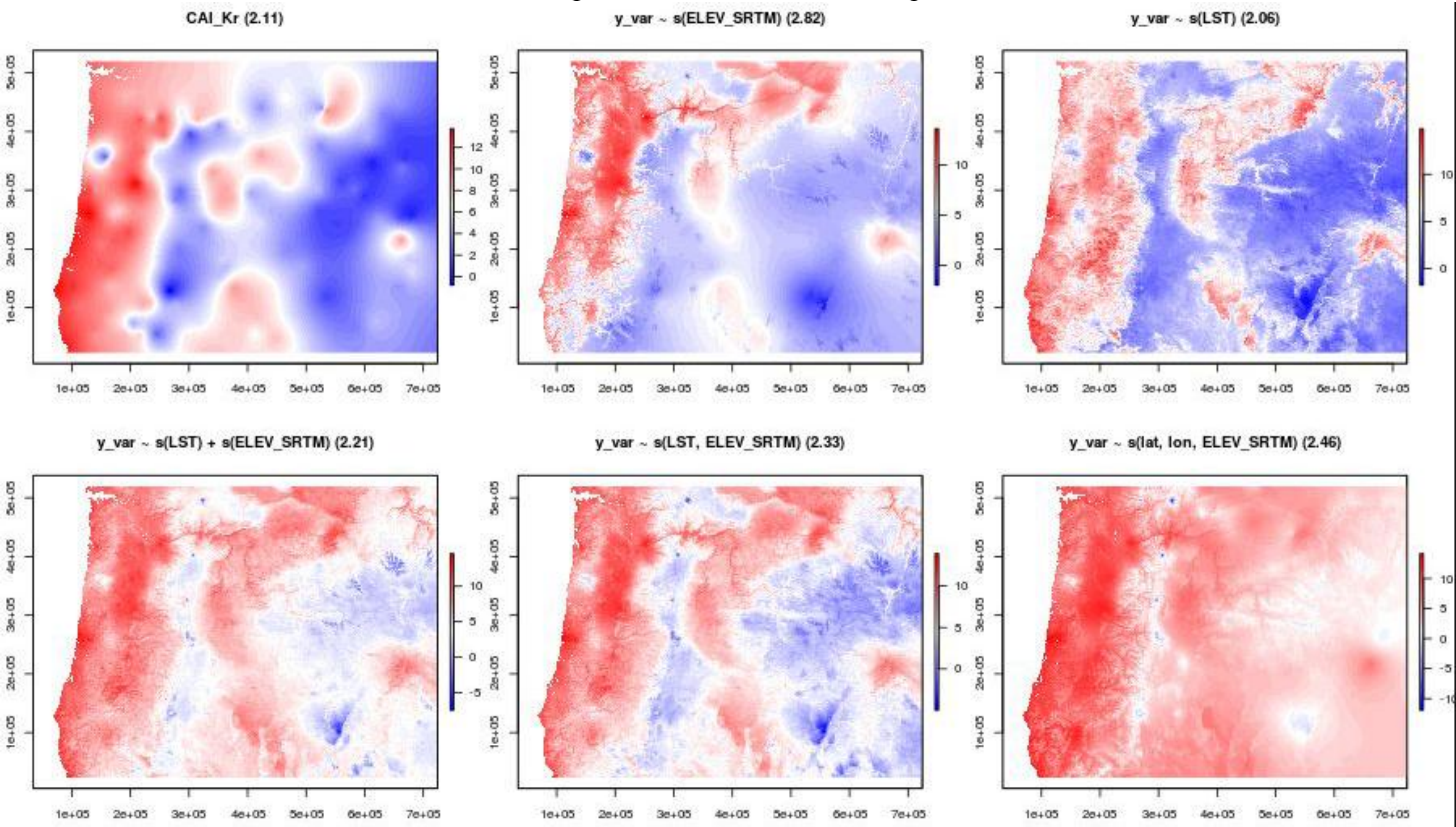
Average RMSE
For 365 dates:

	CAI_Kr	mod1	mod2	mod3	mod4	mod5
avg	2.29	3.13	2.87	2.68	2.69	2.59
sd	0.65	0.73	0.61	0.62	0.62	0.64

CAI3: SIMPLIFIED MODELS-MAPS FOR JAN 3 2010

*in parenthesis:RMSE

Ranges varies for each image

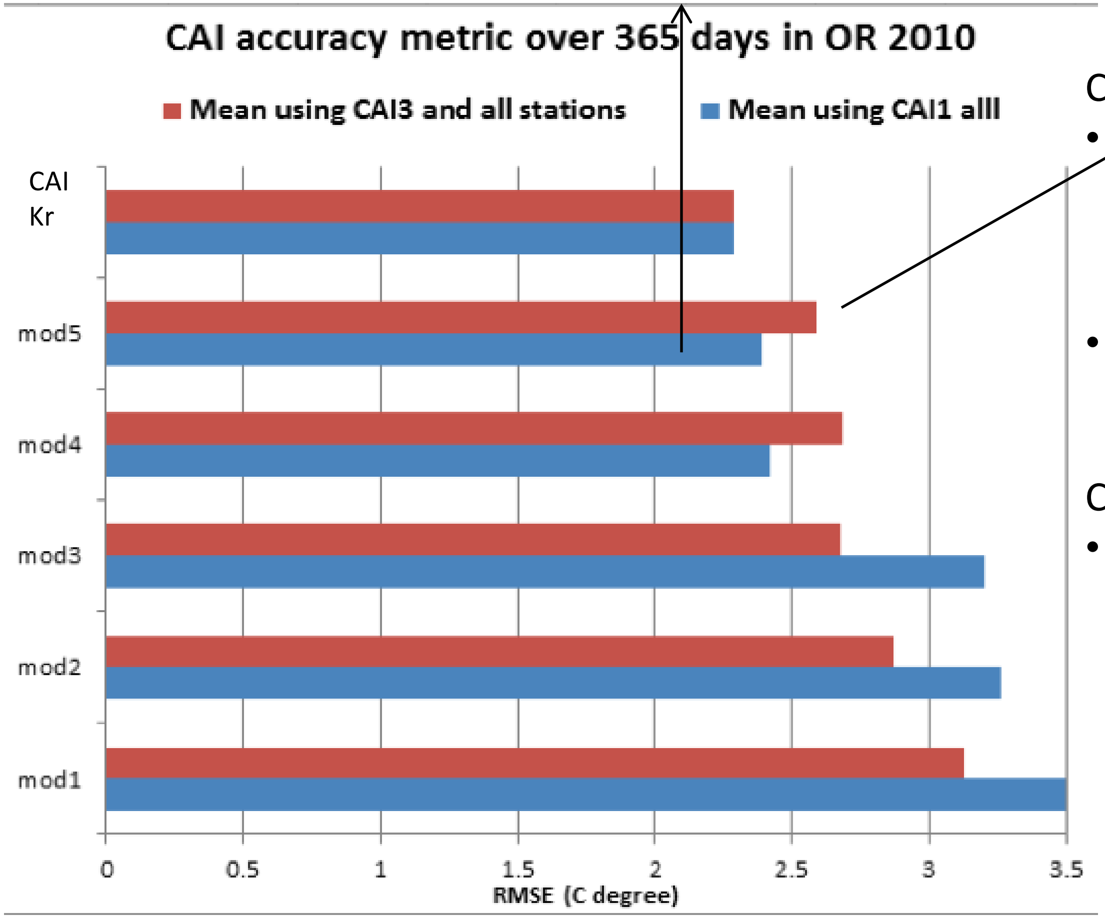


Average RMSE
For 365 dates:

