Boreal Avian Modelling Project

Biophysical **D**atabase - **D**escriptive **S**tatistics

Steve Cumming, Mélanie L. Le Blanc

June 22, 2010

Contents

1 Introduction **4**

2 Climate Variables **5**
2.1 Annual climatic indices . . . 6
2.2 Monthly climatic indices . . . 7
2.2.1 Monthly Means of Primary Variables . . . 7
2.2.2 Monthly Standard Deviations of Primary Variables . . . 23
2.2.3 Correlation between Monthly Standard Deviations of Primary Variables . . . 31
2.2.4 Comparison of Means and Standard Deviations for the Summer Period . . . 38
2.3 seasonality variables . . . 48
2.4 ecophysiological variables . . . 50

3 Remothly sensed variables **51**
3.1 Landcover . . . 51
3.2 Vegetation Continuous Fields . . . 52
3.3 Productivity . . . 54
3.4 Greenness Indices . . . 55
3.4.1 Monthly Leaf Area Index . . . 55
3.4.2 Normalized Difference Vegetation Index . . . 55
3.4.3 Comparison of Monthly Greeness Indices for the summer period . . . 55

The Biophysical Dataset: Descriptive Statistics

1 Introduction

The Boreal Avian Modelling Project and the collaborating Canadian BEACONs project have assembled an extensive library of spatial data, which we refer to as "the Biophysical Dataset." From BAM's perspective, the Biophysical Dataset is primarily intended to provide habitat covariates for developing statistical models of the avian data; they could of course be used in any other statistical modeling exercise.

Many geodatabases now exist containing vast amounts of spatial data. Many of BAM's data partners would have used specialized or local spatial data sets of their own in sample design or analysis. Because acquiring and working with any large dataset entails costs, it is necessary to be selective. BAM's main criteria for selecting spatial data sets were as follows:

1. selected variables must correspond to known or hypothesized processes effecting the distribution and abundance of forest birds in the boreal region—no fishing!;
2. each covariate must be available in a consistent spatial coverage that includes the entire BAM study region;
3. the data must be available at little or no charge;
4. all data must be freely usable in publications (subject to appropriate acknowledgments) and must be well documented.

The most readily available datasets that satisfy these criteria are spatially interpolated climate data and remote-sensed products. The current version of the database now contains 106 climatic covariates, and 25 remote-sensed covariates. The data are summarized at the level of the point count station, not the round or year. That is, we present a view of the spatial locations sampled by the BAM data base. The number of unique spatial locations is HOW MANY? Because the spatial resolutions of the covariates differ, there is a certain degree of non-independence. For example, for data at the 1km resolution, several stations may be located within the same pixel. Such spatial dependence is not accounted for here.

The purpose of this part of the AutoDocumentation report is to describe these covariates and to present a partial view of their correlation structure. Among 131 covariates, there exist 8515 unique pair-wise correlations amongst this set of covariates, and it is not possible to examine them all. We focus here on correlations within between pairs judged a priori to be potentially highly correlated. For example, we illustrate the relations between different monthly climate covariates within months (e.g. June Minimum and Maximum temperate), and between consecutive months (e.g. May and June precipitation).

We use correlograms to show the univariate distributions and pairwise correlations among groups of from two to five covariates. The correllelogram is a commonly-used tool for exploratory visual display and for depicting the patterns of relations among variables in matrices directly, particularly when the number of variables is moderately large (Firendly, 2002). A correlogram is a matrix of graph panels, in this document structured as follows. The diagonal elements show the univariate distribution of each covariate in the group, as a histogram of optimal bin width. The lower-diagonal elements are pairwise scatterplots, each using a random subsample of 1000 locations. The subsampling is necessary simply to limit the size the images. Associated with each scatter plot is an estimate of the Spearman rank correlation coefficient (given the sample sizes, we consider significance levels to be superfluous). The upper-diagonal elements show, for each pair or covariates, a loess curve fit to the bivariate data, and an ellipse which includes 95% of the points. The Spearman rank coefficients, loess curves and 95% ellipses use the full data set, rather than a sample. The corelograms were produced using a customized version of function *Correlogram* from R package *corrgram*.

The statistic software program used for this document was R version 2.7.2 (2008-08-25)and the platform version was i386-pc-mingw32 and LaTeX 2_ε.

2 Climate Variables

All the climatic covariates described in this section were produced by the Canadian Forest Service, Natural Resources Canada (NRCAN), based on data from weather stations maintained by Environment Canada. These weather stations record hourly temperatures and total precipitation measured from solar noon to solar noon. Hourly temperatures are rolled up to a daily minimum and daily maximum temperature, and then to monthly means of the daily minima and maxima. Annual mean temperatures and total precipitation are then calculated from the monthly values. Additional indicators are calculated from the daily data. For example, the annual minimum temperature at a station is calculated as the minimum of daily minimum temperatures over the year. Seasonality indicators can also be derived from the daily or monthly data. For example, if conditions for the start of the growing season are defined as the date when the mean daily temperature excess some threshold, the date of this event in any given year may be determined. Climate normals are conventionally 30 year means of annual values. Thus, the normal for mean minimum May temperature is the mean over 30 years of May monthly means of daily minimum temperatures. Normals for seasonality and other indices are similarly calculated as 30yr means of annual values. Unless otherwise noted, we selected the period 1971-2000 as the 30yr normal period, as this period covers most of the data collected by BAM.

The station level data are available as recalculated normals and as annual values. These are spatially interpolated to non-sampled locations using elevation-correct cubic-splines via the ANUCLIM system (McKenney et al., 2006a). A number of standard interpolated products are available; when a regular grid is required, we use the standard 300 arc-second grid provided by NRCAN. Most of the data used in this Part of the report were custom interpolated to the exact locations of the point-count stations. BAM gratefully acknowledges the courtesy of Dan McKenney and Pia Papadopol (Canadian Forest Service, Sault Ste. Marie)in providing this data.

2.1 Annual climatic indices

As annual data, we used 1971-2000 normals of:

- Mean Annual Mean Temperature (MANNTEMP)
- Minimum Temperature (MINCOLD)
- Mean Annual Precipitation (ANNPRC)

For the latter index, see (McKenney et al., 2007).

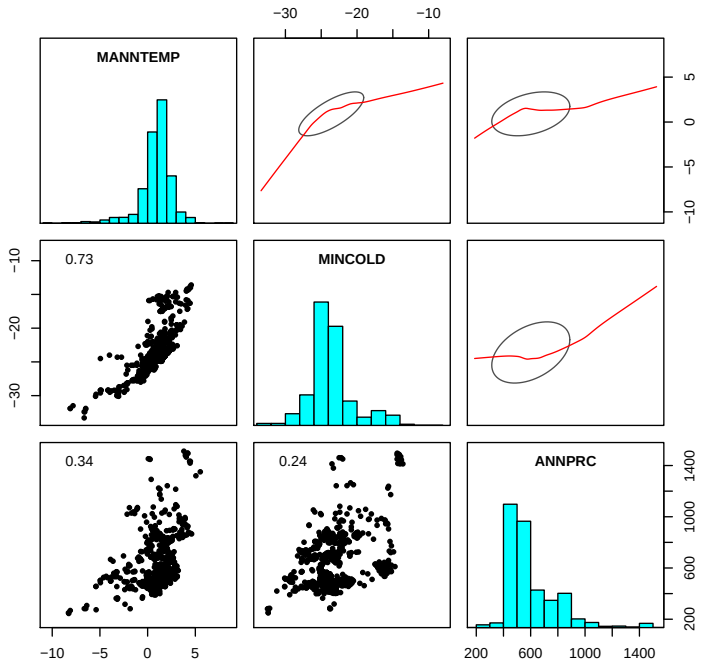

Figure 1: Correlogram of Annual Data

2.2 Monthly climatic indices

Monthly climate indices were based of four primary monthly values:

1. Mean Monthly Minimum Temperate Tmin
2. Mean Monthly Maximum Temperate Tmax
3. Mean Monthly Mean Temperate, estimated as 0.6 Tmax + 0.4 Tmin.
4. Mean Monthly Total Precipitation

We calculated 30-year normals for each of these four primary variables for each of the 12 calendar months, for 48 variables in total. We also calculated standard deviations of monthly values over the normal period for each of the four main variables for each of 12 calendar months, yielding an additional 48 covariates. These calculations are done for the point-level interpolated data, and for the trans-boreal coverage of the 300 arc-second grid. In this section, we summarise the station-level values only. There are thus two main groups of monthly climate variables: monthly means and monthly standard deviations.

For each group of monthly covariates, we report the pairwise covariances of each primary variable between months within seasons. For this purpose, we define 3 overlapping seasonal periods of five months each:

- Summer: April, May, June, July, and August
- Fall: August, September, October, November, and December;
- Winter: December, January, February, March and April

We also report the pairwise covariances among each group of four primary means and standard deviations within months. Finally, we report the bivariate correlation between monthly means and standard deviations of each of the four primary variables within the five summer months—the other monthly comparisons would require an additional 28 figures, and are neglected here for brevity. These results are presented as series of corellograms, with minimal commentary.

2.2.1 Monthly Means of Primary Variables

Correlation of monthly means of primary variables within seasons

Mean *Monthly* Temperature (TeM*month*): Average of Maximum and Minimum monthly temperature.

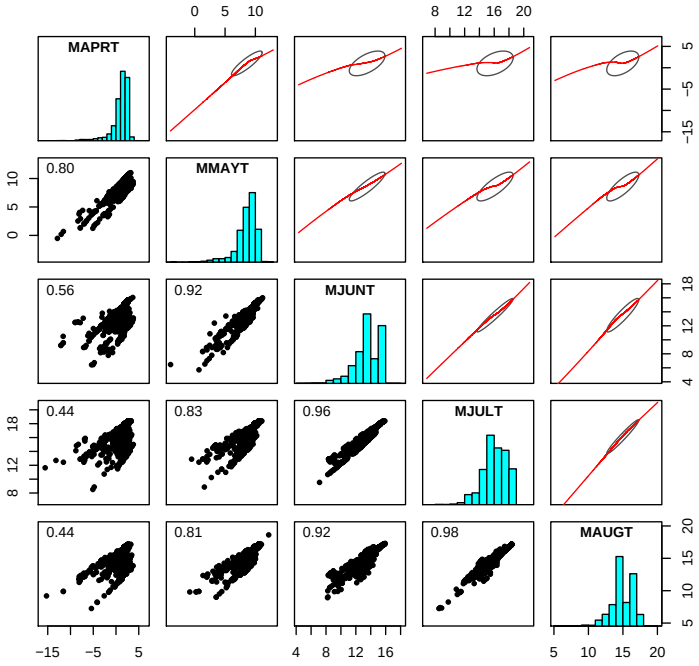

Figure 2: Mean Temperature for the summer period

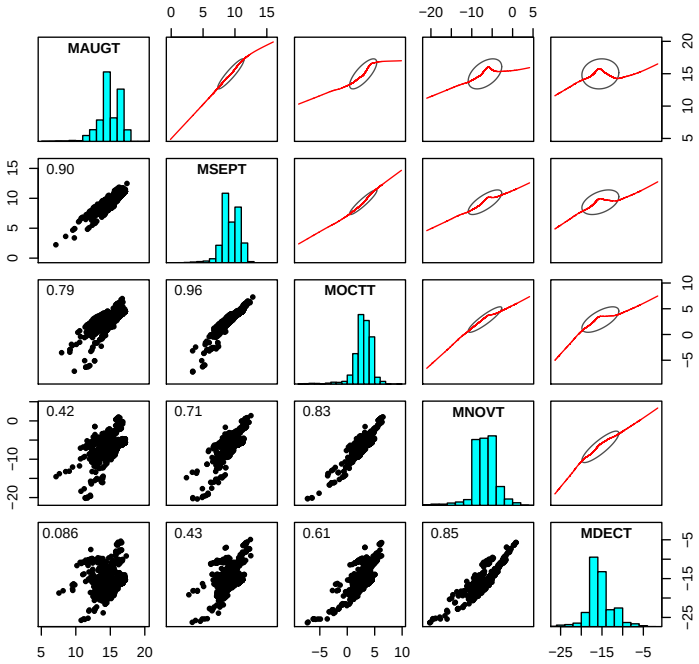

Figure 3: Mean Temperature for the fall period

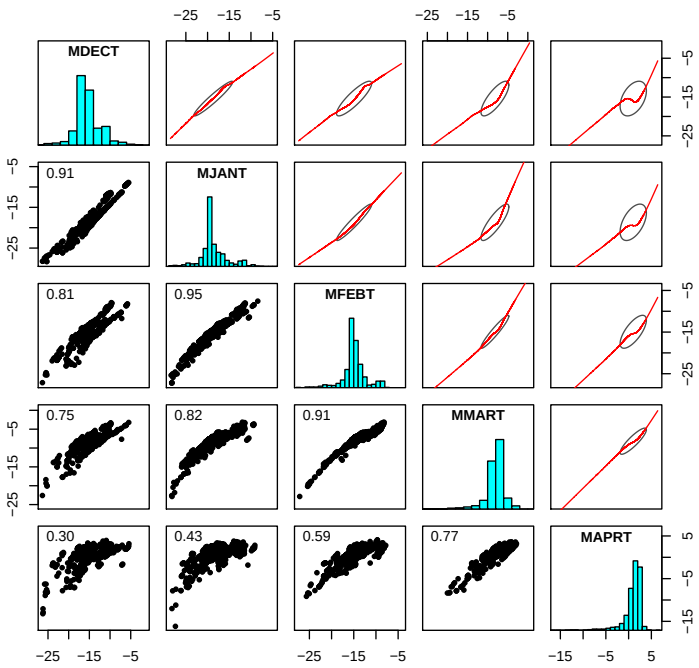

Figure 4: Mean Temperature for the winter period

Maximum *Monthly* Temperature (TeMx*month*): Maximum Temperature all months and by month.

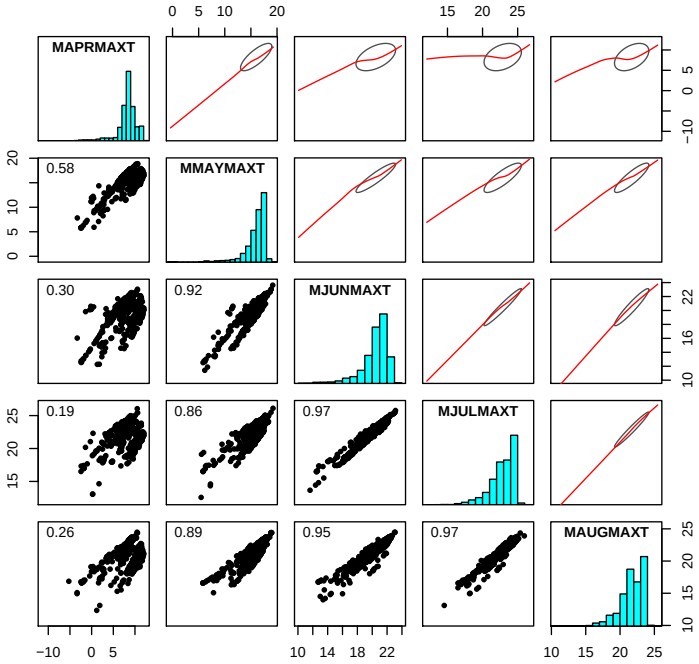

Figure 5: Maximum Temperature for the summer period

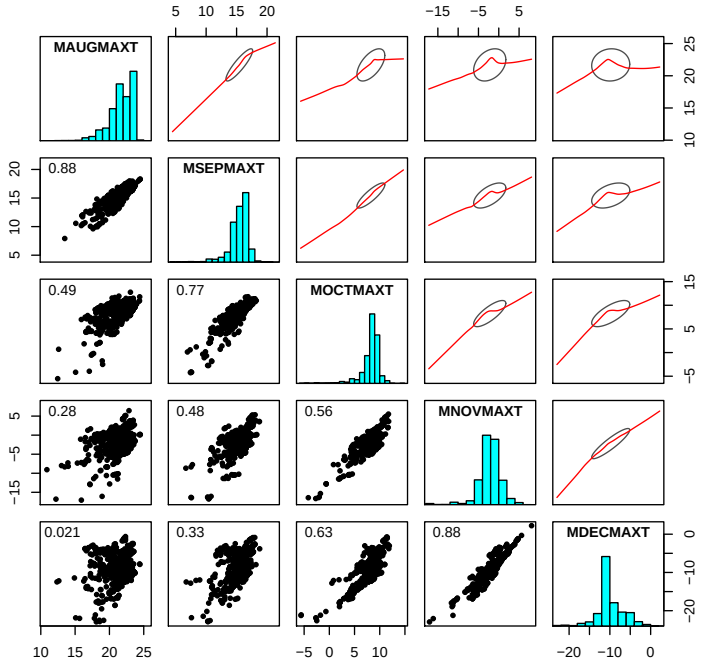

Figure 6: Maximum Temperature for the fall period

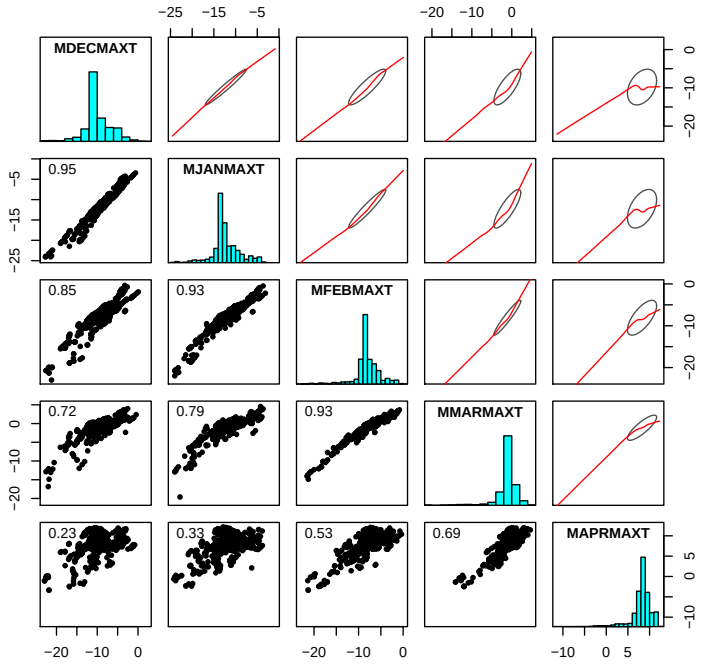

Figure 7: Maximum Temperature for the winter period

Minimum *Monthly* Temperature (TeMx*month*): Minimum Temperature all months and by month.

