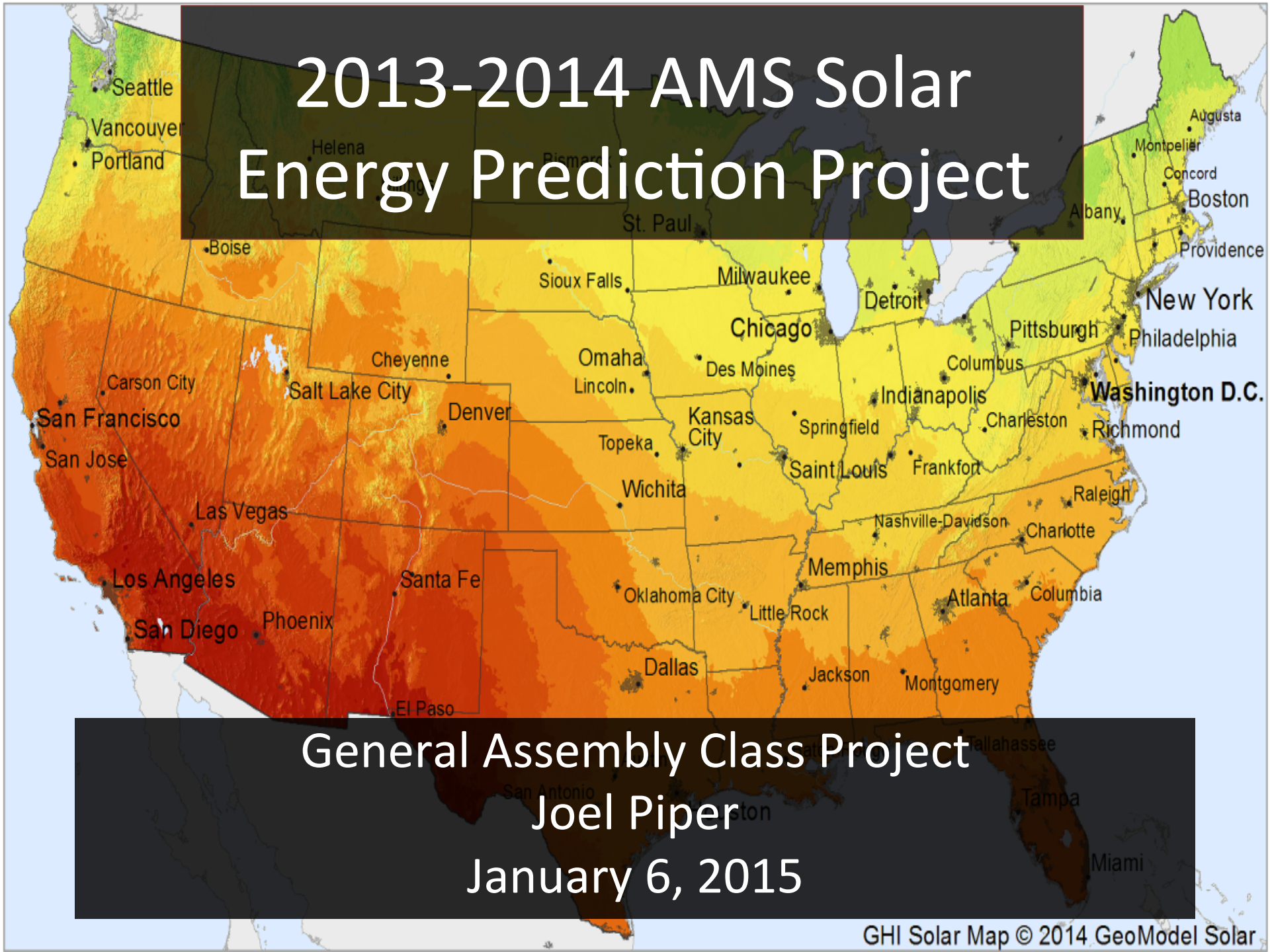


2013-2014 AMS Solar Energy Prediction Project

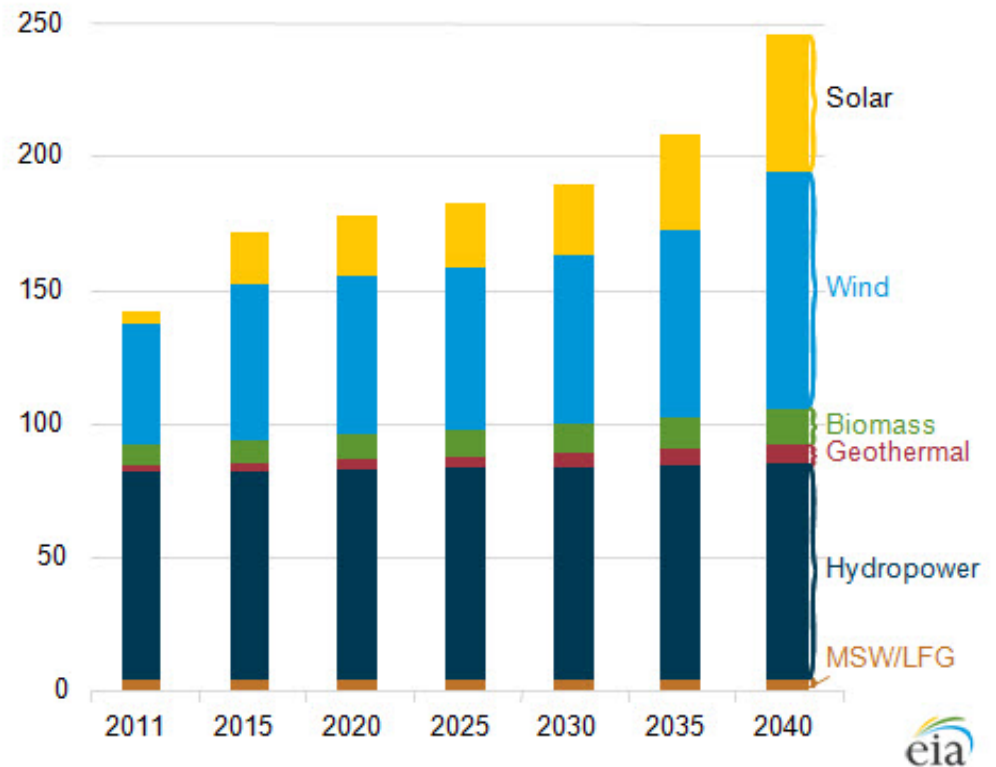


General Assembly Class Project
Joel Piper
January 6, 2015

Solar power is a small but increasing source of energy in the United States

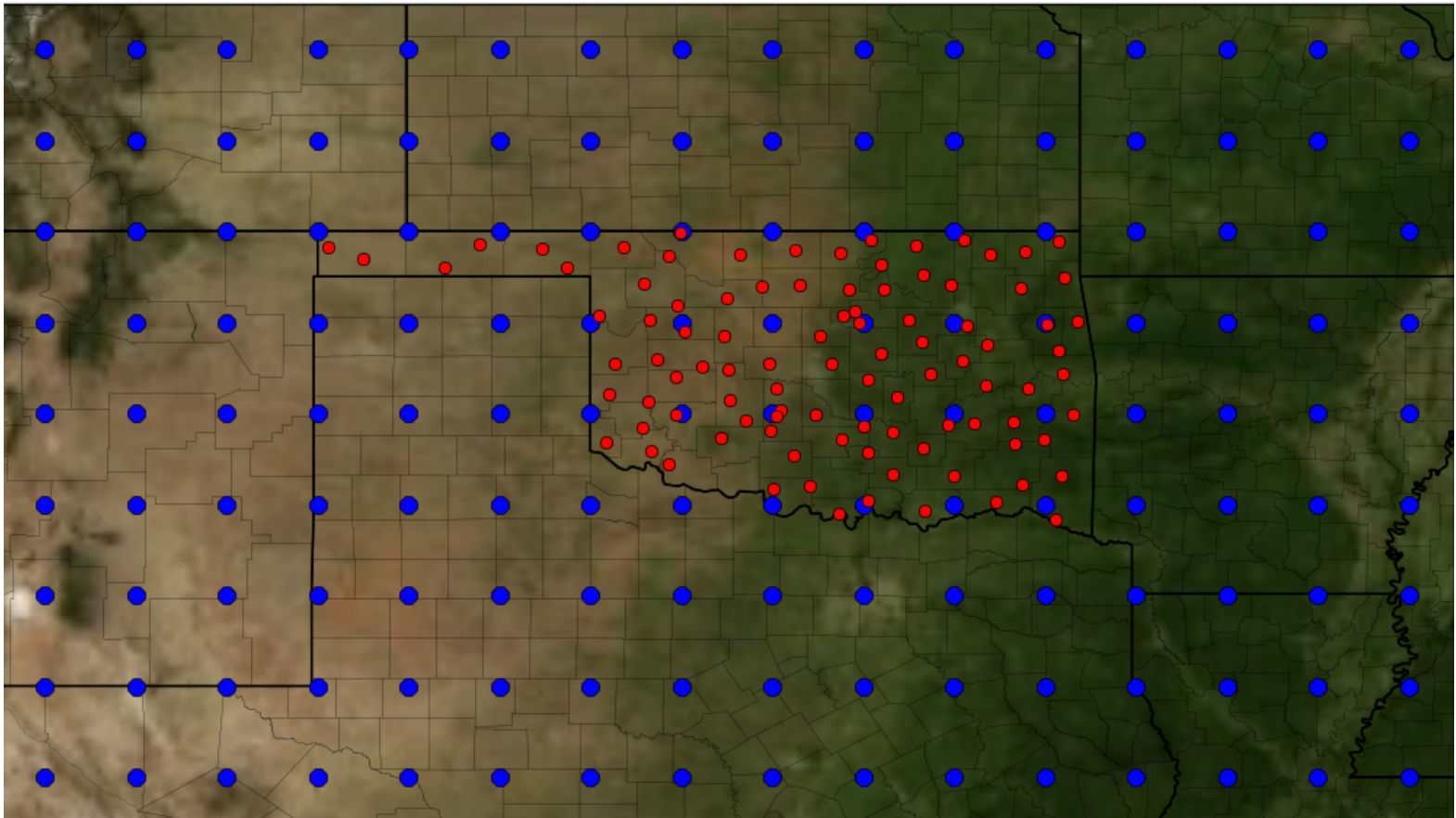
- Two major hurdles in becoming significant energy source:
 - Efficiency
 - Reliability
- Reliability
 - Storage improvements
 - Predictive technologies

Renewable Energy Generating Capacity in the United States (Gigawatts)



Project Goal: Develop Competitive Learning Technique for 2013 Kaggle AMS Solar Energy Challenge

Predictive Question: Predict the daily solar energy deposition at 98 sites across Oklahoma over two years using weather data (including short wavelength and long wavelength radiative flux) from a 16x9 grid of collection stations across the area.