	CAI_Kr	mod1	mod2	mod3	mod4	mod5
avg	2.29	3.13	2.87	2.68	2.69	2.59
sd	0.65	0.73	0.61	0.62	0.62	0.64

CAI3: SIMPLIFIED MODELS

models	CAI1 alll station (in blue)	CAI 3 simplified models, all station (in red)
mod1	Tmax~ f(lat) + f(lon) + f(ELEV_SRTM)	Tmax~ fELEV_SRTM)
Mod2	Tmax~ f(lat,lon) + f(ELEV_SRTM)	Tmax~ f(LST)
Mod3	Tmax~ f(lat) + f(lon) + f(ELEV_SRTM) + f (Northness) + f (Eastness) + f(DISTOC)	Tmax~ f(LST) + f(ELEV_SRTM)
mod4	Tmax~ f(lat) +f(lon) + f(ELEV_SRTM) + f (Northness) + f (Eastness) + f(DISTOC)+ f(LST)	Tmax~ f(LST,ELEV_SRTM)
mod5	Tmax~ f(lat,lon) + f(ELEV_SRTM) + f (Northness,Eastness) + f(DISTOC)+ f(LST)	Tmax~ f(lat,lon,ELEV_SRTM)



CAI3:

- **CAI with Kriging has the lowest mean RMSE (2.29C)** with the second best being three way model corresponding to WorldClim (mod5).
- Model including only elevation (mod1 has Higher RMSE but lower than when

CAI1:

- When using many covariates (CAI1): the best model presented here is model5 which include aspect, lat-lon and distance to ocean (DISTOC) as additional variables.

PART 3: Screening of extreme values in space and time…

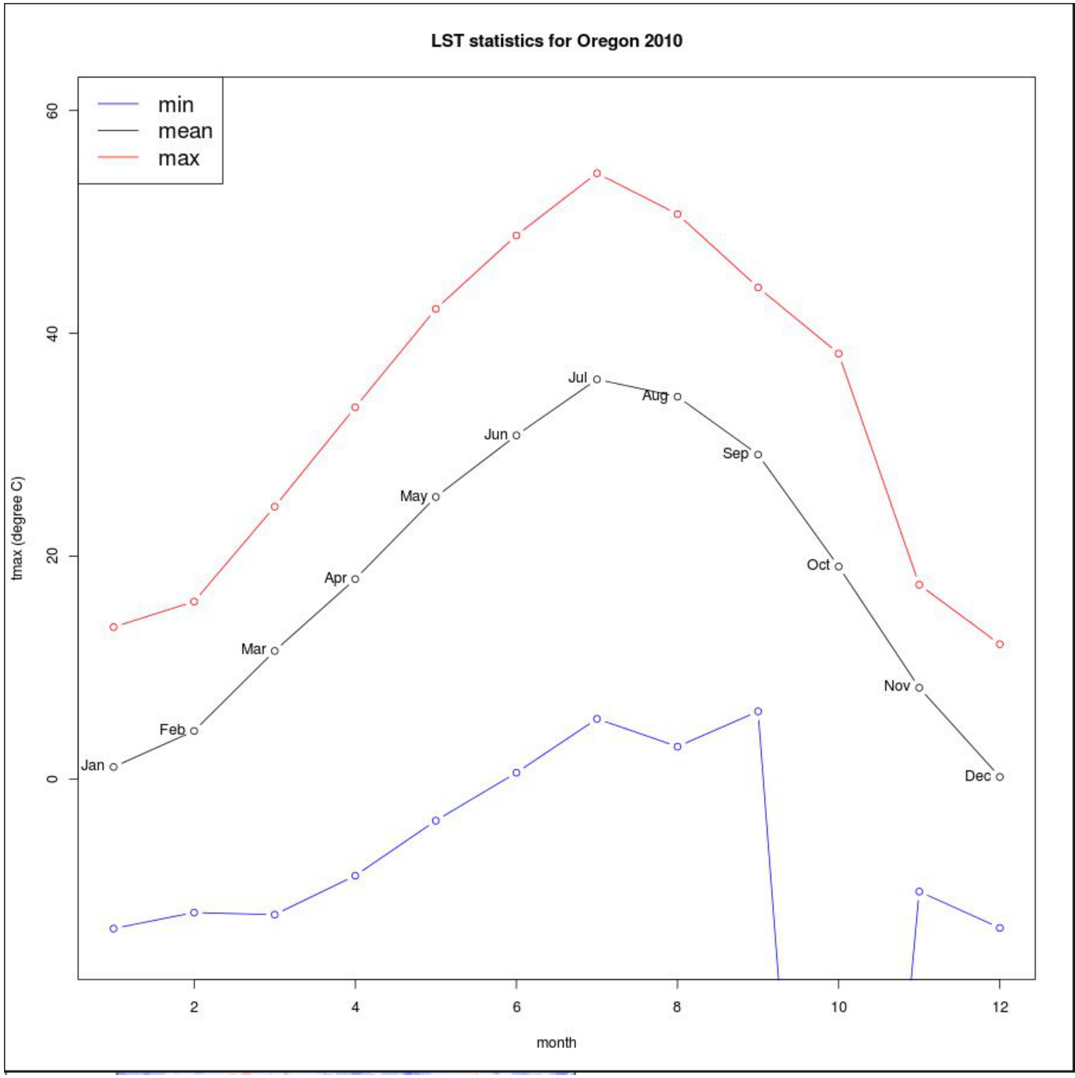
Analysis of LST, Tmax (monthly), LST bias to detect extreme values in space and time.