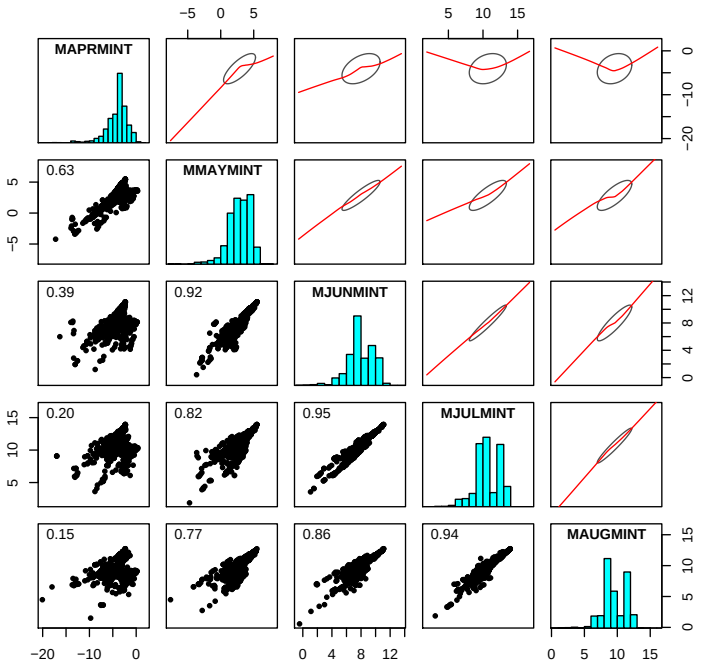

Figure 8: Mimimum temperature for the summer period

Mean *Monthly* Precipitation (PrM*month*): Mean precipitation all months and by month.

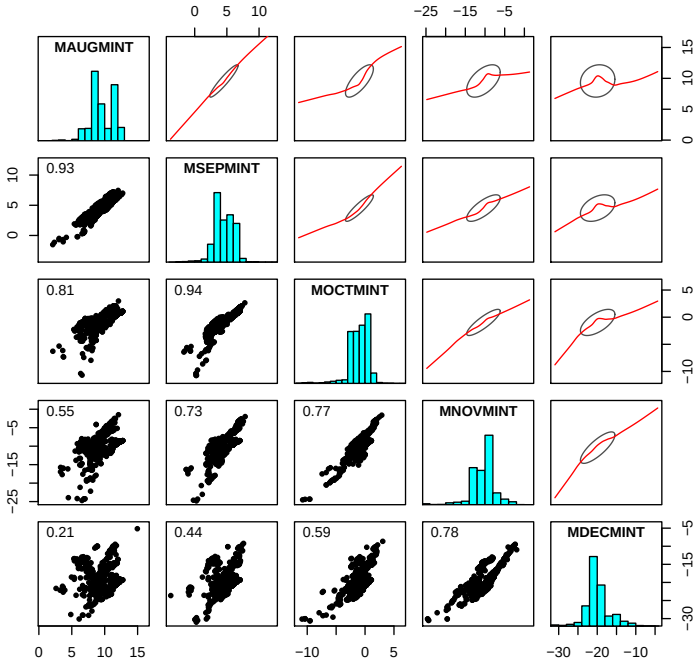

Figure 9: Mimimum temperature for the fall period

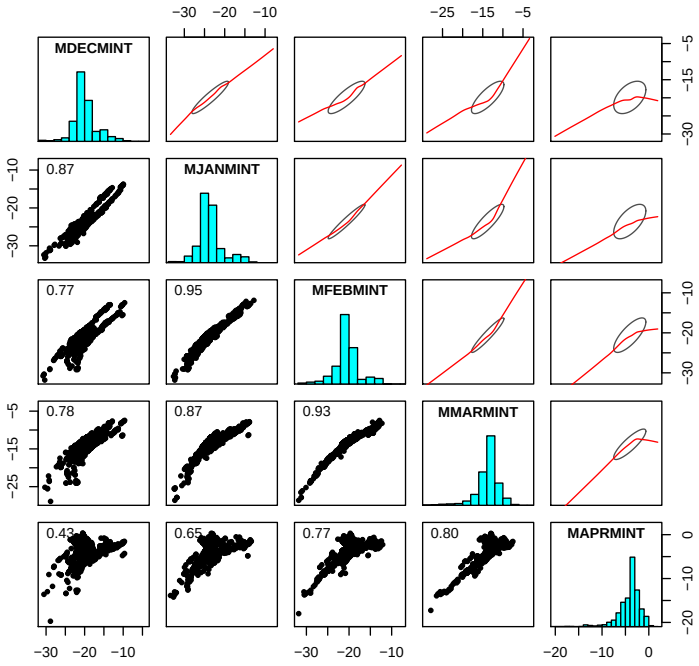

Figure 10: Mimimum temperature for the winter period

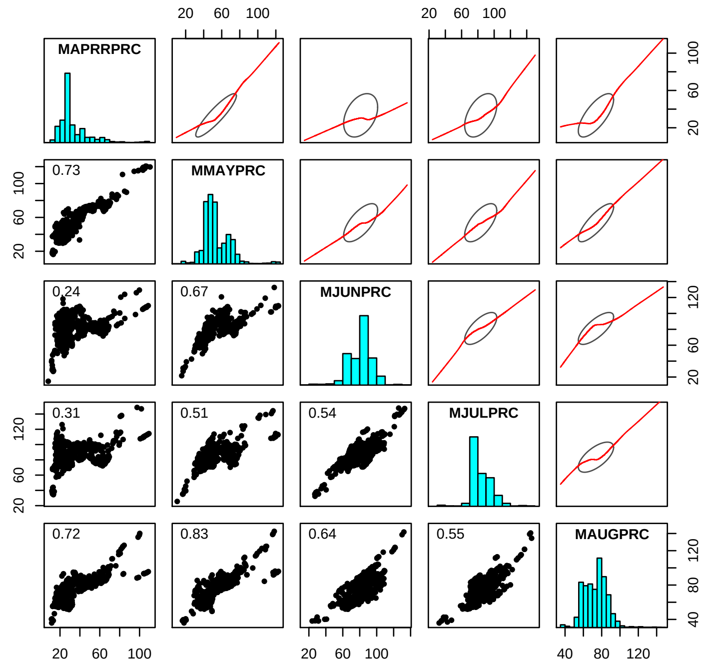

Figure 11: Mean Precipitation for the summer period

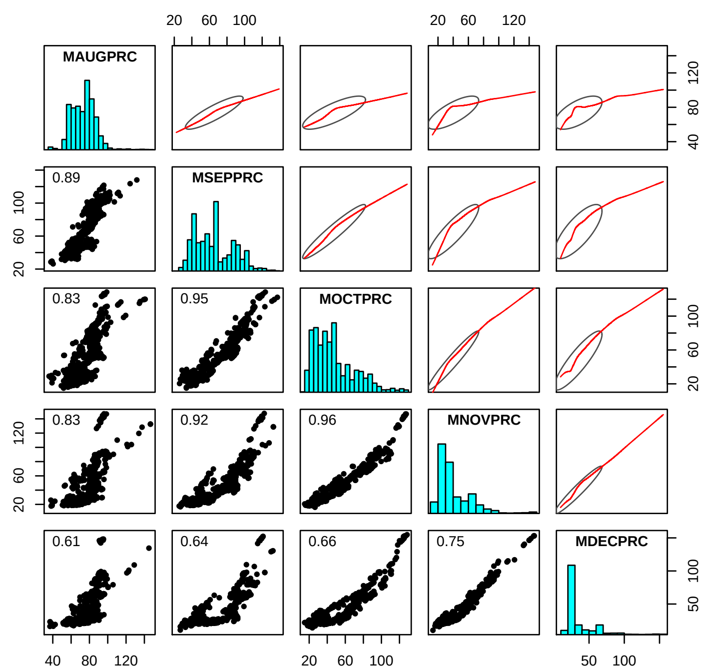

Figure 12: Mean Precipitation for the fall period

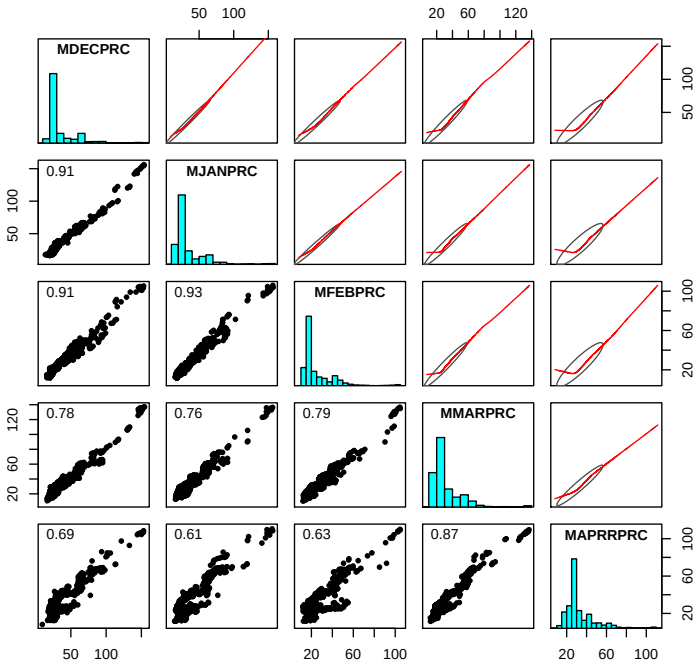

Figure 13: Mean Precipitation for the winter period

Correlation between monthly means of primary variables

Comparison of Mean *Monthly* Temperature, Maximum *Monthly* Temperature, Minimum *Monthly* Temperature and Mean *Monthly* Precipitation.