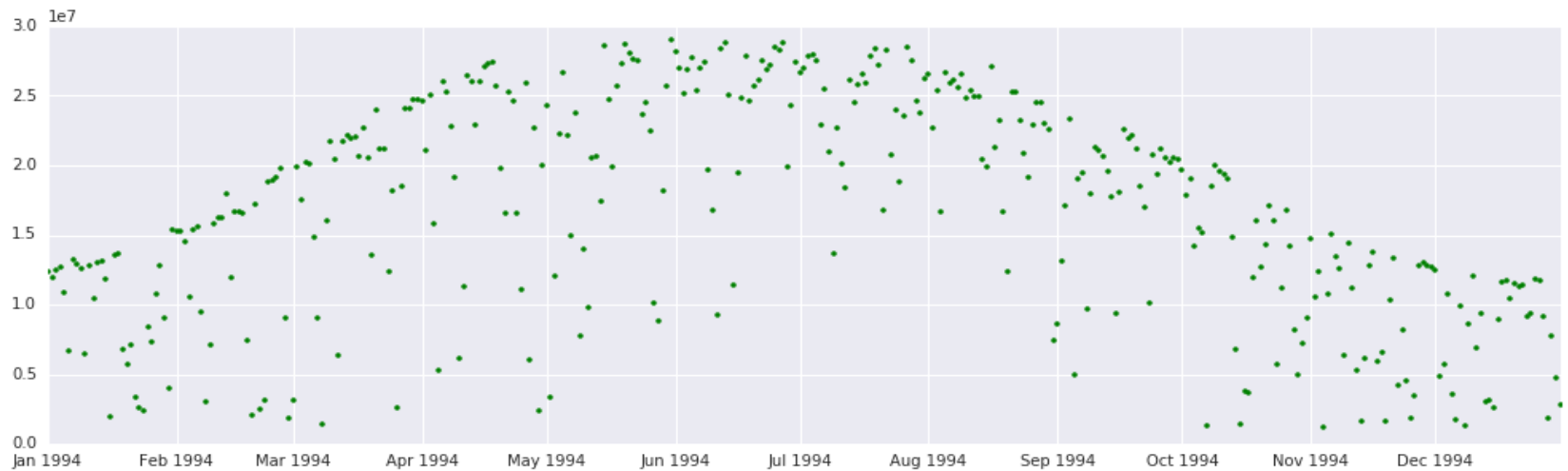
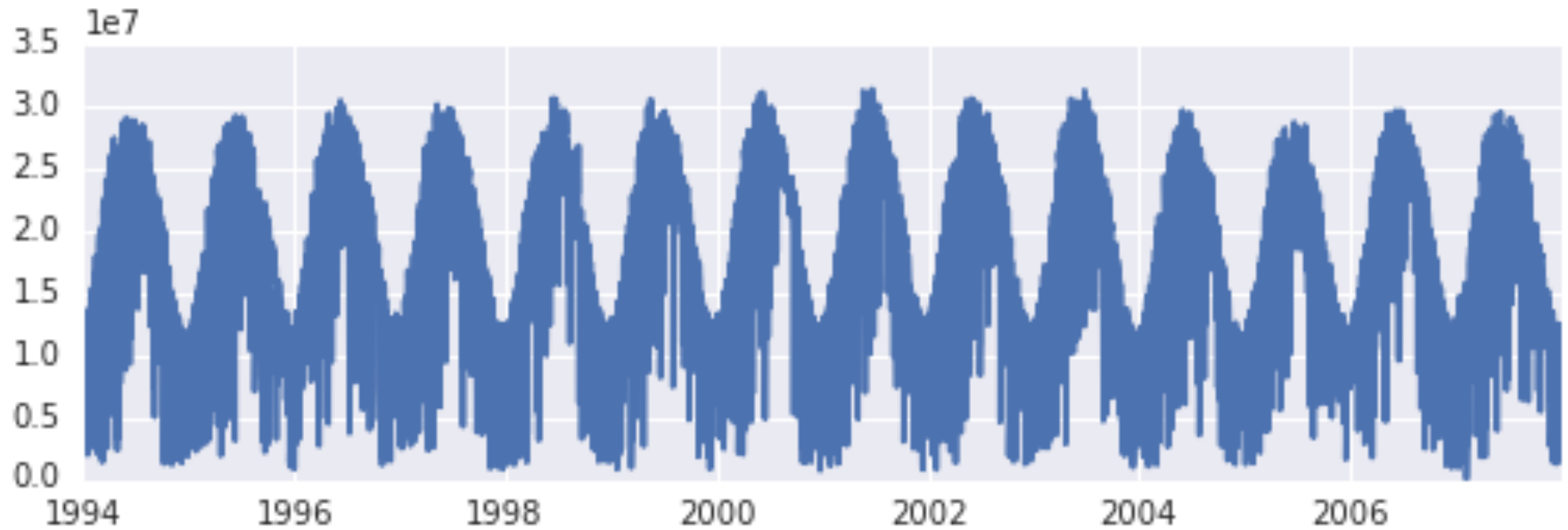
Data- 15 NetCDF input and one CSV output file