LST STATICS FOR STACK…molst

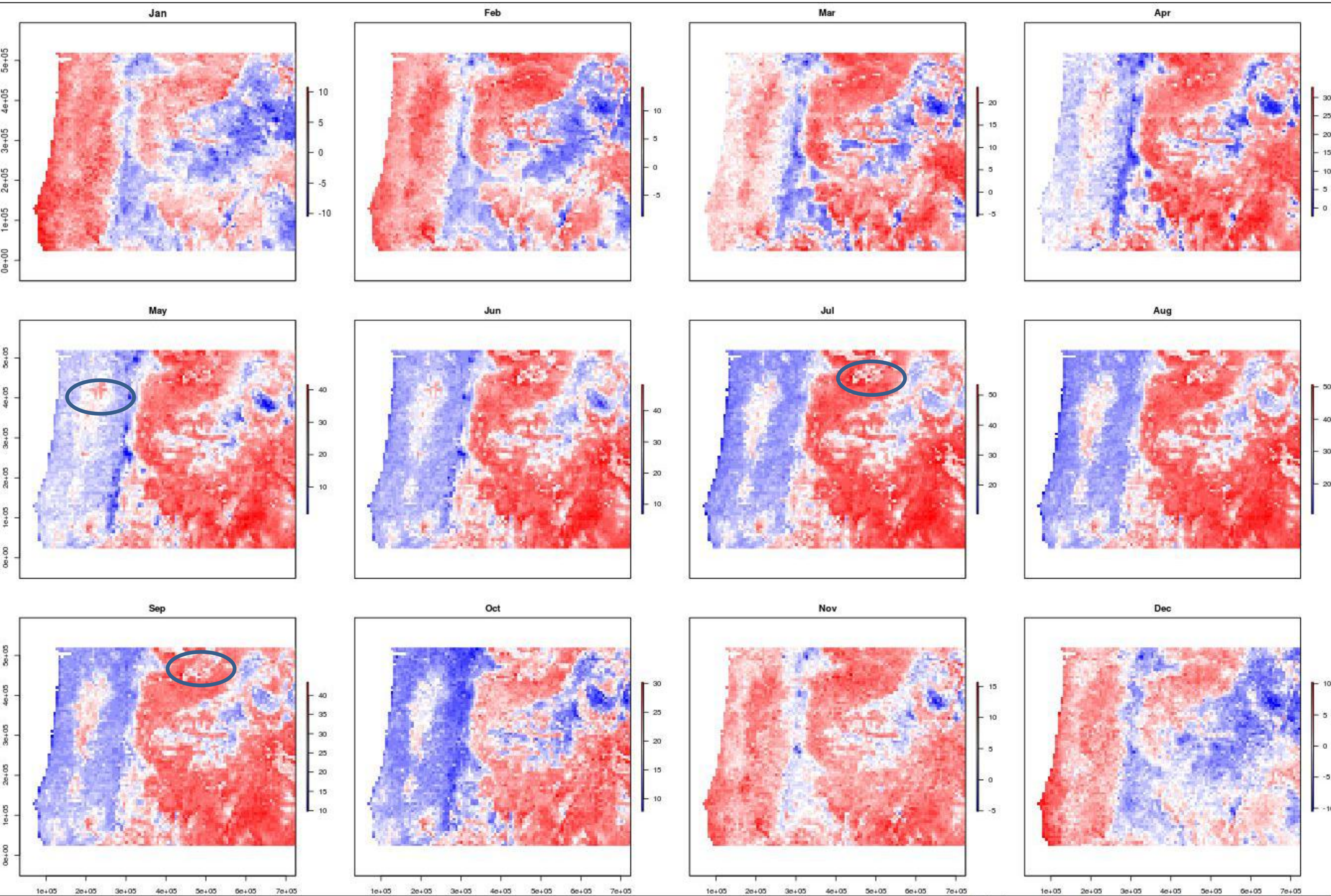
Statistics show that screening is needed for Oct 2010 LST image because its minimum value (-80C) does not follow the expected temporal pattern.

Variability is the highest in summer with Peak of standard deviation at 9.24C in July When the mean is the highest.

	row.names	min_values	max_values	mean_values	sd_values
1	Jan	-13.4258997	13.63272	1.0776955	3.675844
2	Feb	-11.9909937	15.92453	4.3129347	4.093666
3	Mar	-12.1752893	24.44559	11.4947797	5.010748
4	Apr	-8.6800195	33.35834	17.9455053	6.593789
5	May	-3.7420618	42.19107	25.3136717	7.280559
6	Jun	0.5594519	48.78748	30.8430454	8.210558
7	Jul	5.3999976	54.37076	35.8670969	9.241359
8	Aug	2.8999976	50.69645	34.3121028	8.442812
9	Sep	6.0661658	44.11356	29.1146272	6.575432
10	Oct	-88.6400110	38.18000	19.0651518	4.567670
11	Nov	-10.1062280	17.43656	8.1888230	2.659000
12	Dec	-13.3681299	12.09998	0.1852424	3.338995