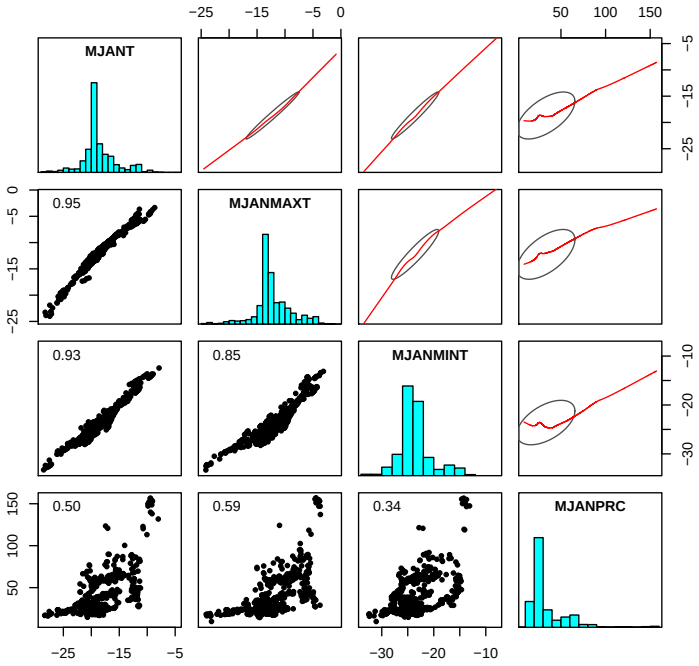

Figure 14: January

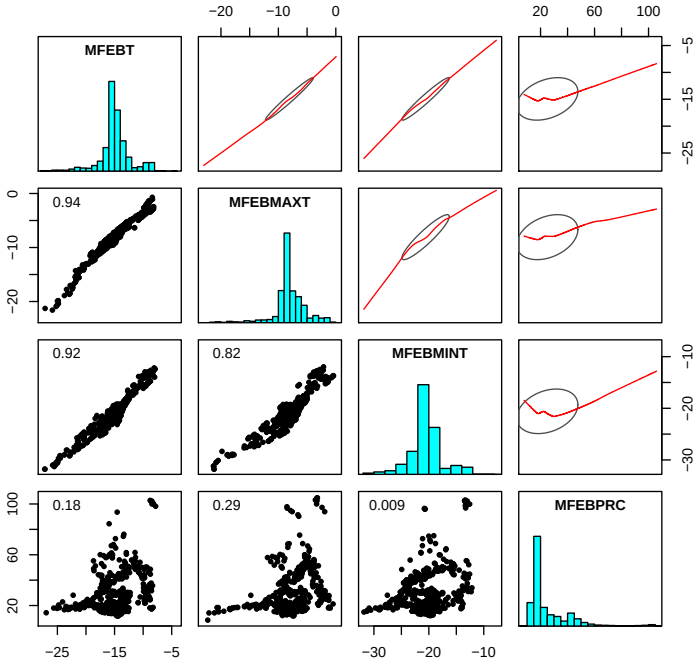

Figure 15: Febuary

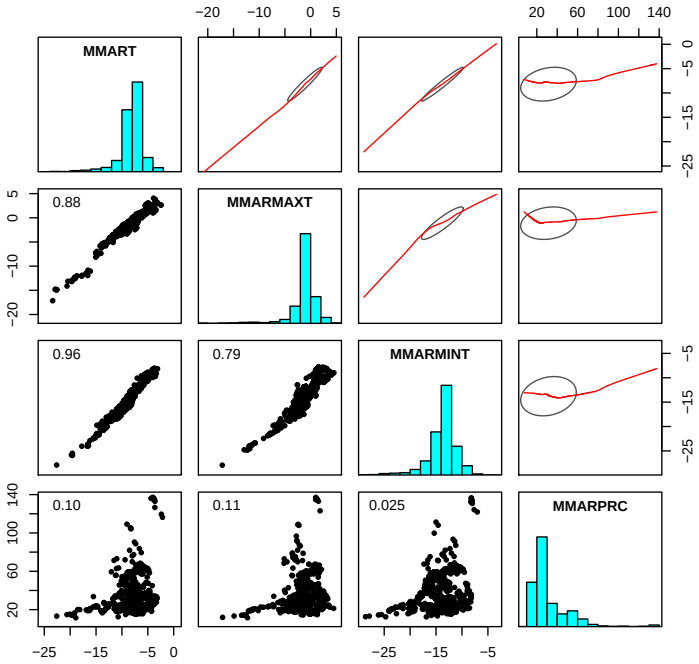

Figure 16: March

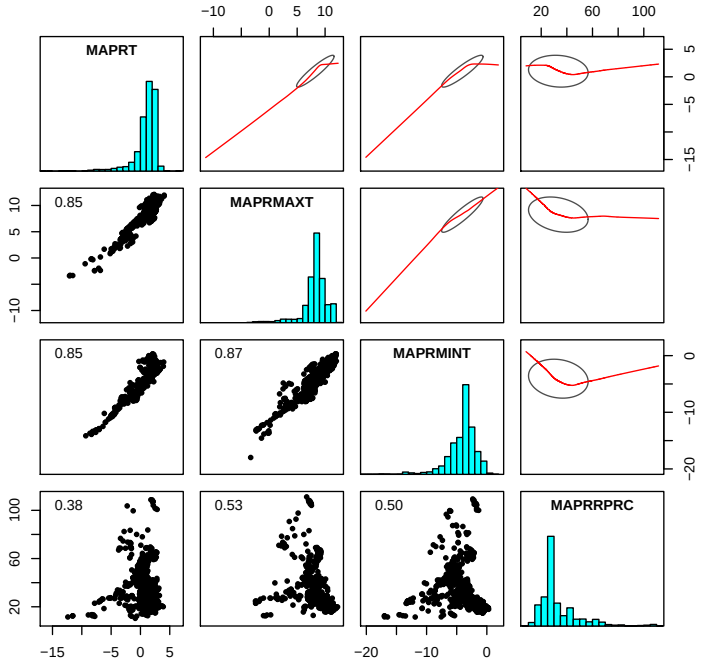

Figure 17: April

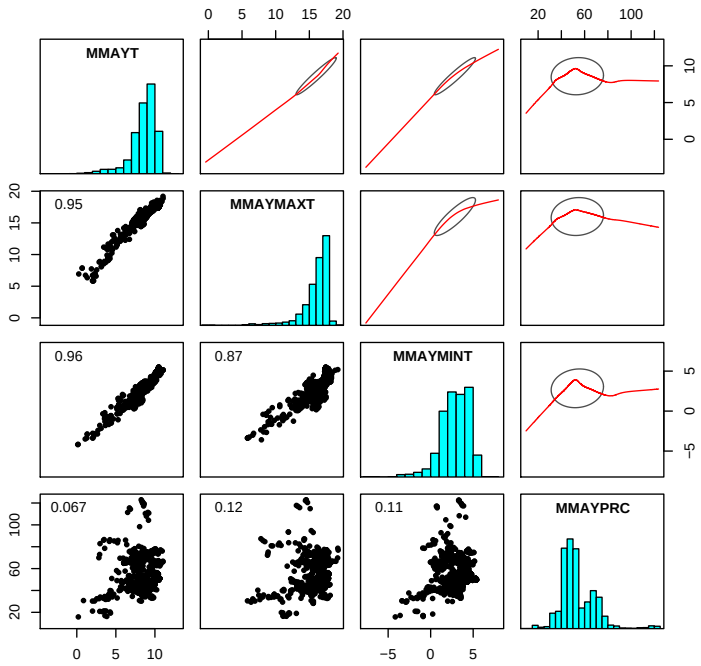

Figure 18: May

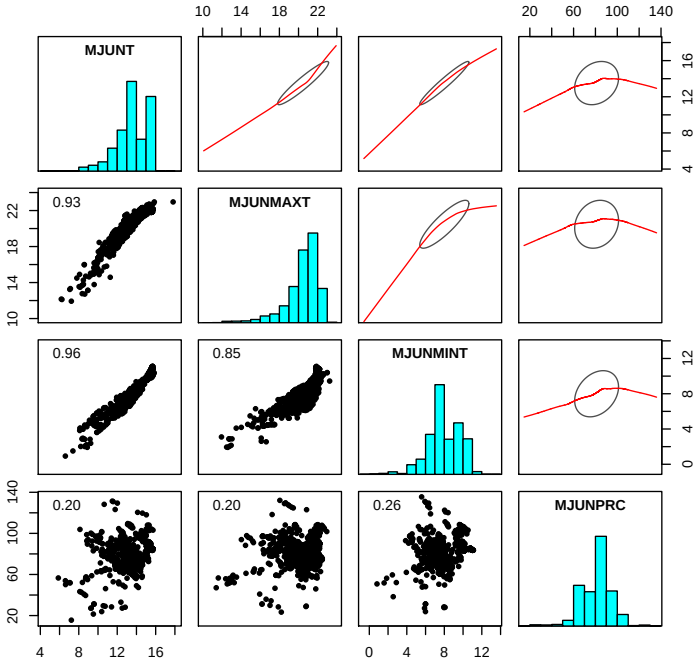

Figure 19: June

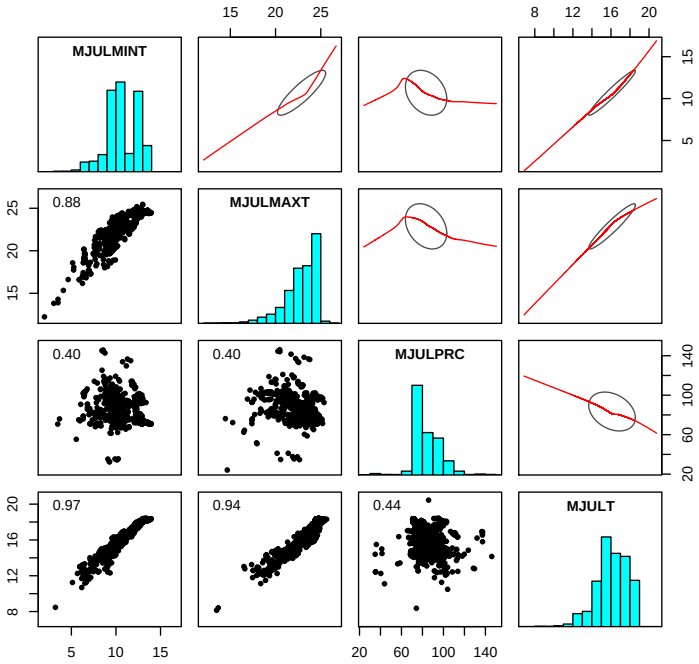

Figure 20: July

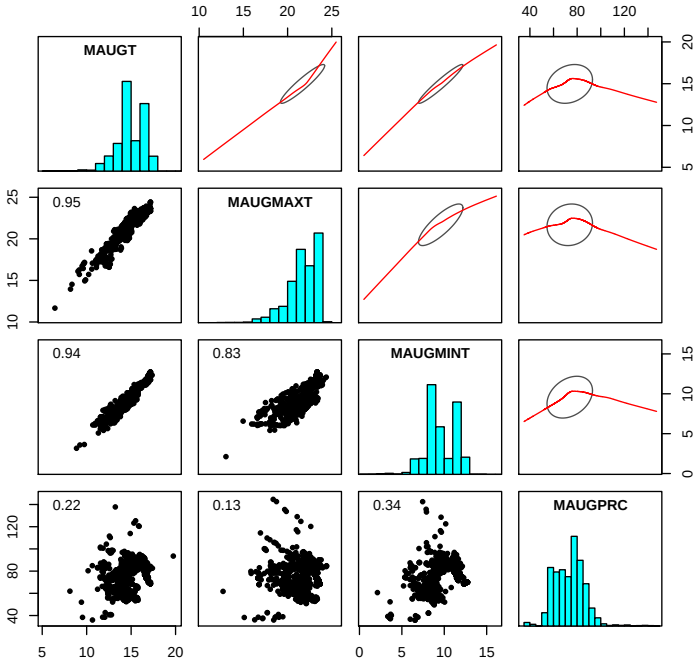

Figure 21: August

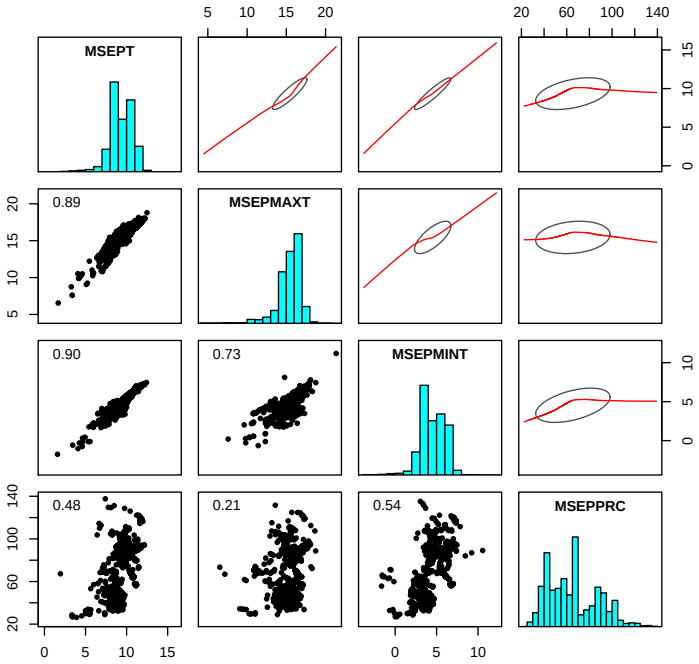

Figure 22: September

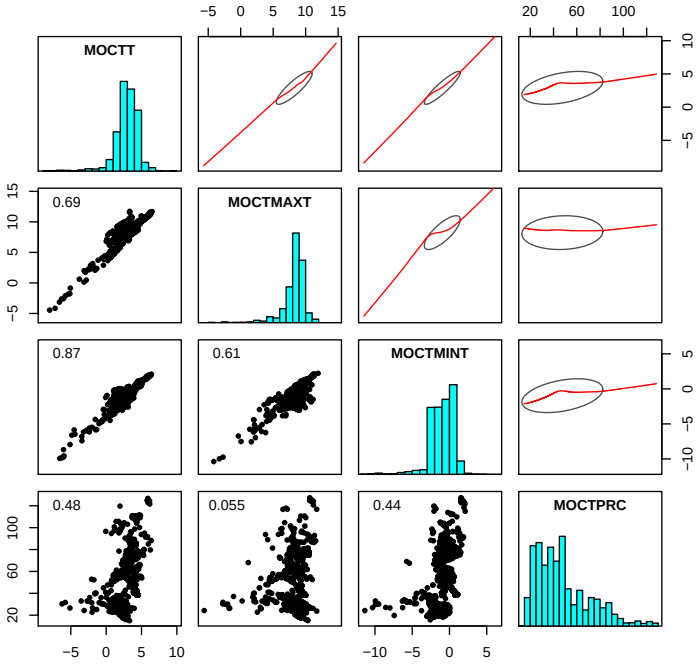

Figure 23: October

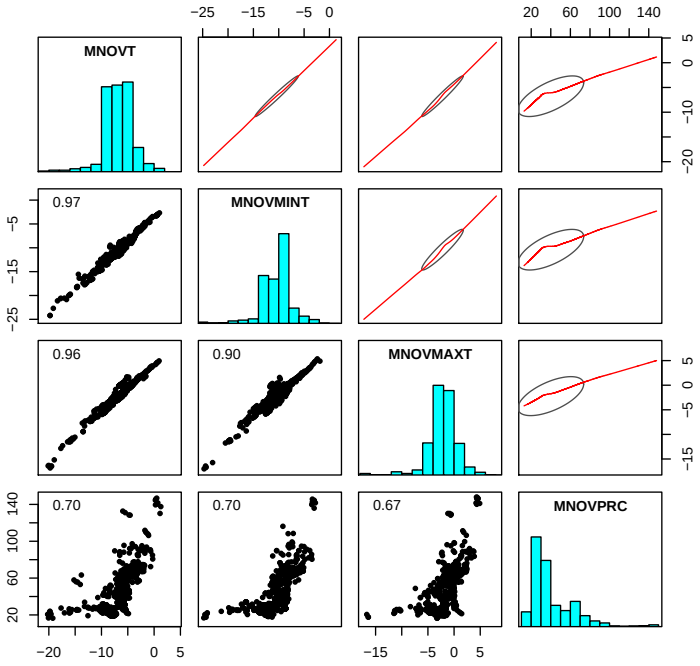

Figure 24: November

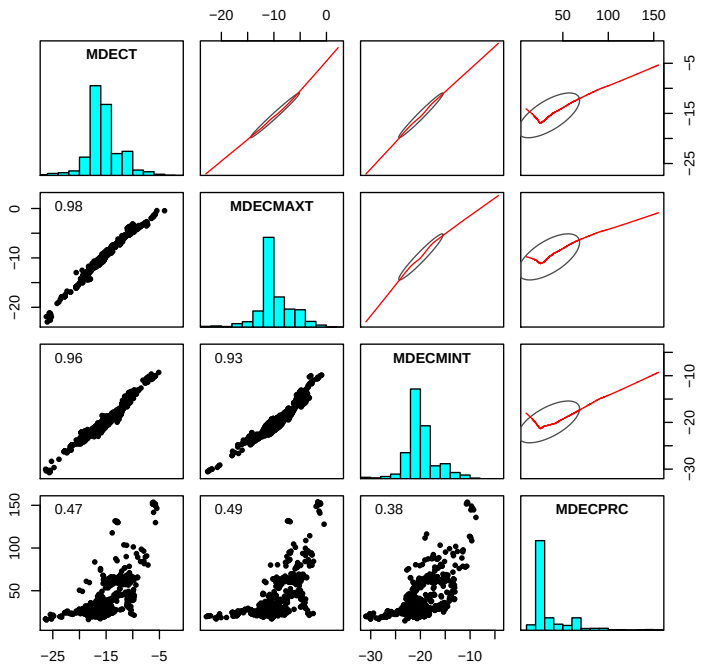

Figure 25: December

2.2.2 Monthly Standard Deviations of Primary Variables

Correlation of monthly standard deviations of primary variables within seasons

Mean *Monthly* Temperature Standard Deviation (TeMxSDmonth): Mean Temperature for Standard Error Maps.

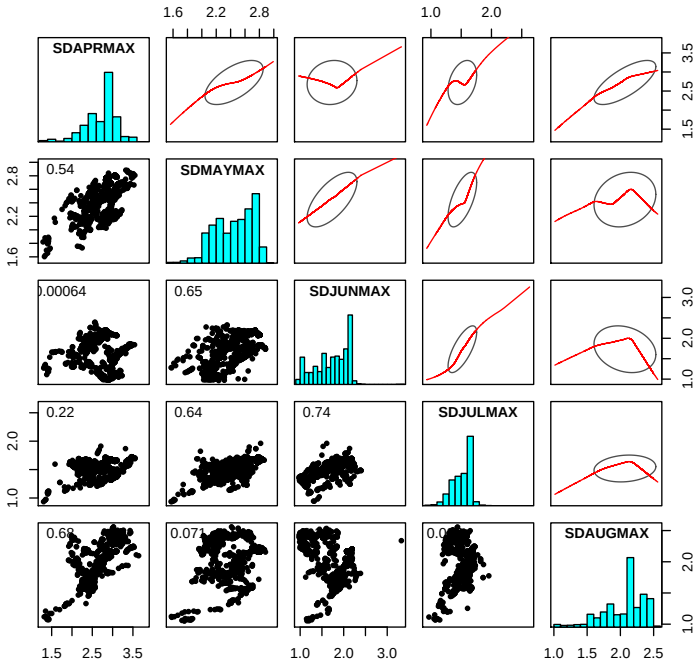

Figure 26: SD of Mean Temperature for the summer period

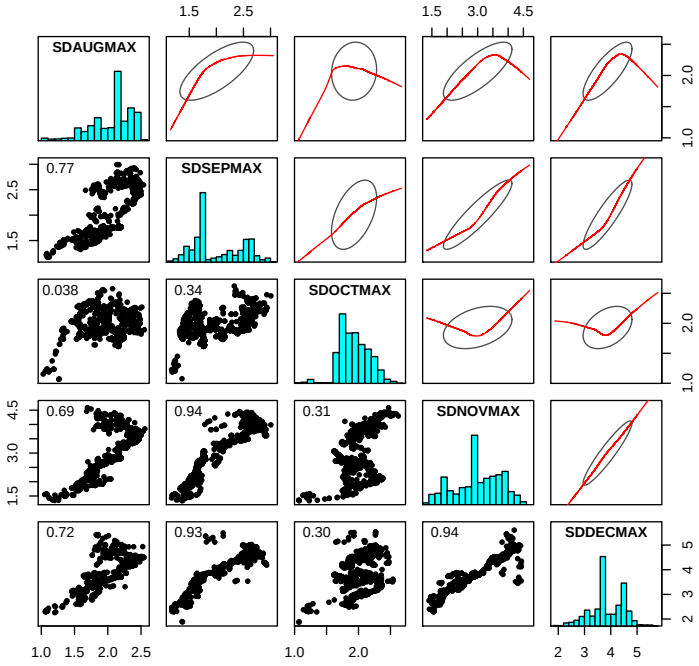

Figure 27: SD of Mean Temperature for the fall period

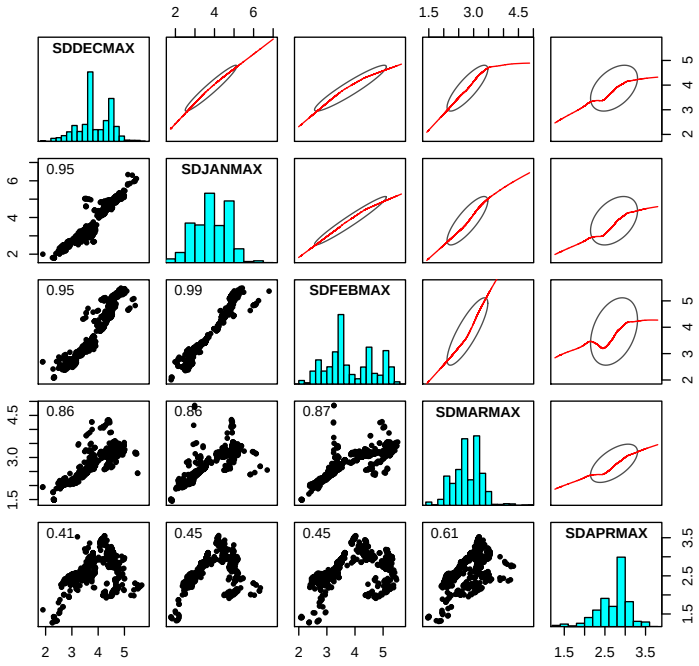

Figure 28: SD of Mean Temperature for the winter period

Maximum *Monthly* Temperature Standard Deviation (TeMxSDmonth): Maximum Temperature for standard Error Maps.