- 15 weather-related variables (including precipitation, cloud cover, downward/upward long-wave/short-wave solar flux, max/min/current temp, humidity, pressure)
 - At 144 locations, five times of the day, and eleven ensemble members
- 5113 daily solar energy measurements (J/m^2) from 1994-2007 at 98 Mesonet locations (~500k predictions)

```
<type 'netCDF4._netCDF4.Variable'>
float32 Downward_Short-Wave_Rad_Flux(time, ens, fhour, lat, lon)
  _FillValue: 9999.0
  units: W m-2
  long_name: Downward_Short-Wave_Rad_Flux_Average (Average for Mixed
Intervals) @ surface
  cell_methods: time: mean
  GRIB_param_discipline: Meteorological_products
  GRIB_param_category: Short-wave_Radiation
  GRIB_param_name: Downward_short_wave_rad_flux
  GRIB_generating_process_type: Forecast
  GRIB_param_id: [ 2 0 4 192]
  GRIB_product_definition_template: 8
  GRIB_product_definition_template_desc: Average, accumulation, extreme
values or other statistically processed value at a horizontal level in a time interval
  GRIB_level_type: 1
  GRIB_level_type_name: surface
  GRIB_interval_stat_type: Average
  GRIB_VectorComponentFlag: easterlyNortherlyRelative
unlimited dimensions:
current shape = (5113, 11, 5, 9, 16)
filling on
```