LST MONTHLY MEAN

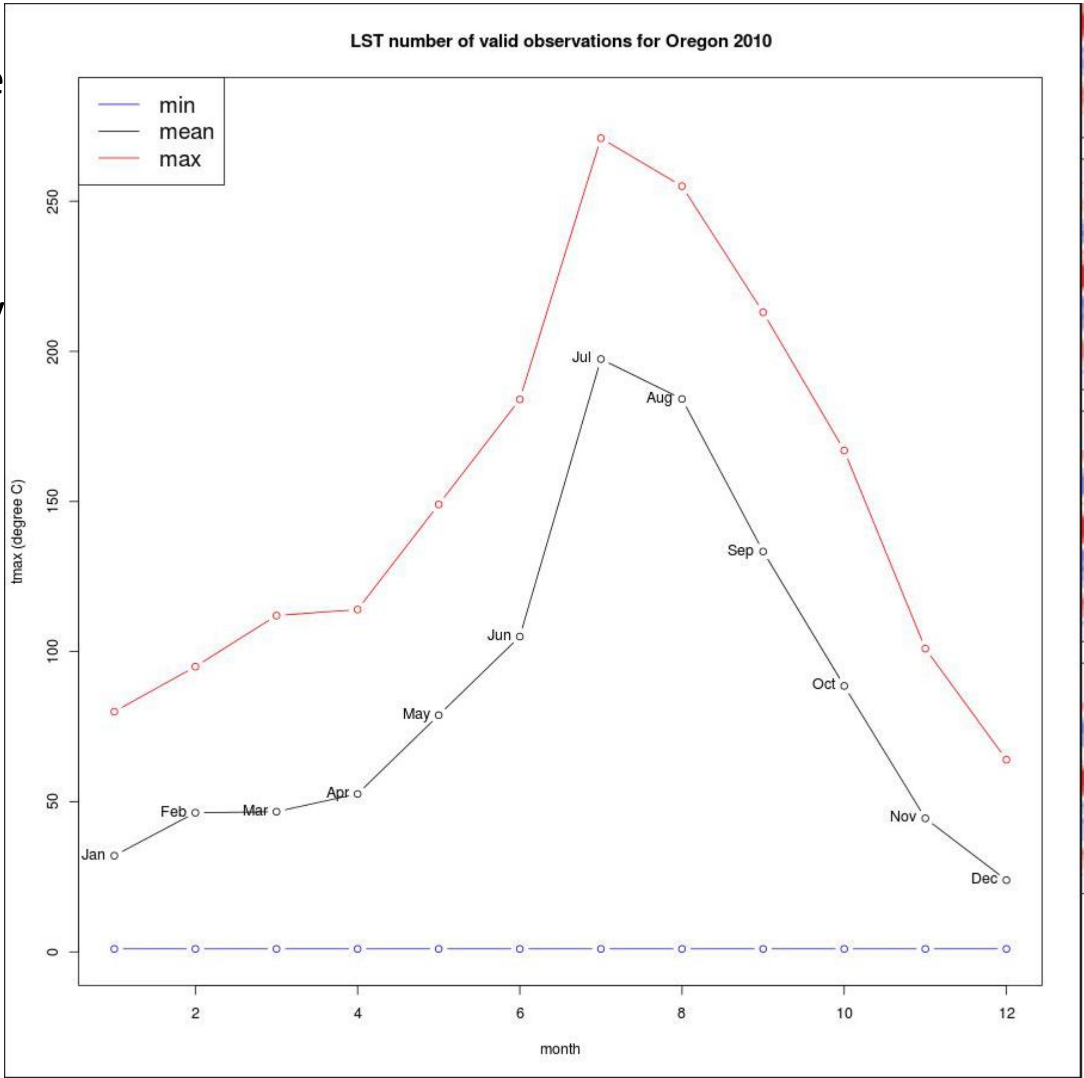


Note the inversion of temperature on the coast compared to inland.

LST STATICS FOR NUMBER OF VALID OBSERVATION

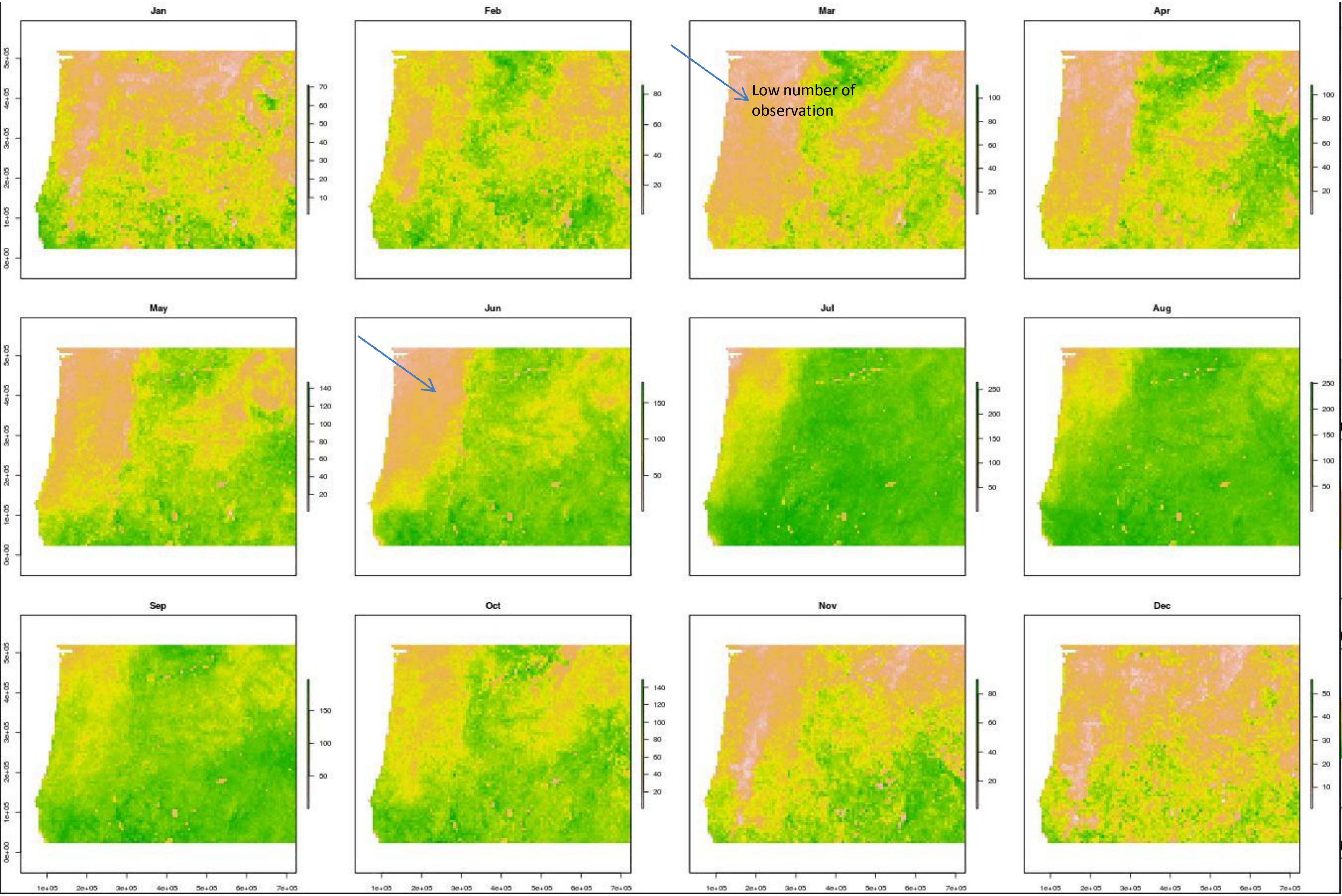
Statistics show that screening is needed for Oct 2010 LST image because its minimum value (-80C) does not follow the expected temporal pattern.

Variability is the highest in summer with Peak of standard deviation at 9.24C in July When the mean is the highest.



	row.names	min_values	max_values	mean_values	sd_values
1	Jan	1	80	32.04959	10.552782
2	Feb	1	95	46.36265	11.372994
3	Mar	1	112	46.69087	17.282301
4	Apr	1	114	52.60424	17.549804
5	May	1	149	78.83438	22.000066
6	Jun	1	184	105.02751	29.166917
7	Jul	1	271	197.43089	36.169408
8	Aug	1	255	184.14487	30.645962
9	Sep	1	213	133.34210	22.594606
10	Oct	1	167	88.57262	18.849616
11	Nov	1	101	44.48281	15.039668
12	Dec	1	64	23.90236	7.891211

LST NUMBER OF VALID OBSERVATION OVER 10 YEARS AND BY MONTH (~310 obs. max)



There are fewer observations in the Northwest part of the Oregon State and in Winter.