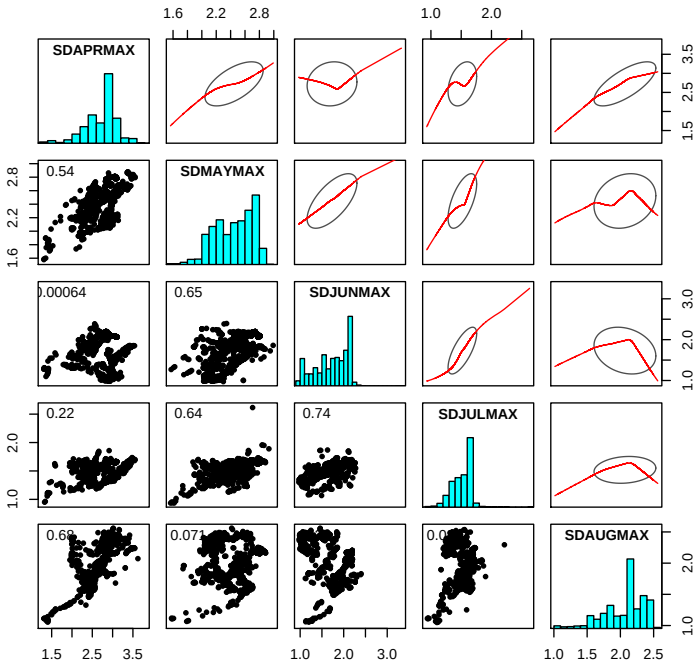

Figure 29: SD of Maximum Temperature for the summer period

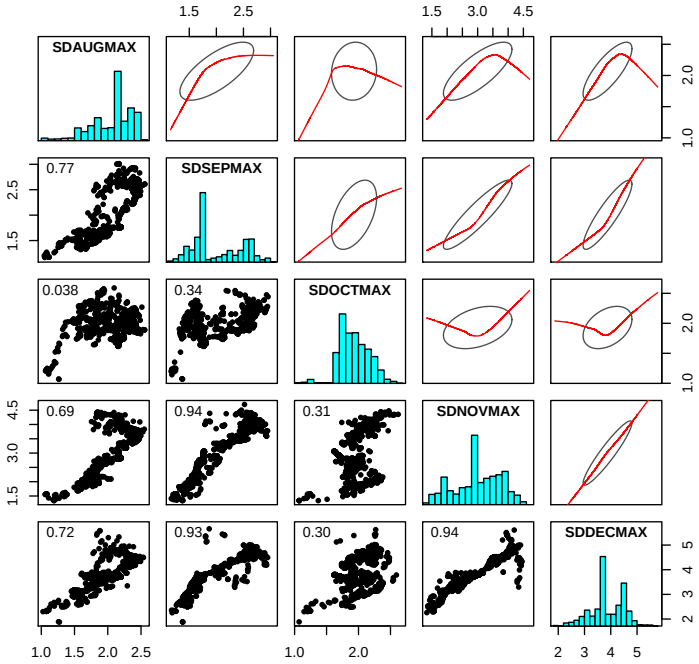

Figure 30: SD of Maximum Temperature for the fall period

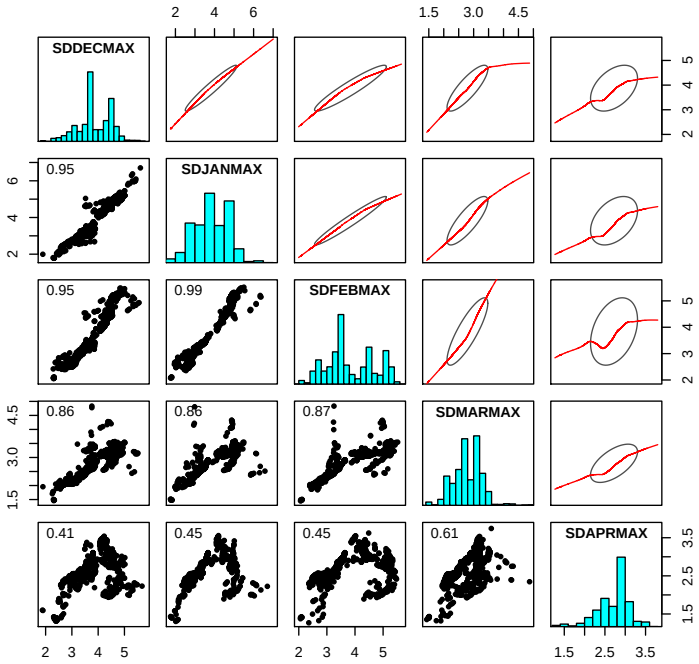

Figure 31: SD of Maximum Temperature for the winter period

Minimum *Monthly* Temperature Standard Deviation (TeMxSDmonth): Minimum Temperature for Standard Error Maps.

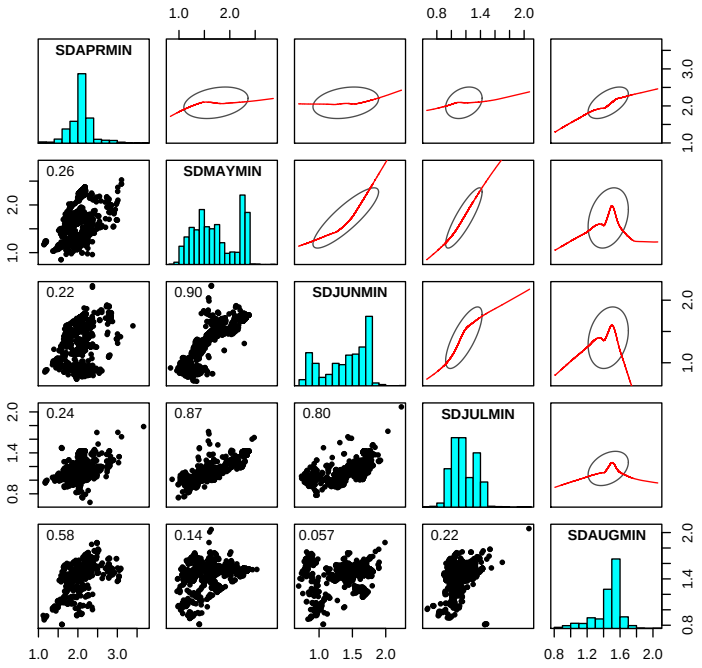

Figure 32: SD of Mimimum temperature for the summer period

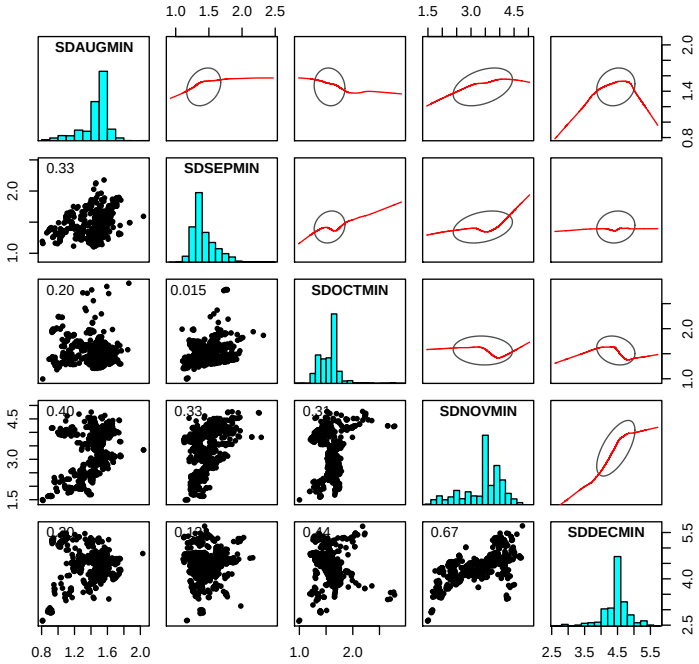

Figure 33: SD of Mimimum temperature for the fall period

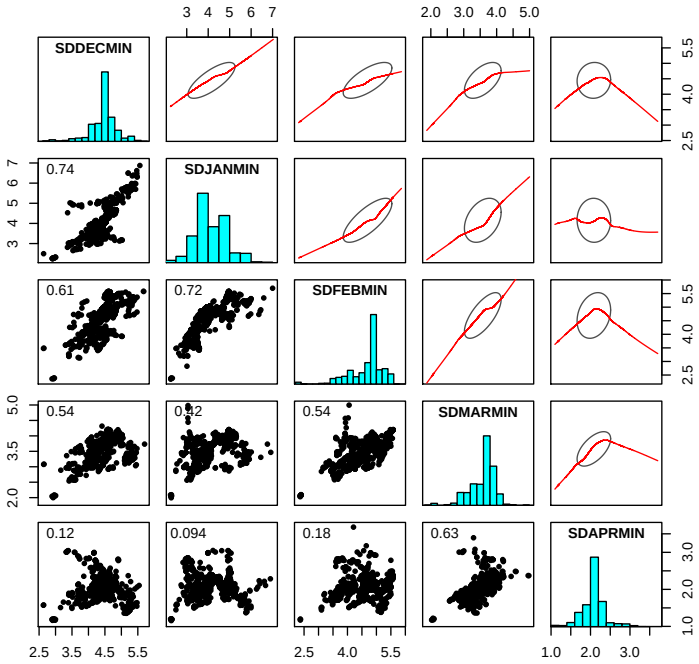

Figure 34: SD of Mimimum temperature for the winter period

Standard deviations *Monthly* Precipitation Standard Deviation (PrMSDmonth): precipitation for Standard Error Maps.

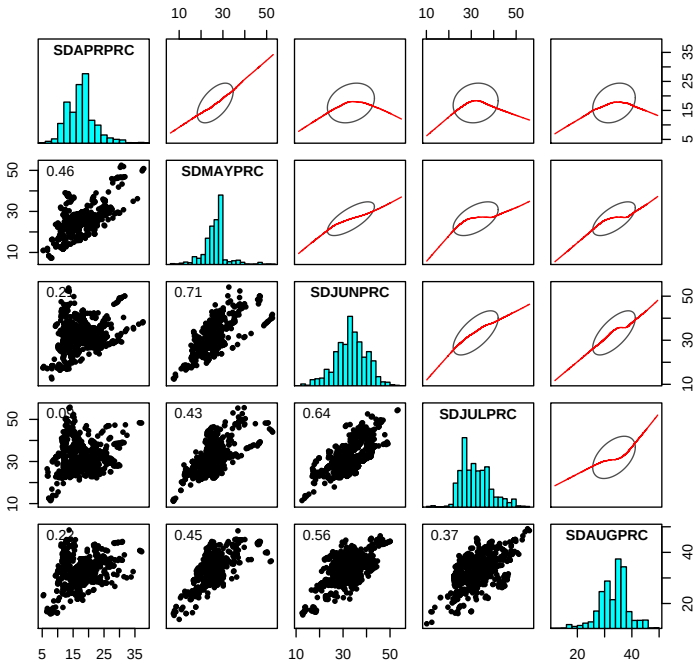

Figure 35: SD of Mean Precipitation for the summer period

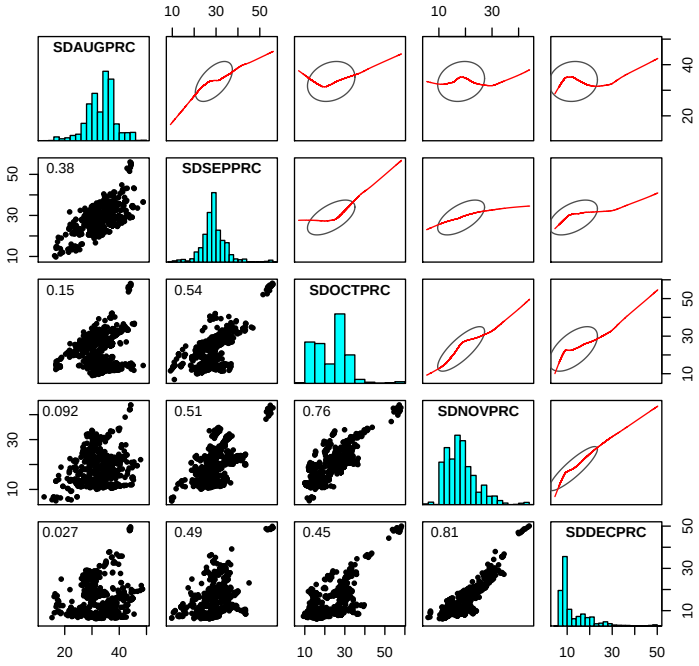

Figure 36: SD of Mean Precipitation for the fall period

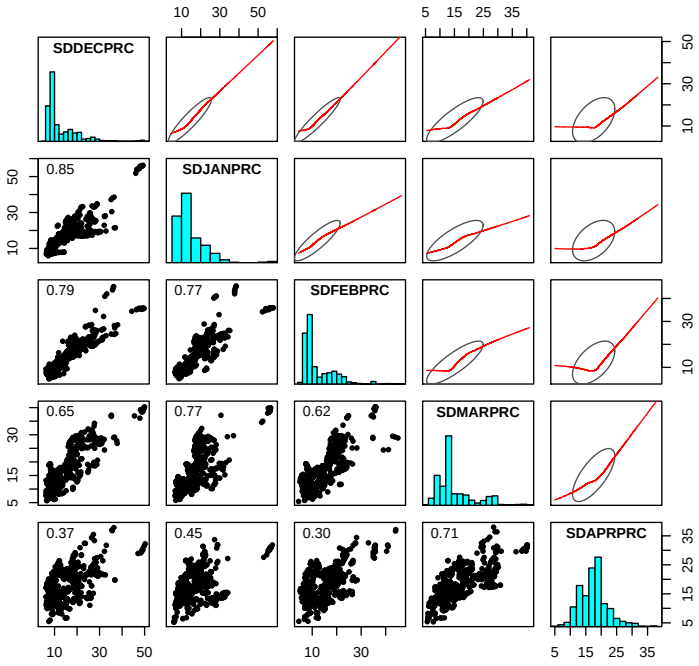

Figure 37: SD of Mean Precipitation for the winter period

2.2.3 Correlation between Monthly Standard Deviations of Primary Variables

Comparison of Standard deviations of Mean Monthly Temperature, Maximum Monthly Temperature, Minimum Monthly Temperature and Mean Monthly Precipitation.