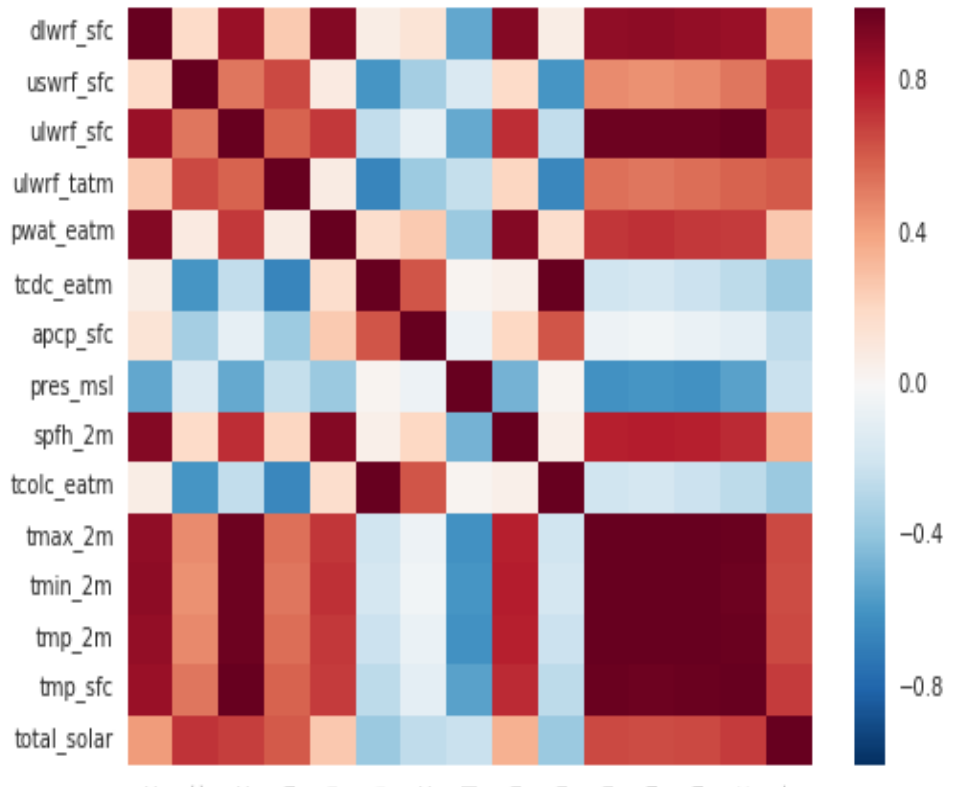
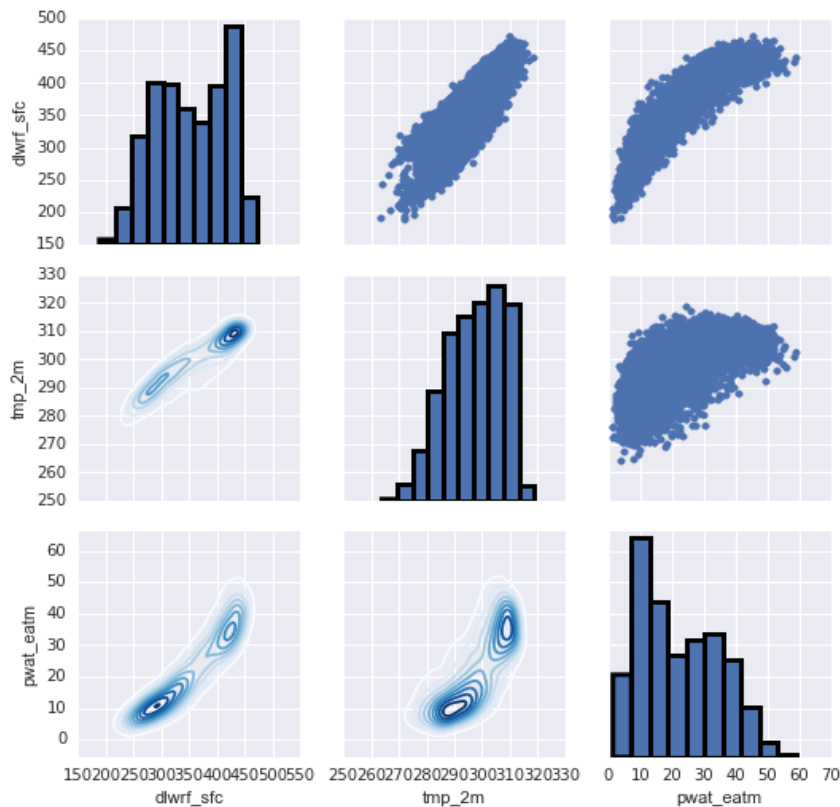
	ADAX	ALTU	APAC	ARNE
Date				
1994-01-02	9778500	10862700	11666400	8062500
1994-01-03	9771900	12627300	12782700	11618400
1994-01-04	6466800	13065300	12817500	12134400
1994-01-05	11545200	8060400	10379400	6918600
1994-01-06	6817200	8157900	7673100	3500400
1994-01-07	12418800	12369900	12873000	12181800
1994-01-08	12375600	12634500	13066500	11608800
1994-01-09	11601000	12156000	12464700	10866000
1994-01-10	3935700	12321900	8164800	11328600

Total Solar Energy (J/m²)



Exploration and Cleaning

Began with a single sample and the surrounding grid of Mesonet points to examine variable correlations.



Next Steps

- Automate tidying data for easier visualization
 - Continue gathering insights from simple visualizations
- Start making predictions at one station
 - Limit the input data points to a surrounding grid
 - Add summary stats across all points or cross-sections (e.g., averages, medians, extremes at all points, similar latitudes, longitudes, elevations)
- Expand to all stations
- Use Apache Spark and MLib to make use of parallel processing algorithms