SCREENING OF MONTHLY MEAN LST and ELEVATION

Frequency in October LST

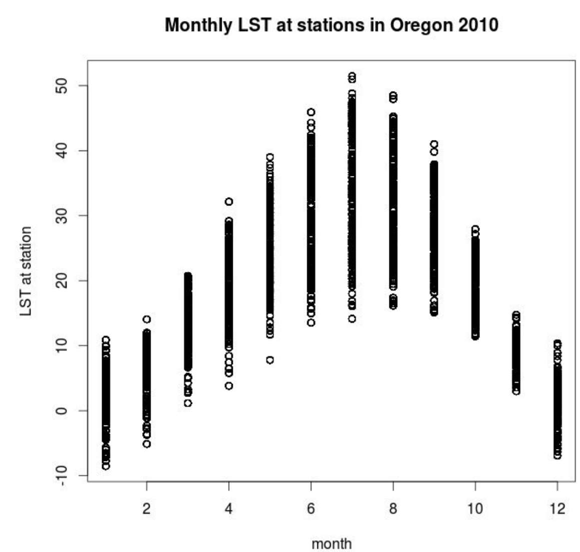
```
     value count
[1,]   -89     3
[2,]   -88     1
[3,]   -87     1
[4,]   -86     1
[5,]    -1     2
[6,]     0     6
```

- Taking into account the cell statistics, I screened out all values less than -80C LST in LST for October. Since there is only a total of 6 pixels with unusual extreme values (less than -80C) in the October LST image, I expect little effect on the modeling. ..

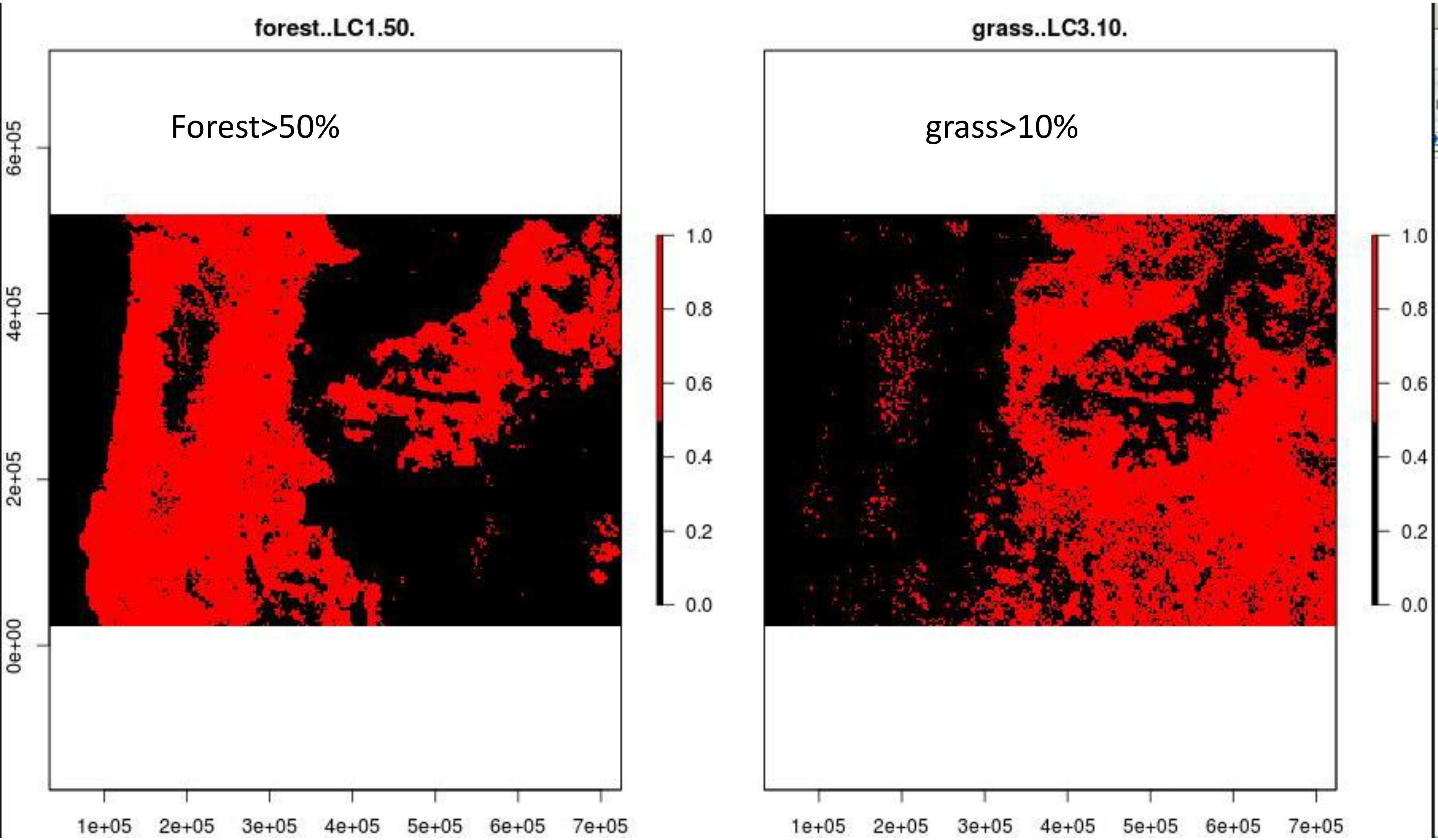
Statistics from monthly mean LST

	row.names	min_values	max_values	mean_values	sd_values
1	Jan	-13.4258997	13.63272	1.0776955	3.675844
2	Feb	-11.9909937	15.92453	4.3129347	4.093666
3	Mar	-12.1752893	24.44559	11.4947797	5.010748
4	Apr	-8.6800195	33.35834	17.9455053	6.593789
5	May	-3.7420618	42.19107	25.3136717	7.280559
6	Jun	0.5594519	48.78748	30.8430454	8.210558
7	Jul	5.3999976	54.37076	35.8670969	9.241359
8	Aug	2.8999976	50.69645	34.3121028	8.442812
9	Sep	6.0661658	44.11356	29.1146272	6.575432
10	Oct	$\boxed{-88.6400110}$	38.18000	19.0651518	4.567670
11	Nov	-10.1062280	17.43656	8.1888230	2.659000
12	Dec	-13.3681299	12.09998	0.1852424	3.338995

- No screening was done for upper (maximum) values because the temporal pattern makes sense. It appears that LST overestimate monthly maximum temperature (TMax) in summer but there are differences in the land cover types (see following slides).