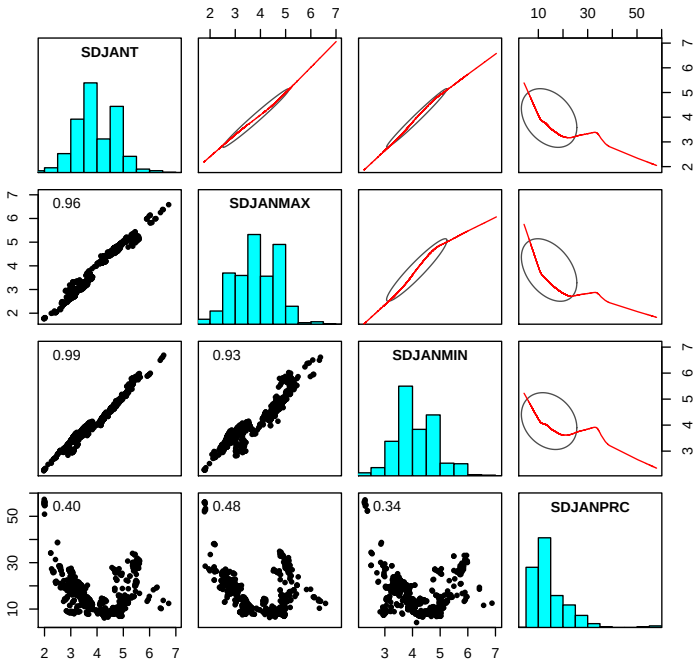

Figure 38: January SD

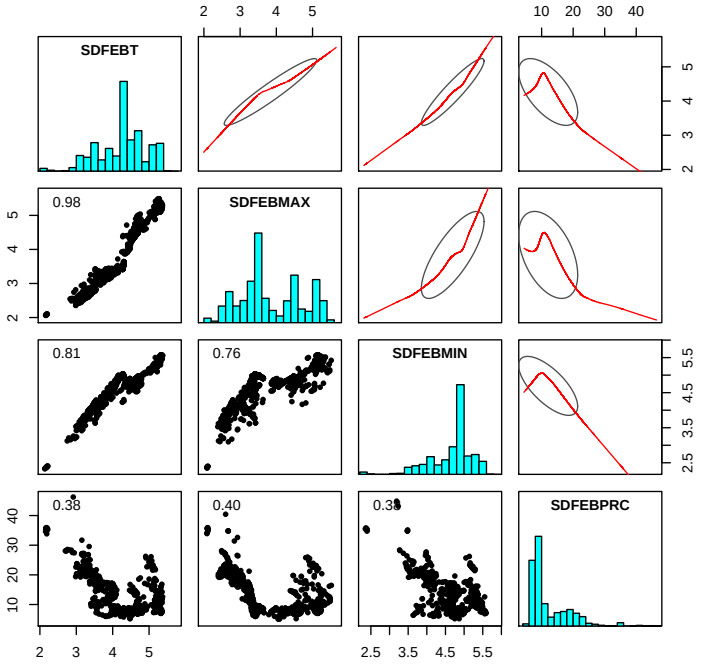

Figure 39: Febuary SD

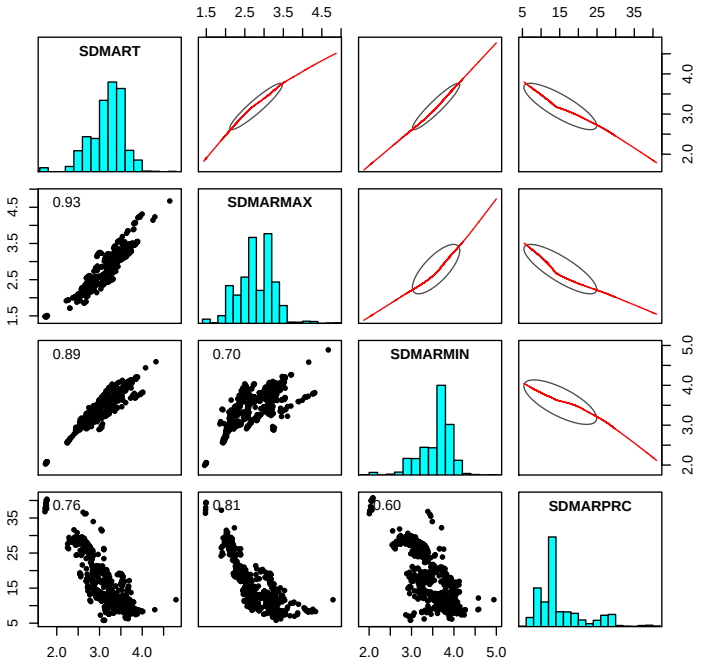

Figure 40: March SD

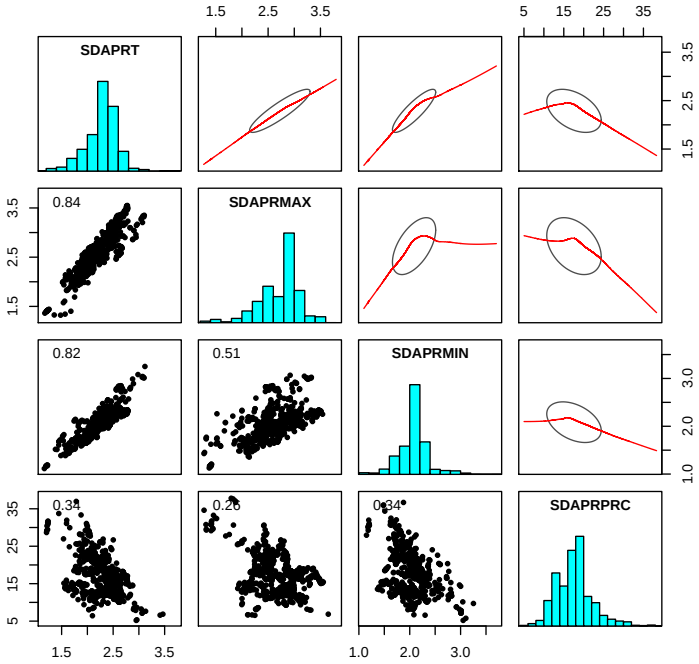

Figure 41: April SD

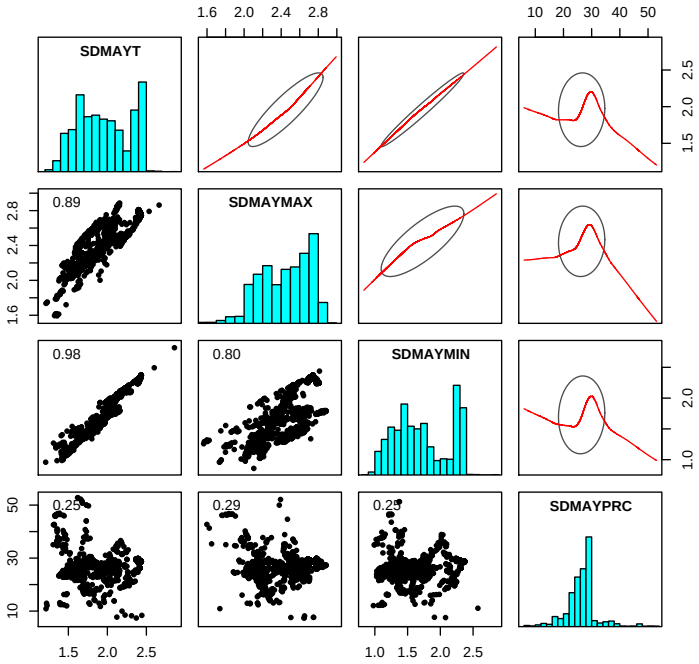

Figure 42: May SD

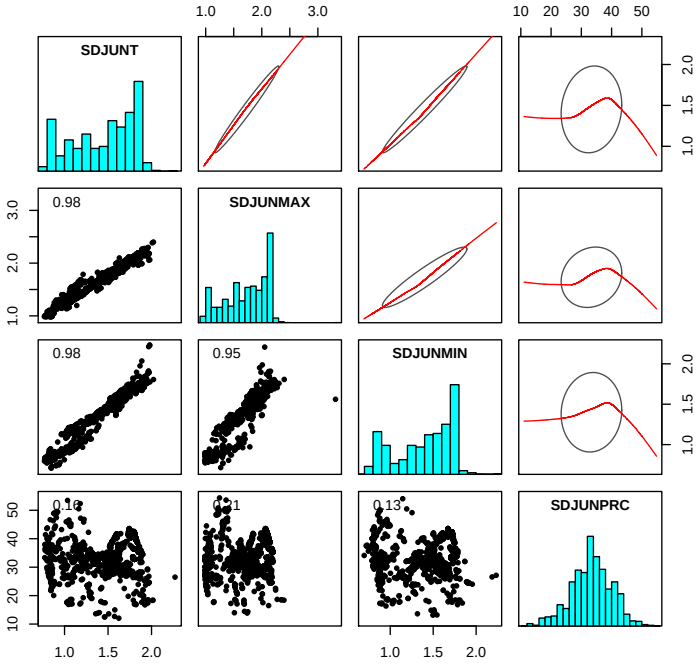

Figure 43: June SD

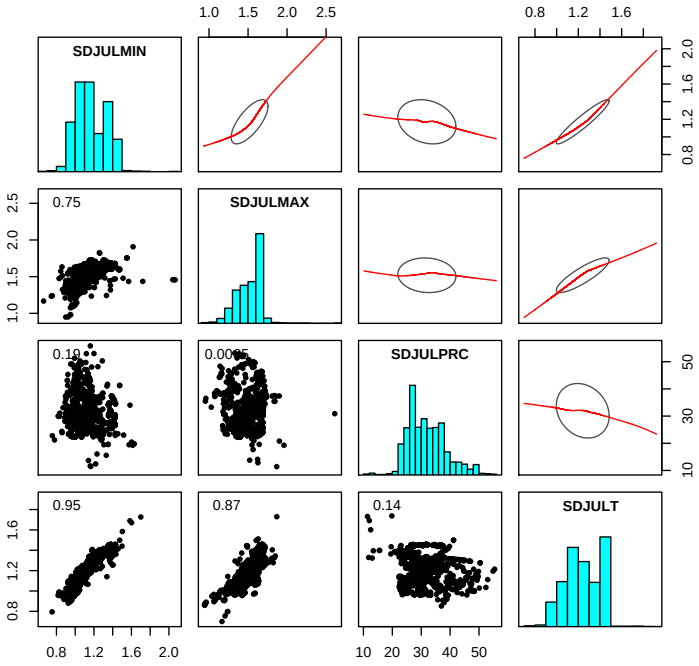

Figure 44: July SD

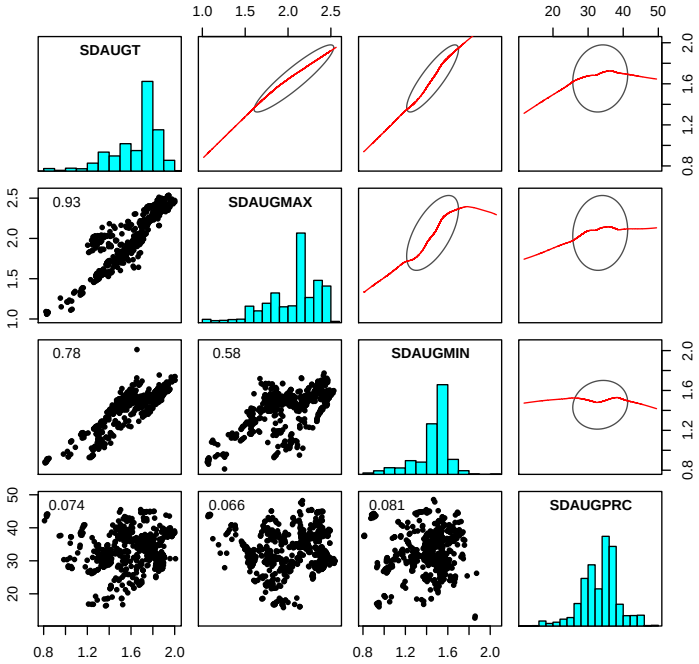

Figure 45: August SD

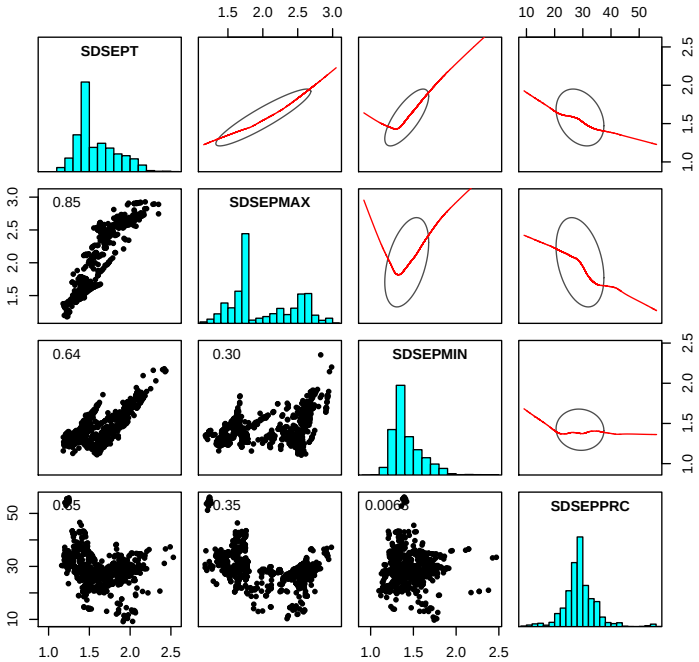

Figure 46: September SD

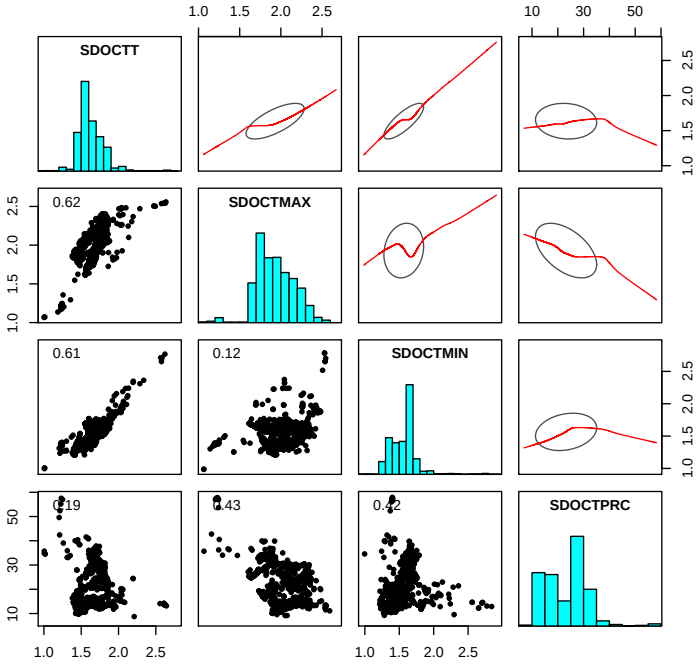

Figure 47: October SD

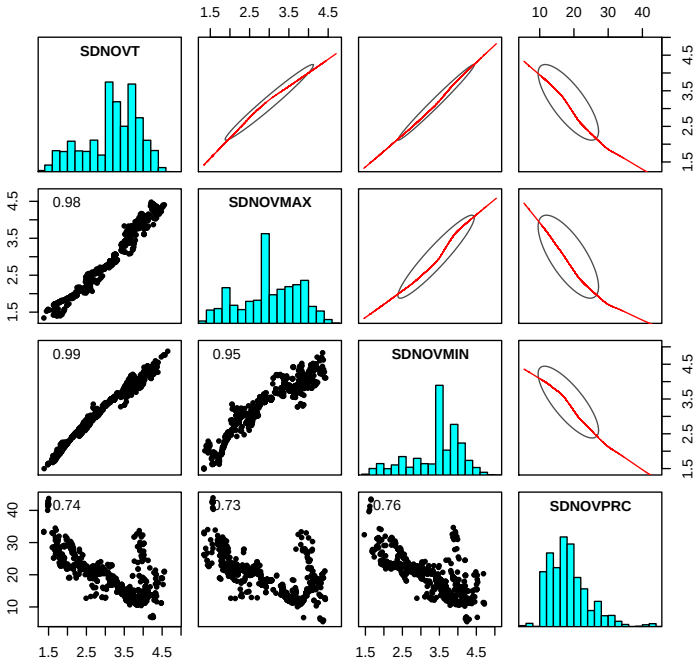

Figure 48: November SD

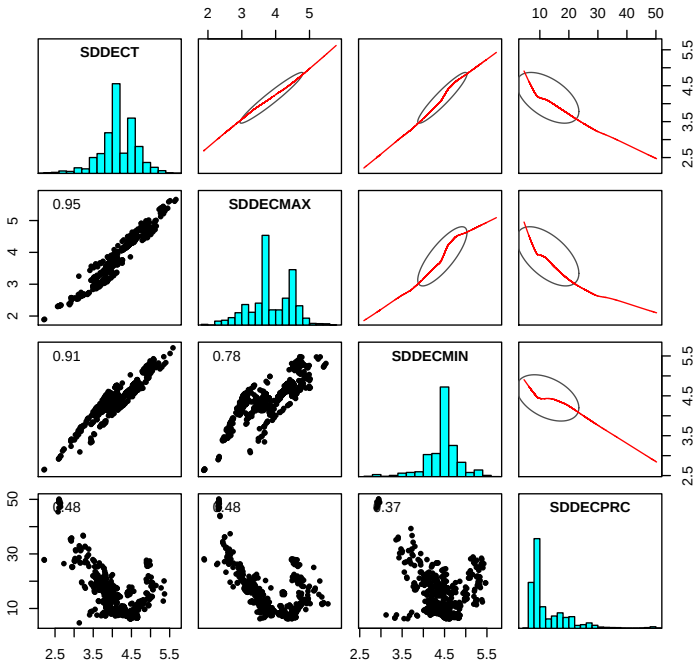

Figure 49: December SD

2.2.4 Comparison of Means and Standard Deviations for the Summer Period

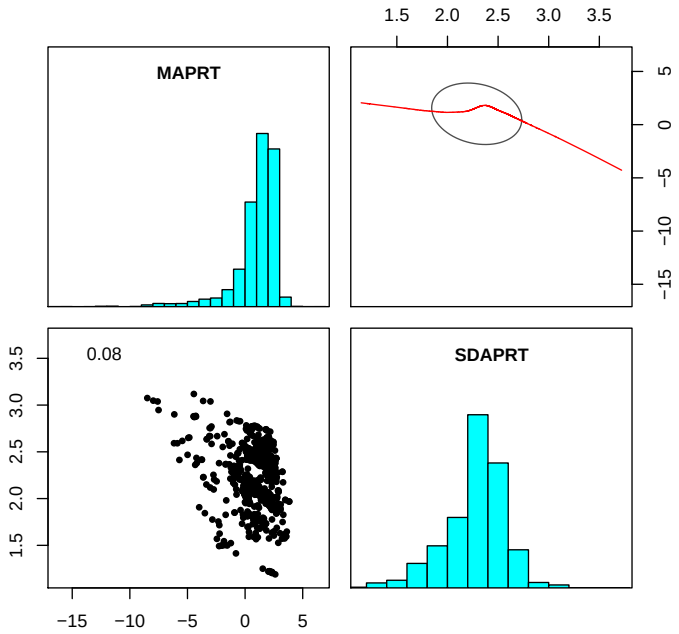

Figure 50: April Mean Temperature

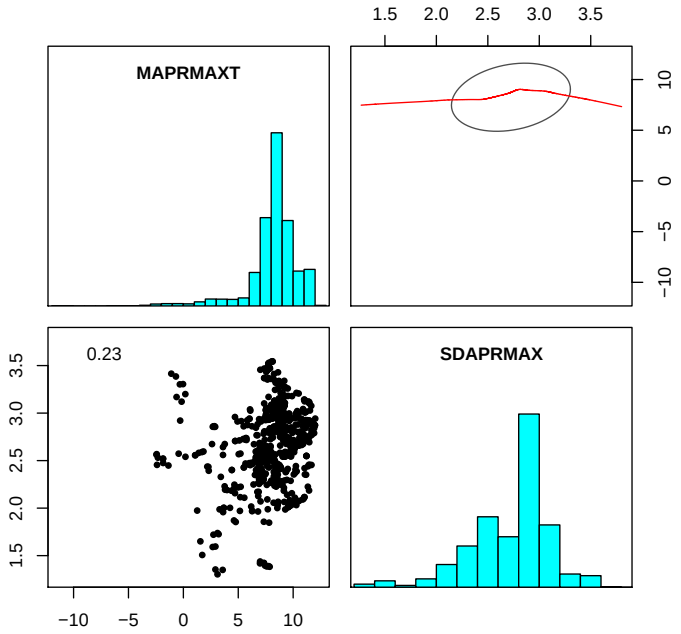

Figure 51: April Maximum Temperature

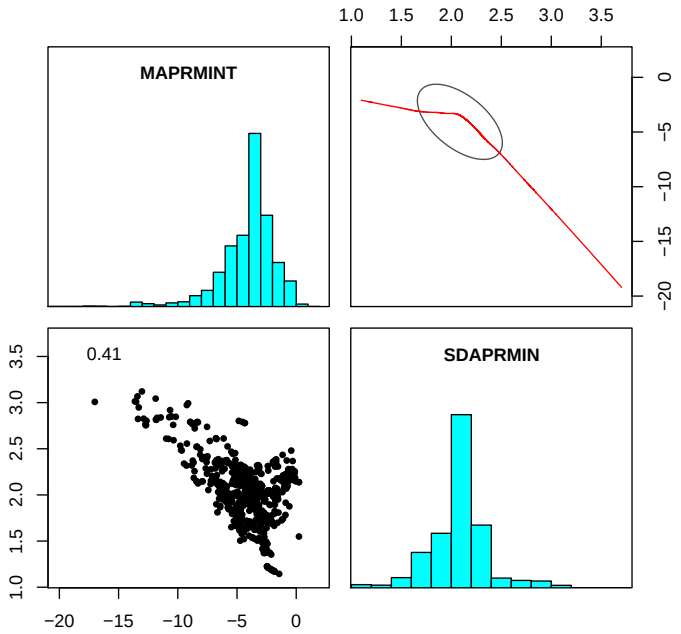

Figure 52: April Minimum Temperature

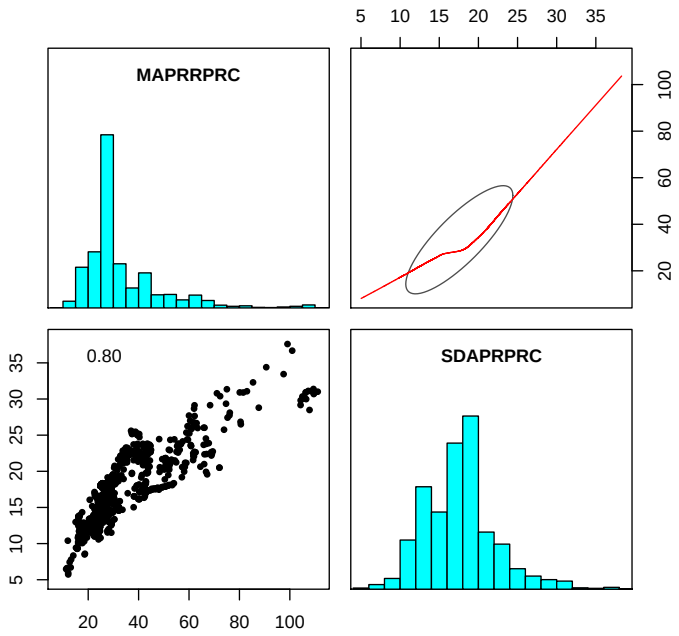

Figure 53: April Mean Precipitation

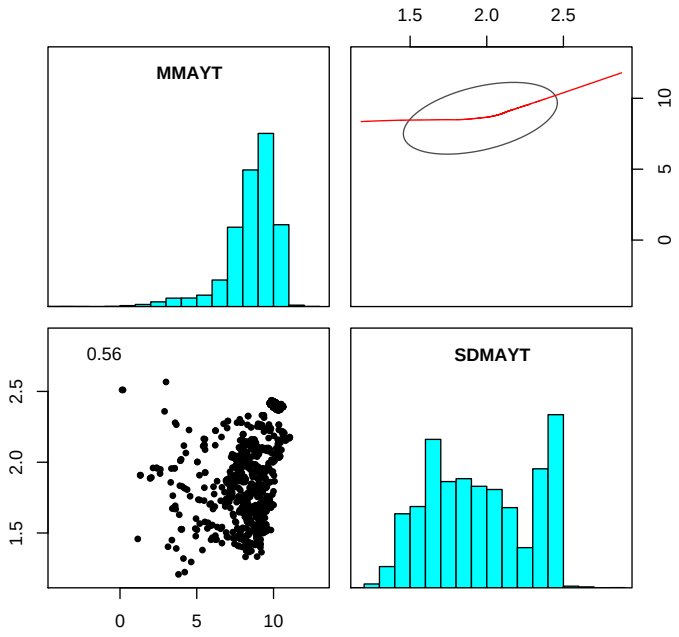

Figure 54: May Mean Temparature

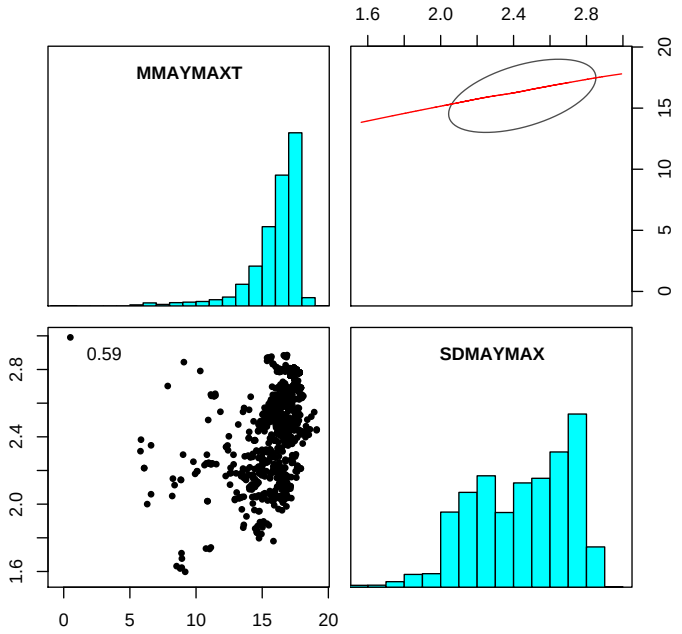

Figure 55: May Maximum temperature

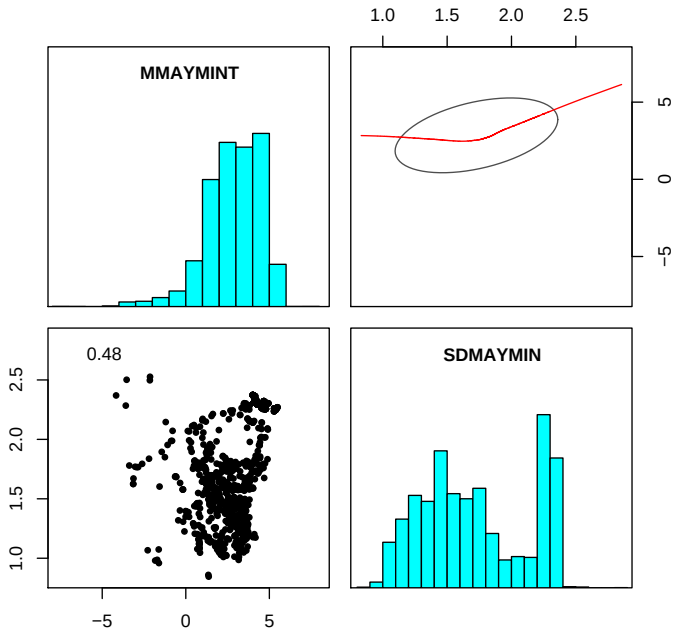

Figure 56: May Minimum temperature

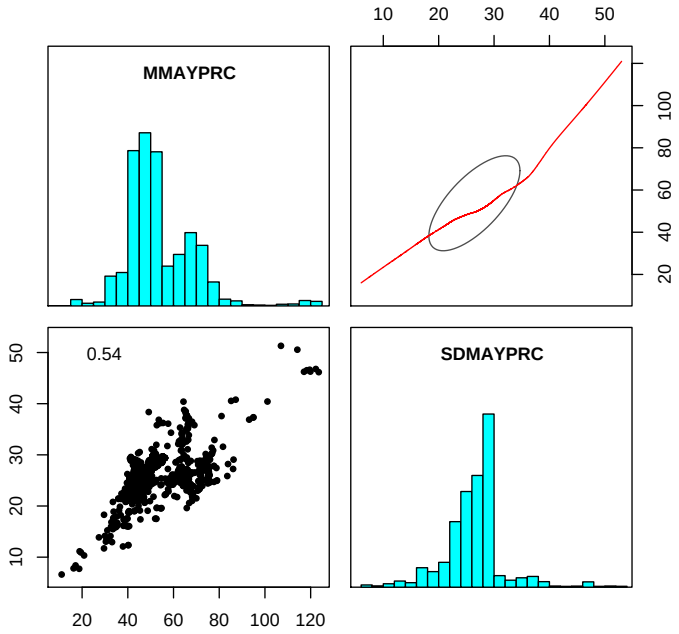

Figure 57: May Mean Precipitation

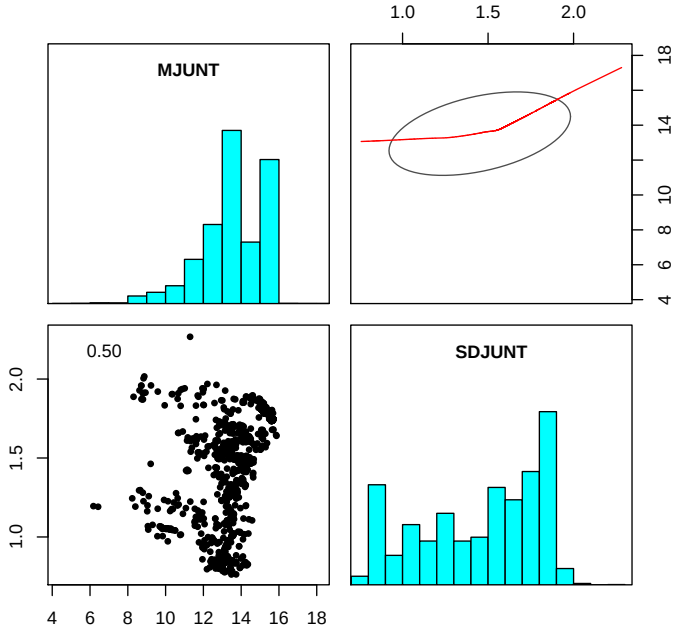

Figure 58: June Mean Temperature

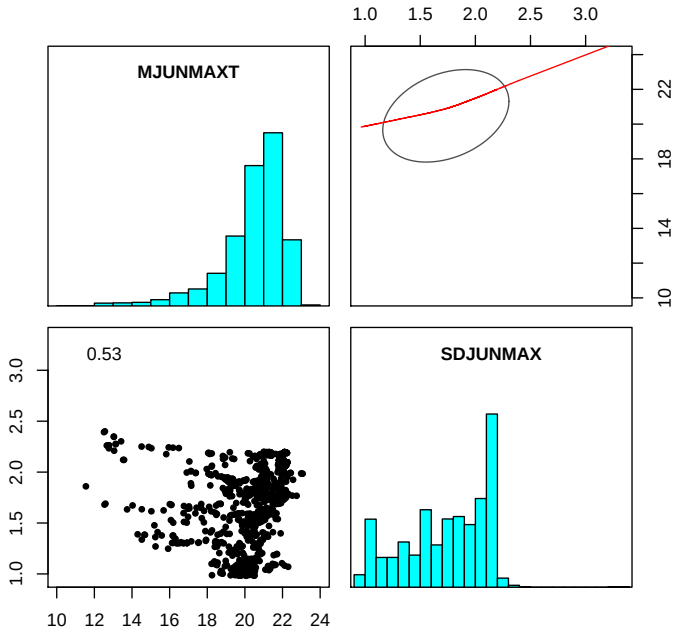

Figure 59: June Maximum Temperature

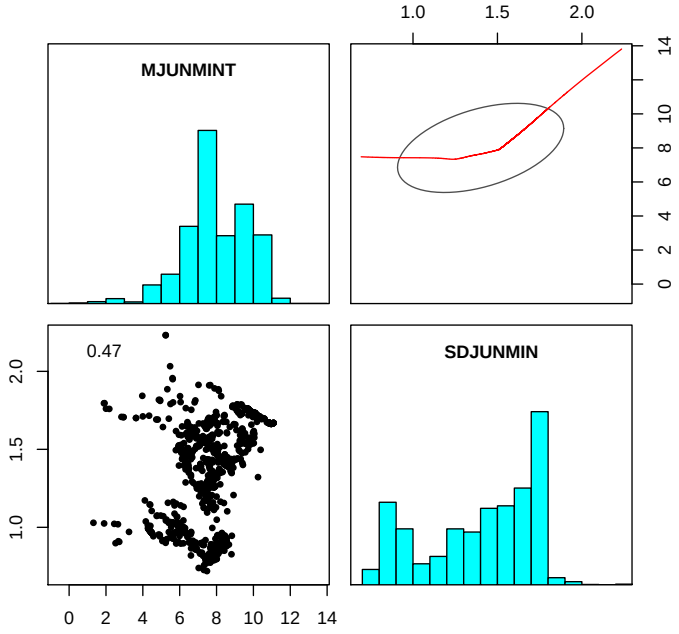

Figure 60: June Minimum Temperature

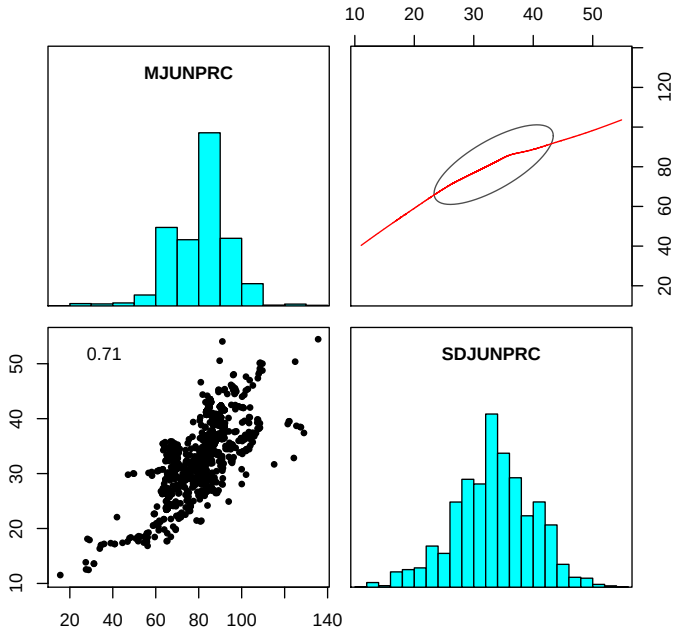

Figure 61: June Mean Precipitation

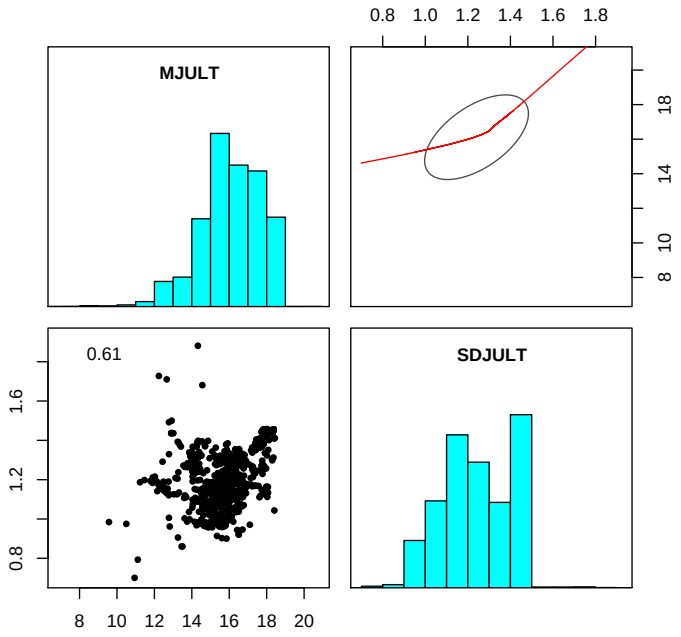

Figure 62: July Mean Temparature

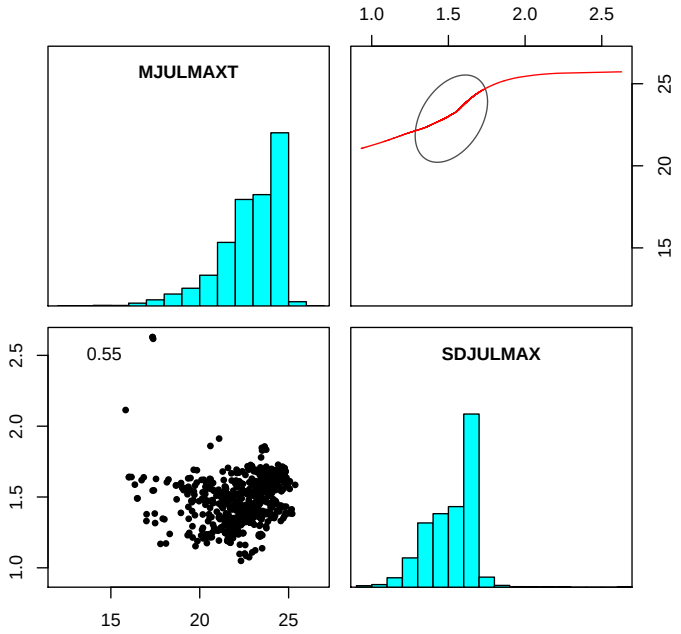

Figure 63: July Maximum Temperature

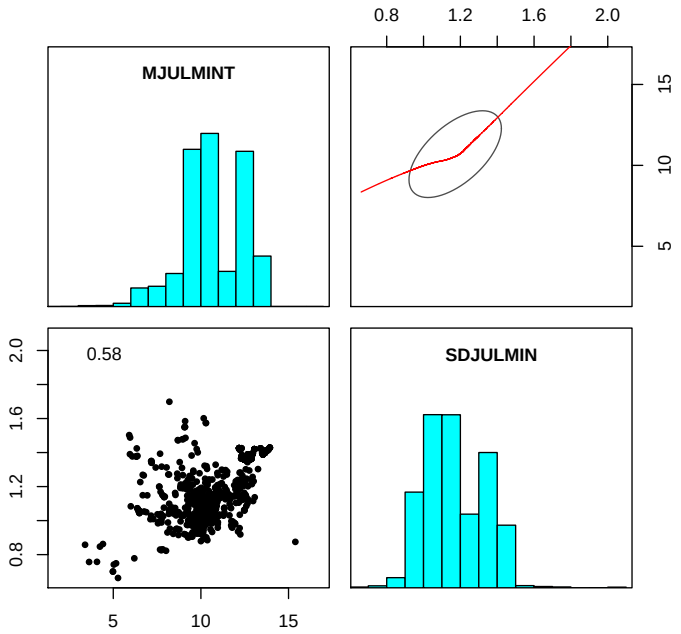

Figure 64: July Minimum Temperature

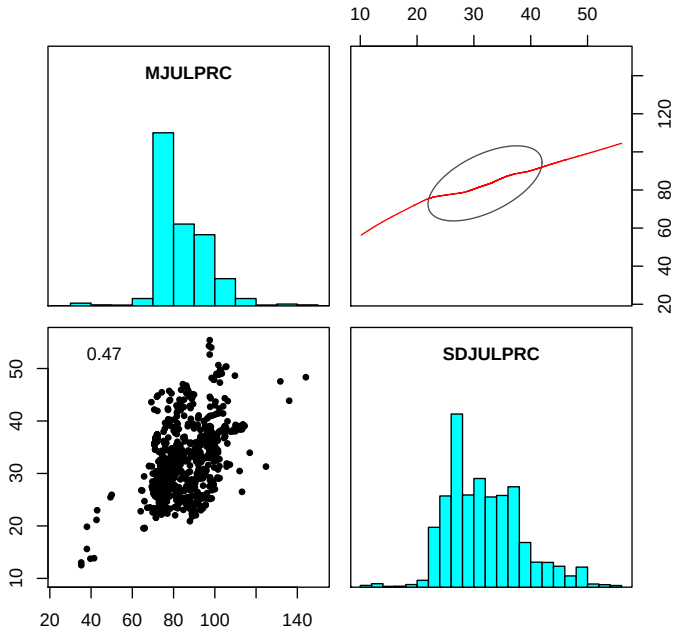

Figure 65: July Mean Precipitation

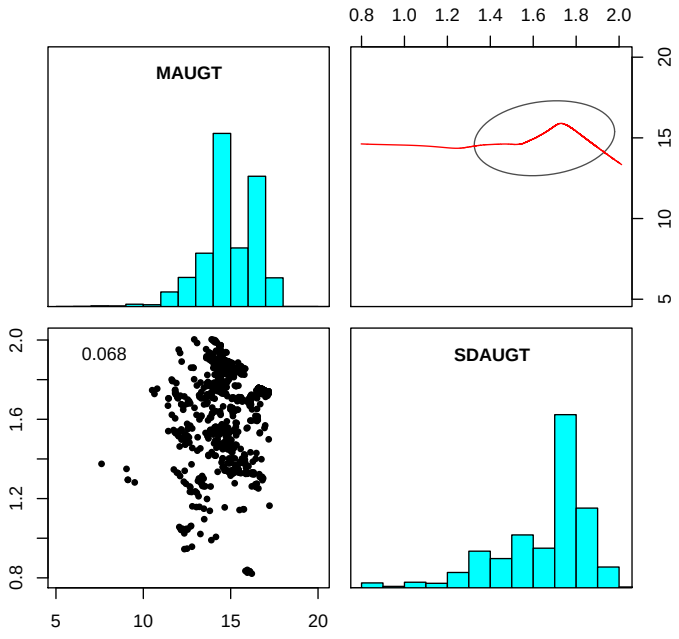

Figure 66: August Mean Temparature

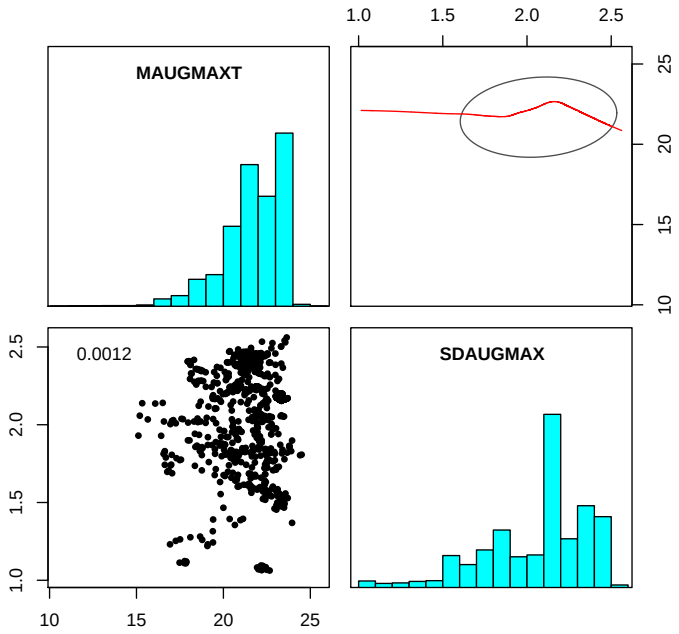

Figure 67: August Maximum Temperature

2.3 seasonality variables

Seasonality refers generally to variation in climate conditions throughout over the course of the year. We have one primary seasonality index in that sense. We also include measures of the start and end and length of the "growing season" as related to the onset of budburst in boreal tree species in spring and the onset of senescence in the autumn. To this we add a growing season heat sums, in terms of degree days above the the threshold temperature for bud-burst over the growing season. There variables are all reported as 30-year normals. Their complete definitions are given in McKenney et al. (2007), which we summarise below.The growing season was determined using temperature-based rules, starting when the mean daily temperature was greater than or equal to 5° Celsius for 5 consecutive days beginning March 1. The growing season ends when the average minimum temperature is less than - 2° degrees Celsius beginning August 1. These rules are aimed more towards defining a growing season for tree species than agricultural crops as they are more clearly related to a frost free period (McKenney et al., 2007). Daily climate estimates needed to calculate Julian day number of start of growing season, Julian day number at end of growing season and Growing Degree Days were obtained by fitting a cumulative value function through the monthly values and then reading off the daily values using a cubic Bessel interpolation (**?**). The growing season variables are approximate measurs of energy balance at a given location (Nix, 1986).For migratory birds, the start or duration of the growing season might also be factors limiting geographic distribution. The specific variables and names are as follows:

- Julian day number of start of growing season (JULDAYST): Julian date when the mean daily temperature is greater than or equal to 5° Celsius for 5 days in a row
- Julian day number at end of growing season (JULDAYEN): Julian date after August 1st when the minimum temperature reaches more than 2° Celsius
- Number of days in growing season (NUDASGS): StartGS - EndGS
- Growing Degree Days (GDDSETHR): Growing Season Heat Sum
- Seasonality (SEAS): Standard deviation of monthly mean temperature estimates expressed as a % of their mean.

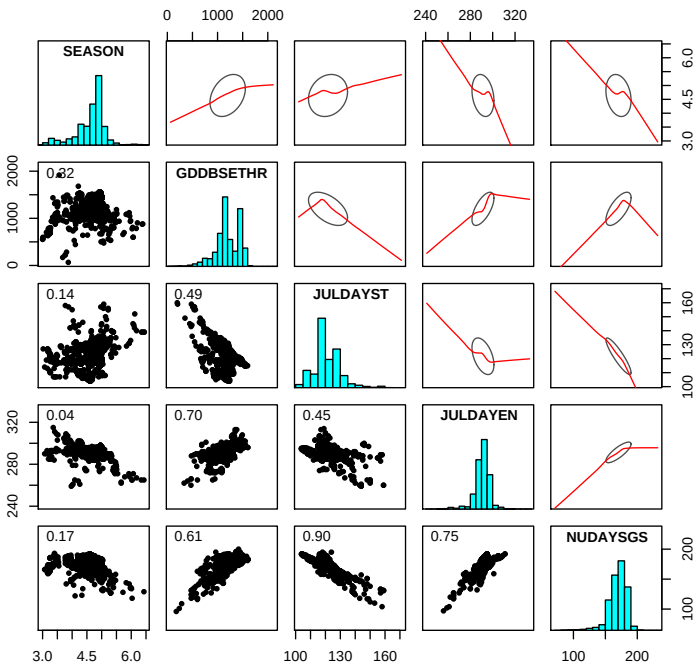

Figure 70: Seasonality Indices

2.4 ecophysiological variables

We use two measures integrated measures of temperature, solar energy and precipitation as indicators of ecophysiological limits on plant productivity. For definitions and details of calculations, see Hogg (1994) and references therein. We use 30year normals of annual values, over the period 1961-1990, as available from NRCAN. These values were not interpolated to the locations of point count stations. Rather, station locations were registered to the standard 300 arc-second climate grid.

Potential evapotranspiration (PET) is a measure of the ability of the atmosphere to remove water from the surface through the processes of evaporation and transpiration assuming no limit on water supply. PET is usually estimated using meteorological parameters, with corrections for latitude to account for sun-angle and day-length, PET was calculated from the Penman-Monteith method.

Climatic moisture index (CMI) is a measure of the water balance of an area in terms of gains from precipitation (P) and losses from potential evapotranspiration (PET). The annual climatic moisture index (CMI) is calculated thus as total mean annual precipitation (P) minus mean annual potential evapotranspiration (PET).

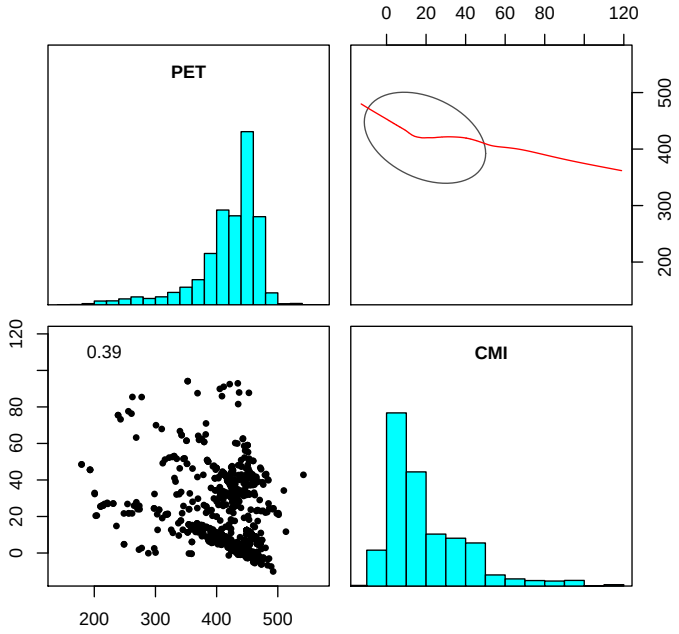

Figure 71: Ecophysiological Indices

3 Remothly sensed variables

Our present suite of remotely-sensed covariates are all derived from Moderate Resolution Imaging Spectroradiometer (MODIS) imagery obtained via satellite. The spatial resolution of these data is 250m or some cases 1km. The data products include landcover, annual measures of gross and net productivity, and monthly measures of leaf area index and a vegetation index. All products are national in extent. Note that many have unique data values coding for missing or invalid data due e.g. to clouds. These special values must be taken into account in all analysis. In the following sections, we represent the data layers as sampled by our station locations, in the form of histograms, correlograms or otherwise as appropriate to the particular data product.

3.1 Landcover