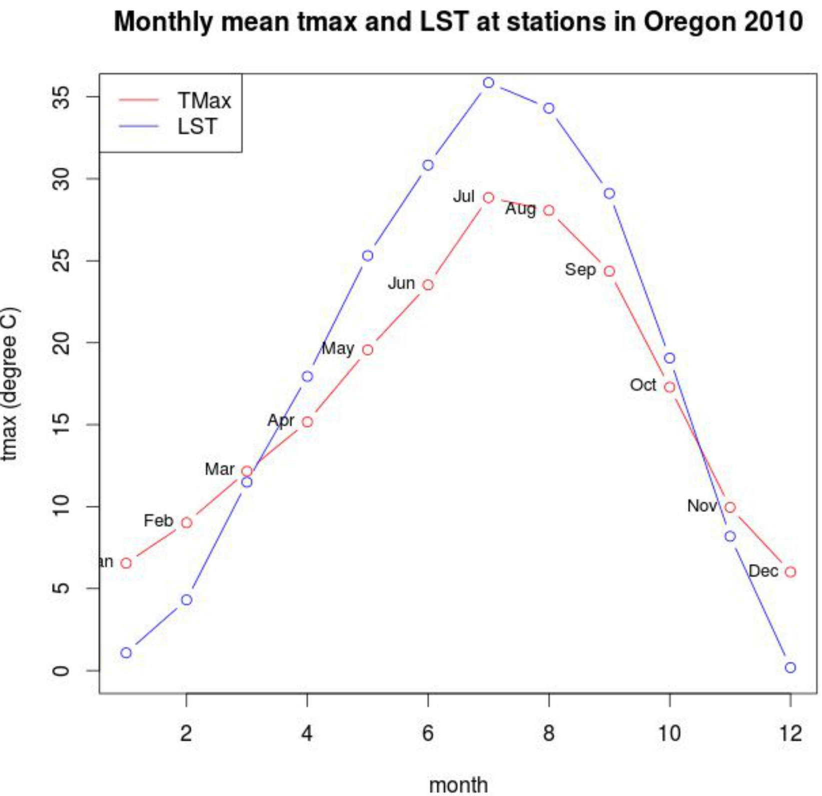


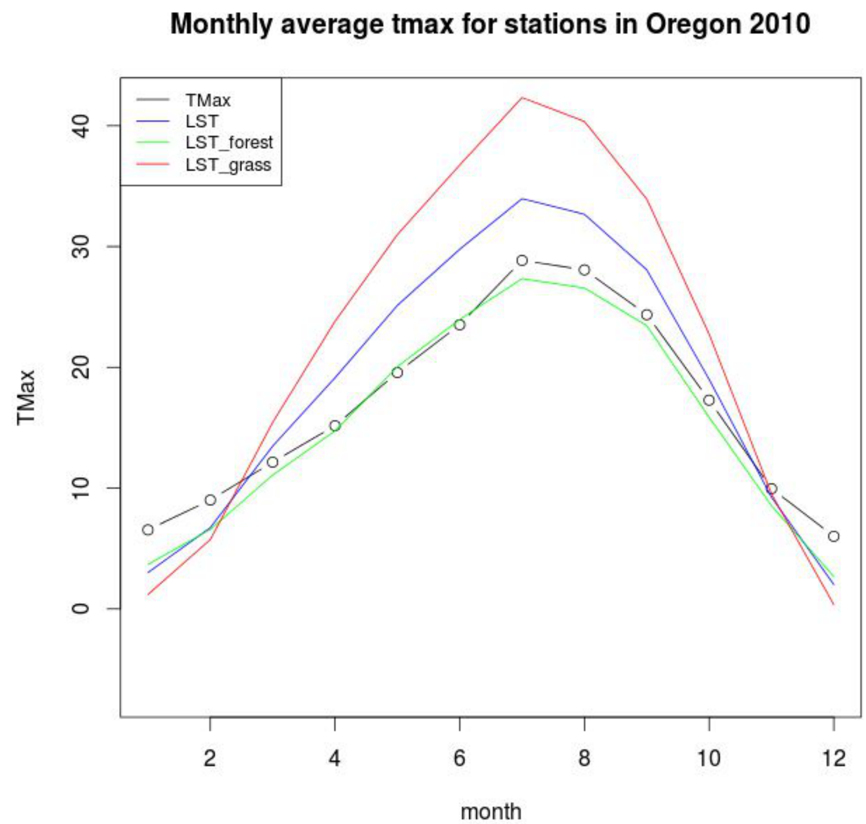
LST BIAS AND LAND COVER



Overlap: ~10,000pixels or 4%

LST AND BIAS: TEMPORAL PROFILES



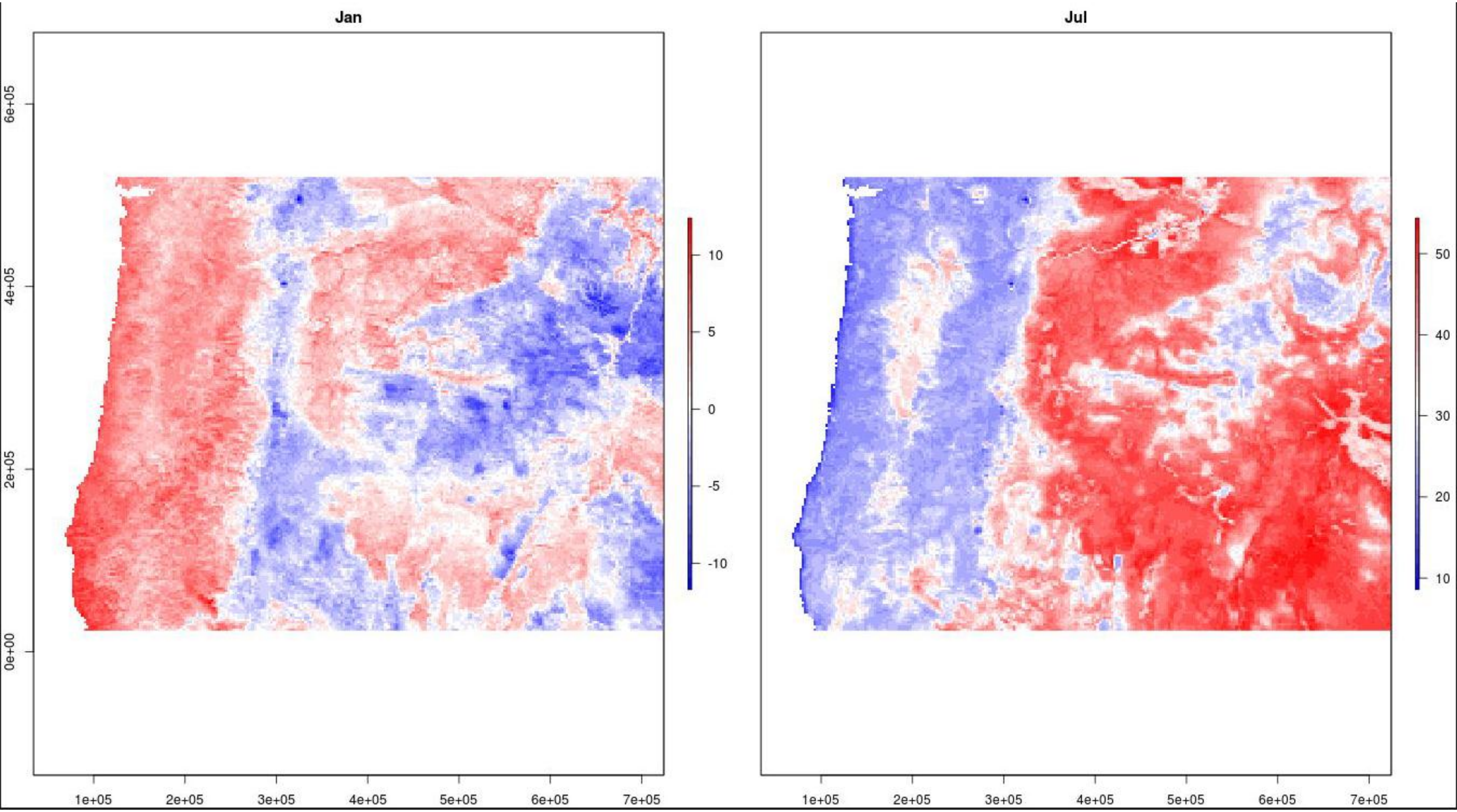


Plots of monthly mean at station show that on average that LST is less than Tmax in Winter and greater than Tmax in summer.

Bias is also influenced greatly by land cover types with:

- LST showing slightly lower temperature than Tmax in summer for forest cover