Our categorical landcover variables are based 2005 Land Cover Classification (LCC05) from Natural Resources Canada. It classifies habitats into 31 different categories obtained by multivariate analysis of MODIS multispectral imagery. To facilitate comparisons, we collapsed those into 17 habitat types shown on the figure 73. The original 31 LCC05 classes include 12 types of forest; 3 of shurbland; 7 of tundra/grasslands; 7 classes of developed land types including cropland, mosaic and built-up areas; and 2 water types.

In our current habitat modelling excercises, we use reclassified the 31 LCC05 classes down to 17 classes. The reclassification was based on the distribution of classes within the boreal and within our sample of stations, general concepts of geography, and inspection of estimated mean abundances by classe for several widely distributed boreal songbird species. The reclassification attempted to reduce the numbre of classes to as few as possible while respecting geographic patterns, the hierarchical structure of the LCC05 legend, and our data's ability to distinguish the classes based on the selected species. For more information on Land cover classification of Canada, please refer to `ftp:.nrcan.gc.ca2005`.

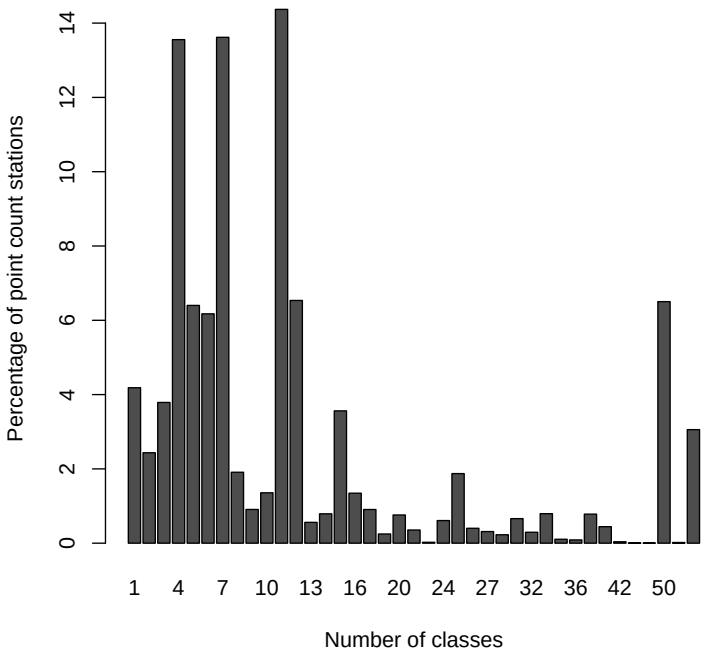

Figure 72: Old Land Cover

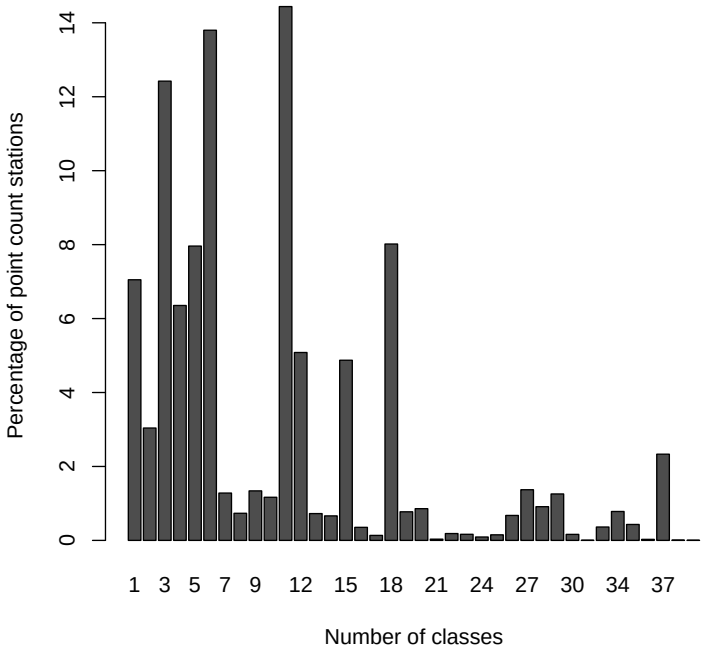

Figure 73: New Land Cover

3.2 Vegetation Continuous Fields

The **Vegetation Continuous Fields**(**VCF**) data are proportional cover estimates for three broad vegetation or landcover types: woody vegetation, herbaceous vegetation, and bare ground. The intent of the VCFs was to reduce some of the effects of arbitrary categorisation in the LCC05, by maintaining a multivariate representation of the cover in each pixel. The data are illustrated by a ternary plot(Figure 74). For more information on Vegetation Continuous Fields (VCF) data, please refer to `http://glcf.umiacs.umd.edu/index.shtml` .

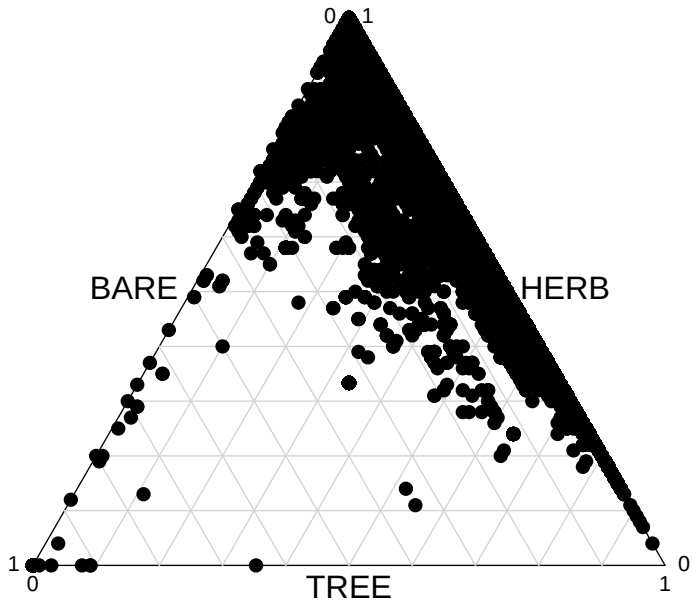

Figure 74: Ternary plot of BARE, TREE and HERB

3.3 Productivity

We evaluated measures of Gross and Net Primary Productivity. **Gross primary production** or **GPP** is the total amount of CO2 fixed by photosynthesis per unit area and time. **Net Primary Production** or **NPP** is the amount of GPP remaining after including the costs of plant respiration. Therefore, **NPP** = **GPP** - **Ra** where **Ra**= Autotrophic respiration. We used pixel-level means of 6yr of data from 2000 through 2006. We here present the birvariate correlation and univariate distributions of the two measures as a correlogram (Figure 75). These two measures are complex model-based derivatives or raw multip-spectral images, and an explanation is beyond the scope of the report. For more information on GPP and NPP, please refer to `http://www.ntsg.umt.edu/`.

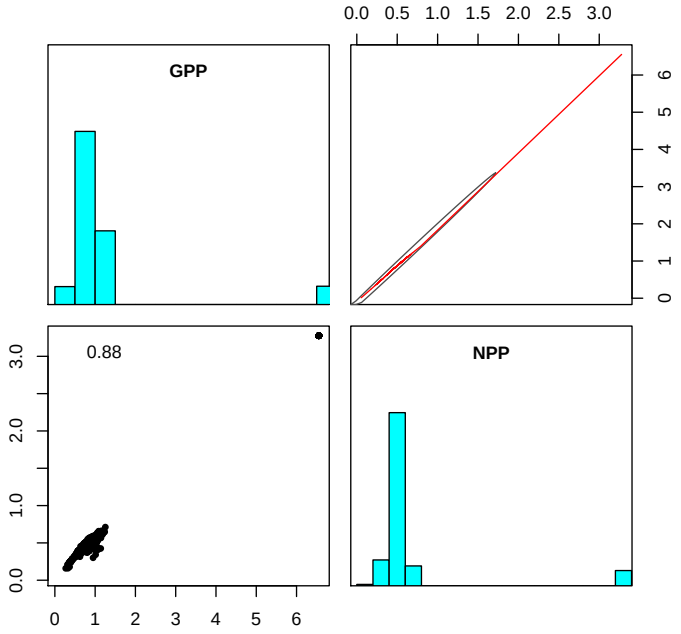

Figure 75: Correlogram of GPP and NPP

3.4 Greenness Indices

3.4.1 Monthly Leaf Area Index

Leaf Area Index or **LAI** (m2/m2) represents the equivalent surface areas of all photosynthetic material in a pixel. It is geometrically defined as the total one-sided area of photosynthetic tissue per unit ground surface area. LAI is a dimensionless value, typically ranging from 0 for bare ground to 6 for a dense forest. Our data were assembled into mean monthly values by acquiring 7 years (1998-2005) of data. LAI was available for all 12 calendar months. We here present correlograms for the summer, fall and winter seasons as defined above. For more information on LAI, please refer to `ftp://ftp.ccrs.nrcan.gc.ca/ad/EMS/LAI/vgtlai/`.