- LST showing strongly higher temperature than Tmax in summer for grass cover.

LST SPATIAL PATTERN

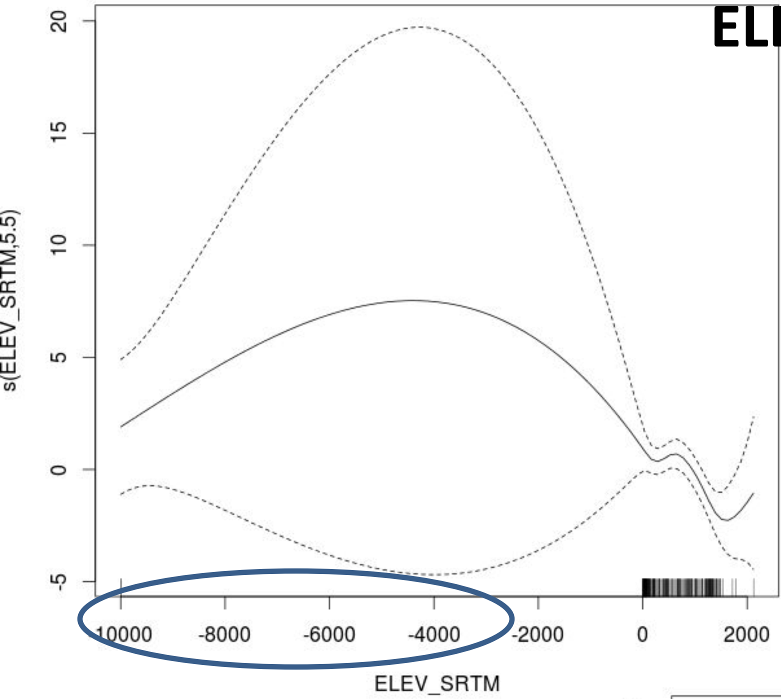


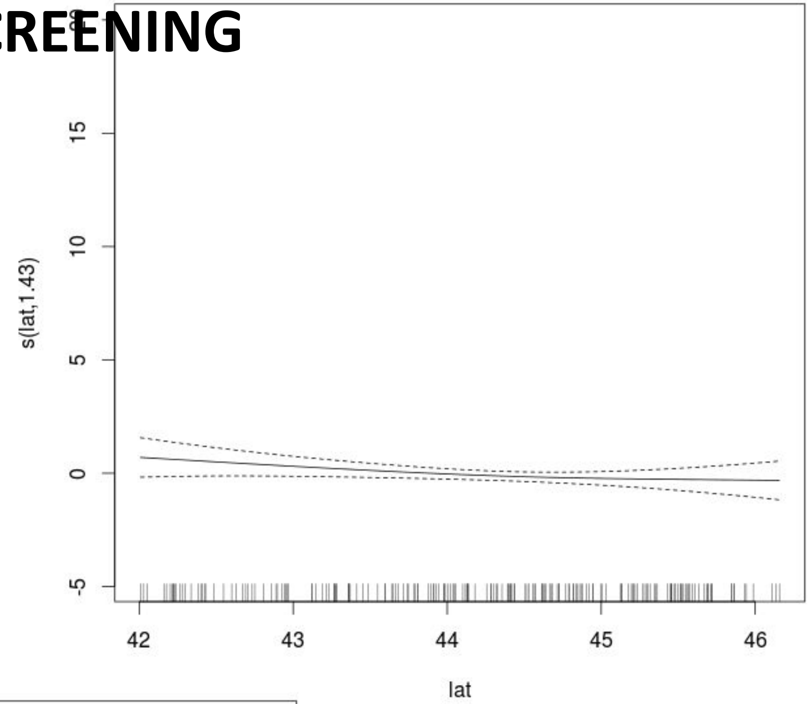
- Spatial patterns in the LST images also make sense with:
- Forest areas cooler than surrounding areas in Summer,
- Area near the coast warmer in Winter
- Valley and crop area standing out in July.

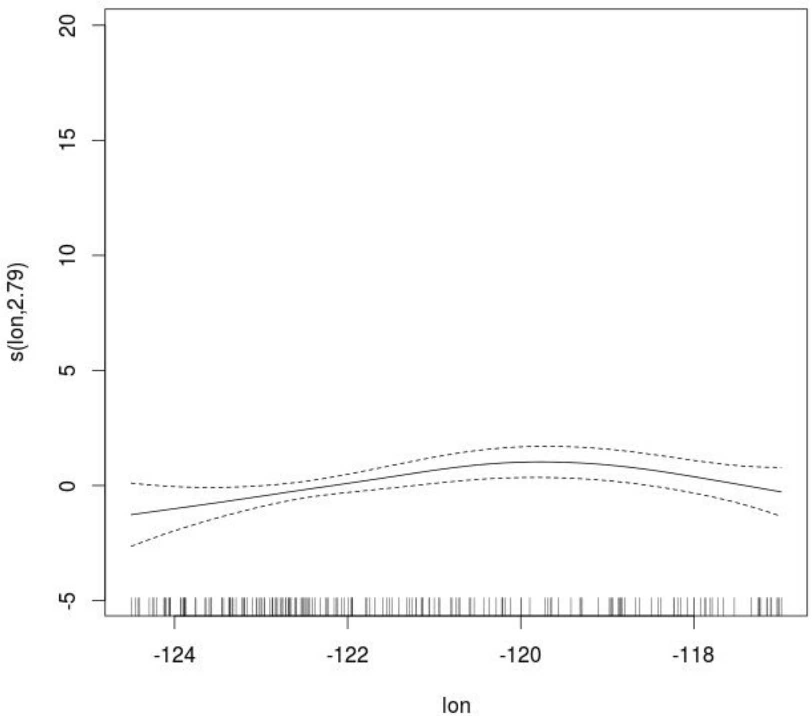
ELEVATION SCREENING

Before screening …

Mod1:



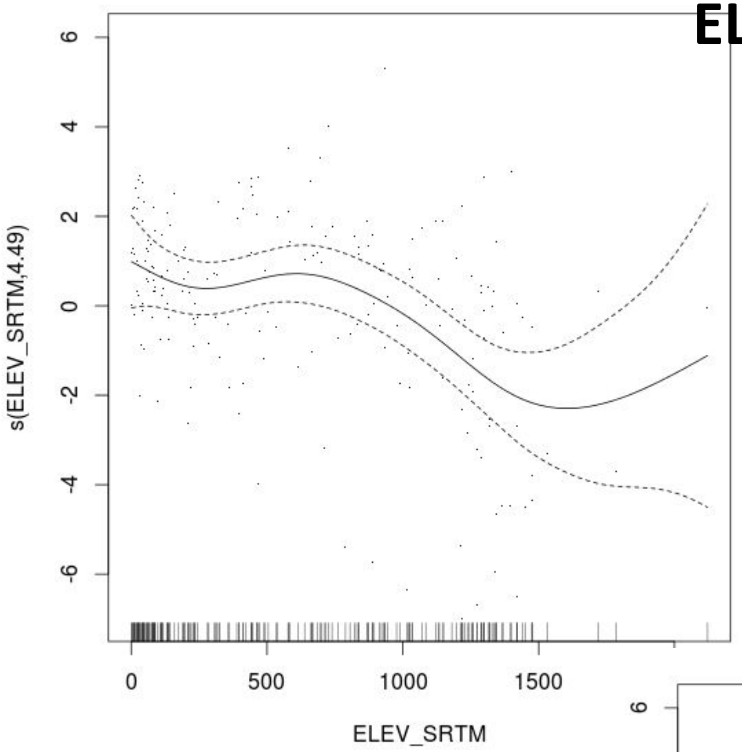


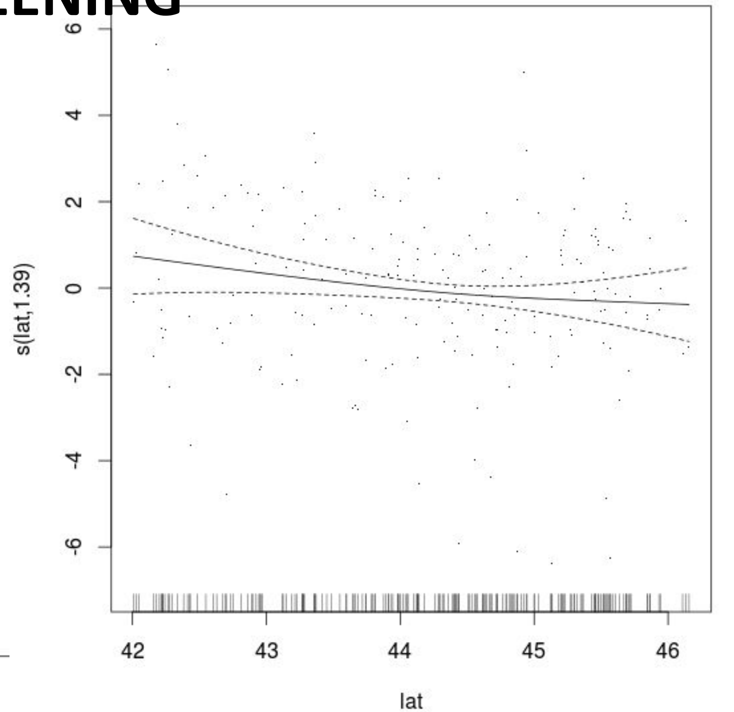


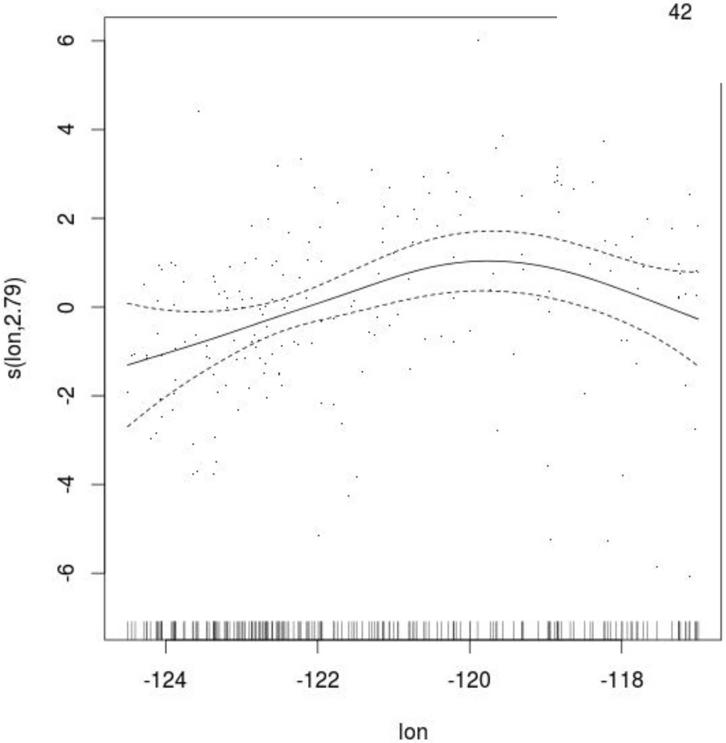
These are plots from gam objects. They show smooth curves for s(ELEV), s(lat) and s(lon).
The independent variable is LST_bias.

ELEVATION SCREENING

After screening…







These are plots from gam objects. They show smooth curves for s(ELEV), s(lat) and s(lon). The independent variable is LST_bias.

PART 4:

Residuals analysis: started the work but yet to be updated with new results

PART 5:

Residuals analysis summarize by season

(code written)

MAIN CONCLUSIONS FOLLOWING THE UPDATED ANALYSIS (ON NOV. 3, 2010

1. There is a slight improvement (decrease of 0.1-0.2 C in RMSE) when more stations are used for the monthly surface estimation.

2. When ELEV_SRTM and LST were screened, results changed slightly on average. Screening may need to be adjusted so that it does not lead to station loss.

3. Spatial and temporal patterns in LST images "make sense" but LST seem to have more seasonality with stronger bias in Summer in particular in basin areas with low vegetation.

4. When using simplified model for the modeling of Tmax climatology, RMSE values were higher but the spatial pattern more sensible than for CAI+Kriging because of the spatial detail.

5. CAI+Kriging and GAM+Kriging remain the "best" method based on the average RMSE over 365 dates. We note however that the gap between methods decreased when more stations were added in monthly time step. It appears the number of station is the most critical factor (need to show this, fits with the literature).