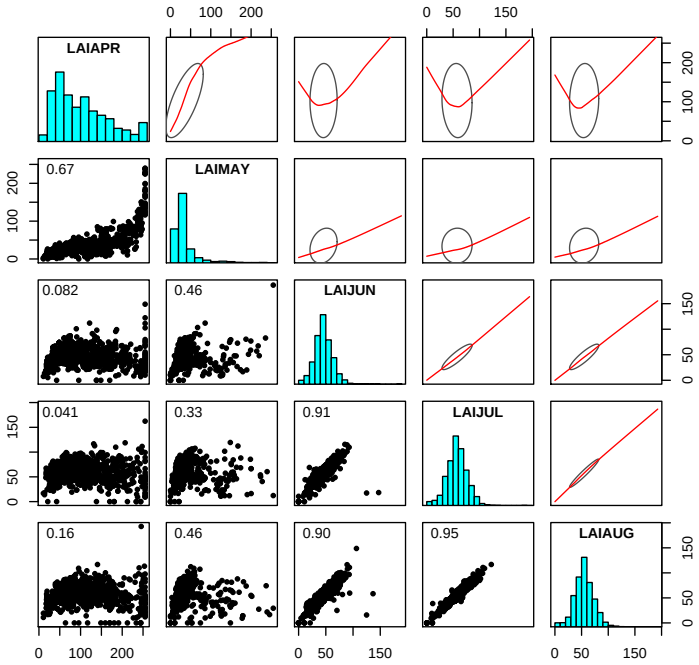

Figure 76: Leaf Area Index for the summer period

3.4.2 Normalized Difference Vegetation Index

The **Normalised Difference Vegetation Index** (**NDVI**) is a simple numerical indicator of photosynthetic activity calculated from the ratios of near and far infrared reflectance. It is a crude measure of the abundance of photosynthetically active material within a pixel. Our data are means of monthly values over 7 years (1998-2005). NDVI was not calculated for all months. For more information on NDVI, please refer to `http://dweb.ccrs.nrcan.gc.ca/rcvcc/index.html`.

3.4.3 Comparison of Monthly Greeness Indices for the summer period

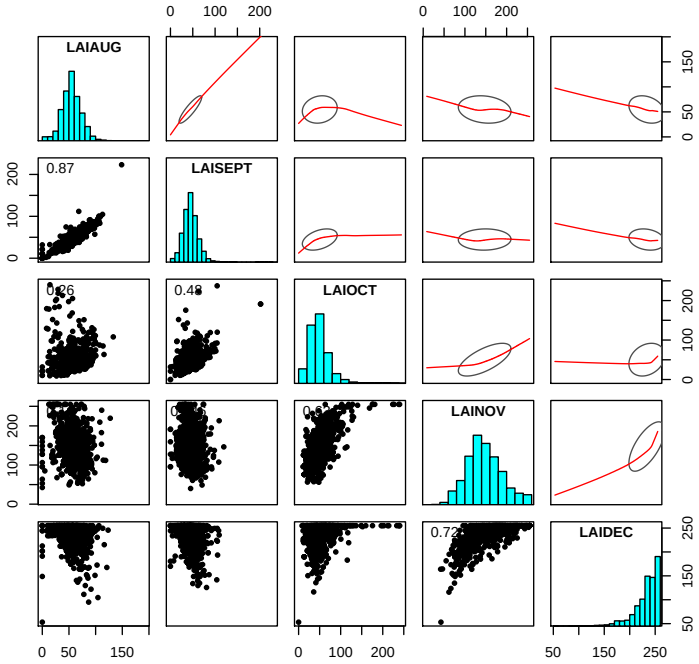

Figure 77: Leaf Area Index for the fall period

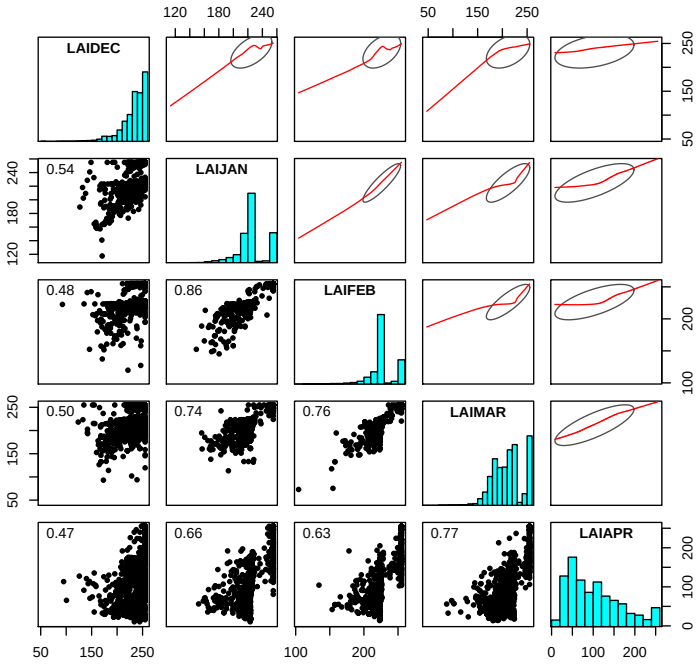

Figure 78: Leaf Area Index for the winter period

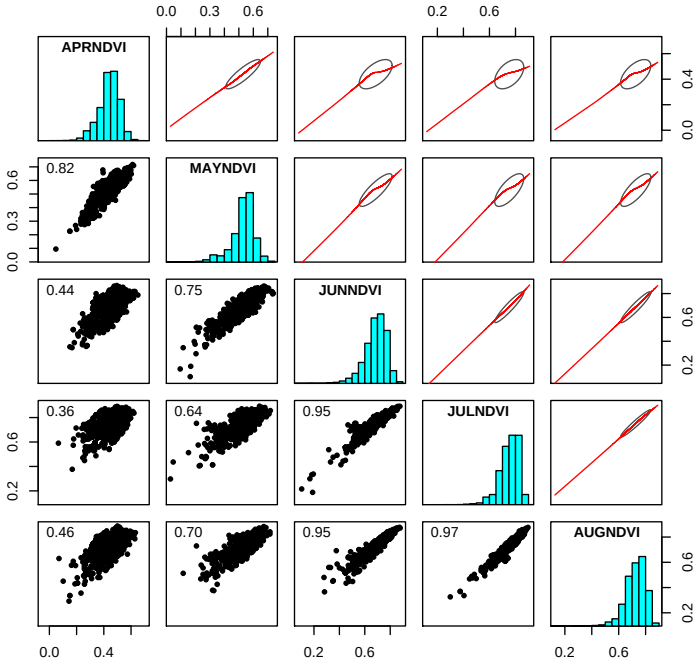

Figure 79: Normalized Difference Vegetation Index for the summer period

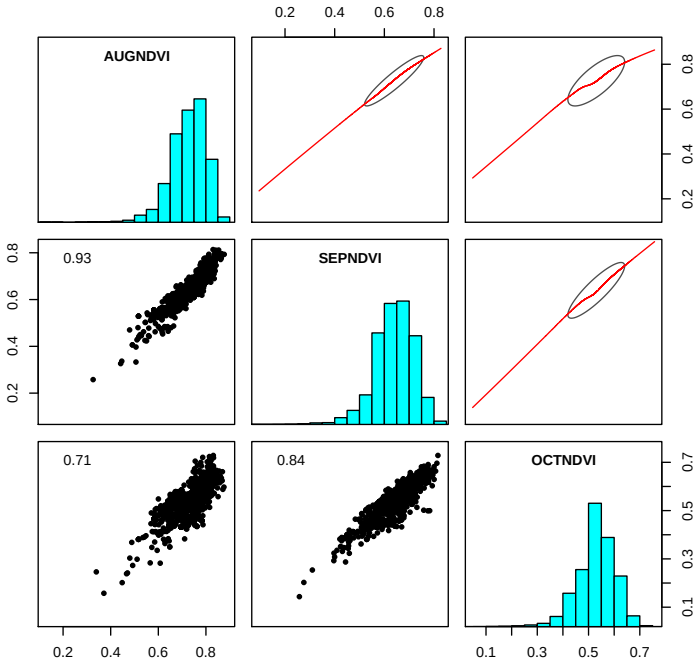

Figure 80: Normalized Difference Vegetation Index for the fall period

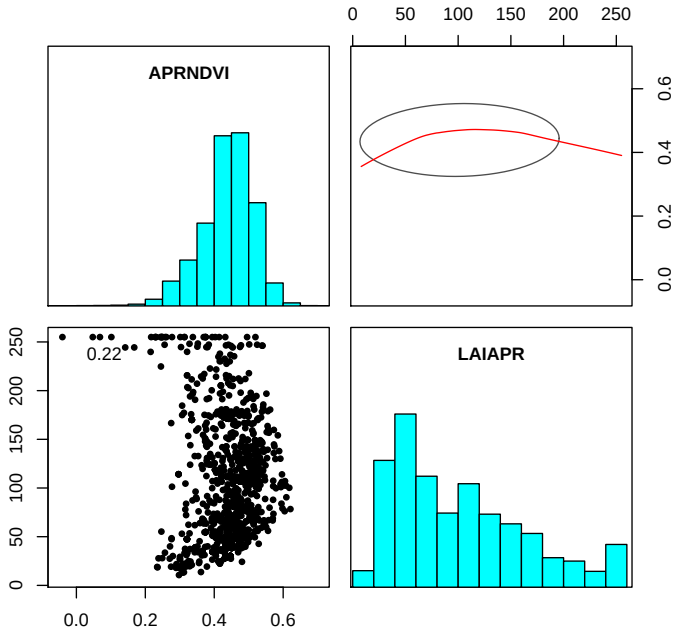

Figure 81: Comparison of LAI and NDVI for the month of April

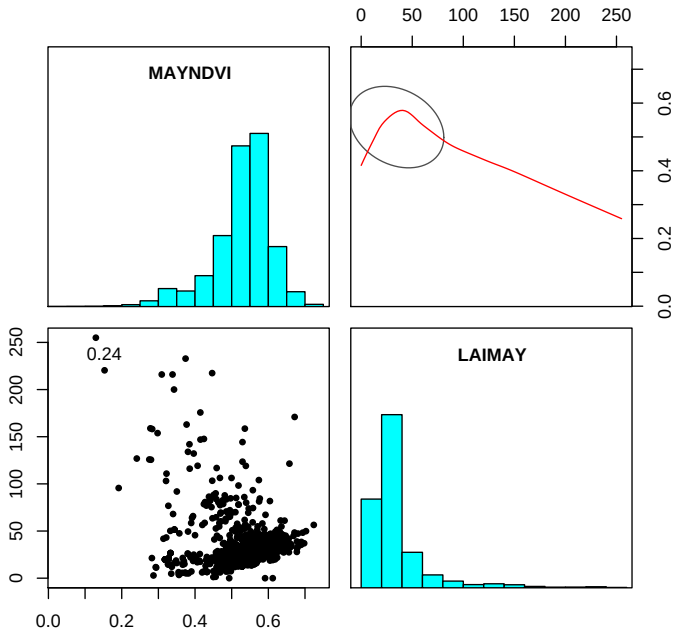

Figure 82: Comparison of LAI and NDVI for the month of May

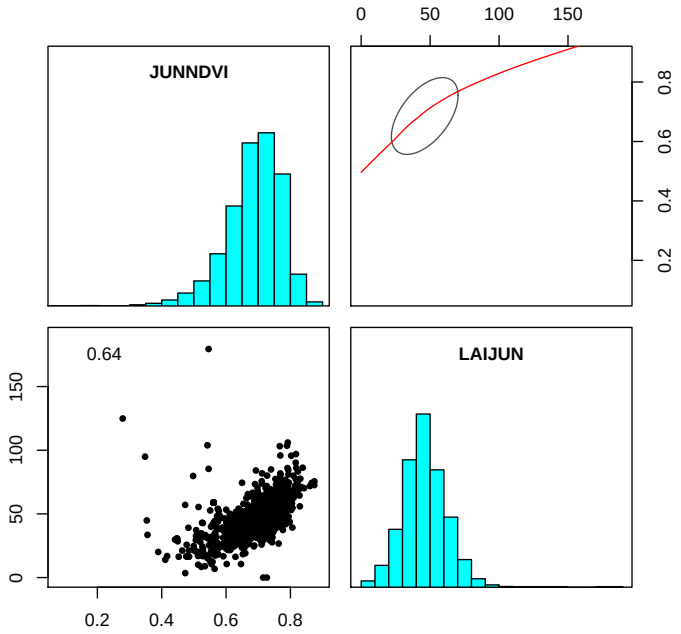

Figure 83: Comparison of LAI and NDVI for the month of June

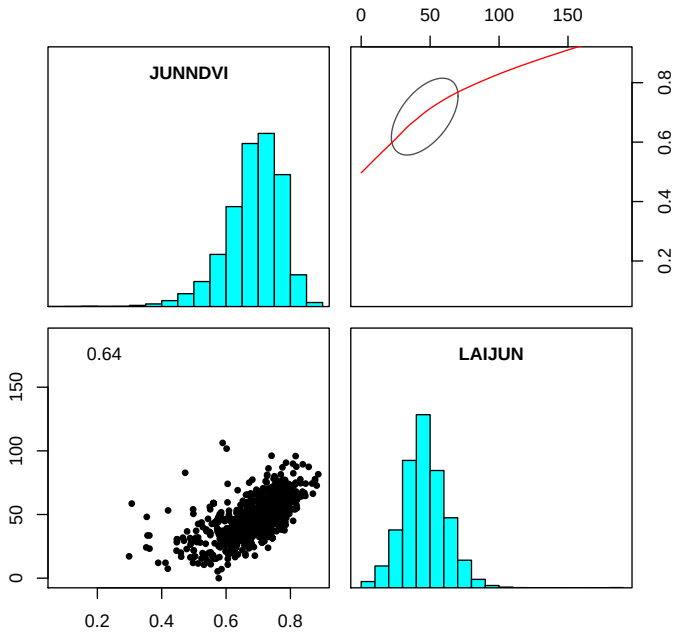

Figure 84: Comparison of LAI and NDVI for the month of July

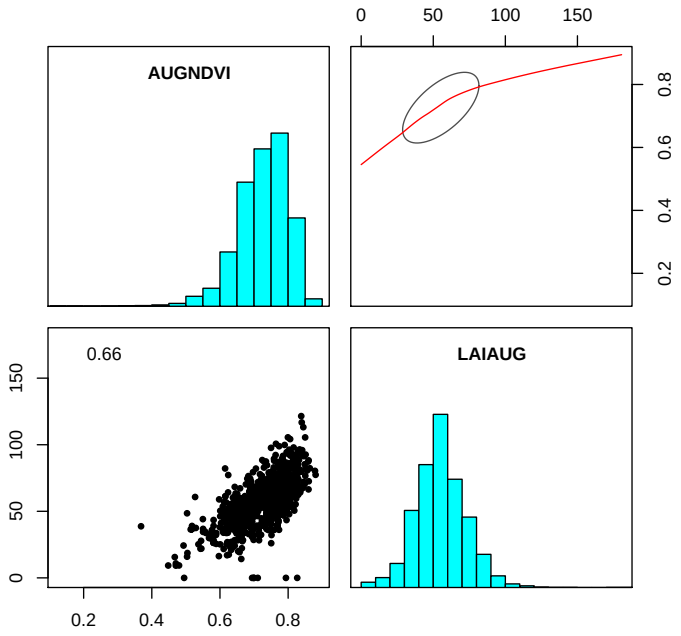

Figure 85: Comparison of LAI and NDVI for the month of August

References

M. Firendly. Corrgrams: Exploratory display for correlation matrices. *The American Statistician*, 15:1–16, 2002.

E.H. Hogg. Climate and the southern limit of the western Canadian boreal forest. *Canadian Journal Forestry Research*, 24:1835–1845, 1994.

D. McKenney, P. Papadopol, K. Campbell, K. Lawrence, and M. Hutchinson. Spatial models of Canada- and North America - wide 1971/2000 minimum and maximum temperature, total precipitation and derived bioclimatique variables. *Frontline Technical Note No.106.*, Great Lakes Forestry Centre: Sault Ste. Marie, ON.:p.9, 2006a.

D. McKenney, P. Papadopol, K. Lawrence, K. Campbell, and M. Hutchinson. Customized spatial climate models for Canada. *Frontline Technical Note No.108.*, Great Lakes Forestry Centre: Sault Ste. Marie, ON.:p.9, 2007.

H. Nix. A biogeographic analysis of Australian elapid snakes. *In: Longmore, R. (Ed.), Atlas of Elapid Snakes of Australia*, Australian Flora and Fauna Series, No.7.(Bureau of Flora and Fauna, Canberra):4–15, 1986.