Final Project Report

Title: Comparing different methods for interpolating weather data
Notice: Dr. Bryan Runck
Author: Timothy Johnson
Date: 16 December 2024

Project Repository: https://github.com/tjjohnson1415/GIS5571
Time Spent: 50 hours

Abstract

Interpolation is used to estimate the value of weather elements (temperature, humidity, pressure, wind, etc.) at locations where it is not measured to produce a continuous surface. There are many methods of interpolation that can be used. First-order inverse distance weighting (IDW), second-order IDW, natural neighbor, spline, and ordinary Kriging were used to interpolated surface based on weather station data retrieved from NOAA and NDAWN. Leave-one-out cross validation was used to calculate RMSE to quantitatively describe the accuracy of each interpolation method for each interpolated weather element. Spline interpolation performed the worst for every weather element. Of the other four methods, the results were comparable and there was not a clear best performance.

Problem Statement

Weather phenomena, such as temperature, pressure, humidity, and wind, form continuous surfaces. No matter how many measurements we take, we cannot measure them everywhere. Instead, weather is measured at designated weather stations. The observations at this discrete number of stations are used to interpolate what would be observed at unknown locations. There are many methods of interpolation. Common deterministic methods include inverse distance weighting, natural neighbor, and spline. Kriging is a family of stochastic interpolation methods that often produce more accurate results than the deterministic methods.

Table 1: Data Requirements

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Weather Station Data	Weather Station Data from NDAWN	Point locations of weather stations	Temperature, Humidity, Pressure, Wind Speed & Direction	NDAWN	
2	Weather Station Data	Weather Station Data from NOAA	Point locations of weather stations	Temperature, Humidity, Pressure, Wind Speed & Direction	NOAA	Pressure converted from Pa to mbar.

Input Data

The datasets listed in Table 2 are weather records for weather stations operated by different organizations. The datasets will be combined to get a denser field of points to interpolate the rest of the surface with.

Table 2: Data

#	Title	Purpose in Analysis	Link to Source
1	Weather Data from NDAWN	Station data for interpolation.	NDAWN
2	Weather Data from NOAA	Station data for interpolation.	NOAA

Methods

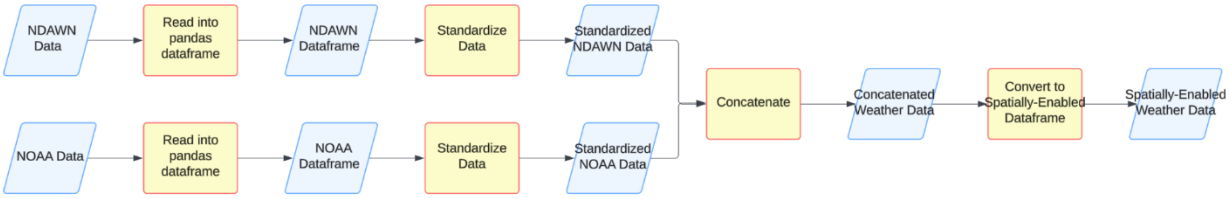


Figure 1: Data flow diagram for preparing data for analysis.

Hourly weather station data containing temperature, humidity, pressure, and wind data was retrieved from both NDAWN and NOAA for March 25, 2024. The data was read into Pandas dataframes and standardized. Many columns needed to be renamed, and the units of pressure needed to be converted before the data could be concatenated. The concatenated weather data was converted to a spatially enabled dataframe.

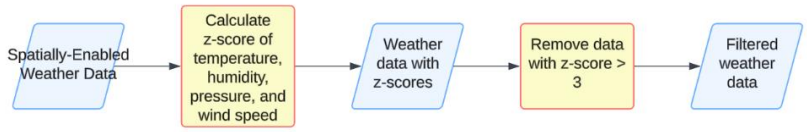


Figure 2: Data flow diagram for quality control.

The data clearly contained some errors, most noticeably in temperature, where a few readings were upwards of 200°F. To filter these out, the z-scores for each temperature, humidity, pressure, and wind speed were calculated and data with z-scores over 3 were removed. This was done individually for each weather element, so that if a station has an errant temperature recording but reasonable recordings for other elements, that station will only not be used when interpolating temperature.

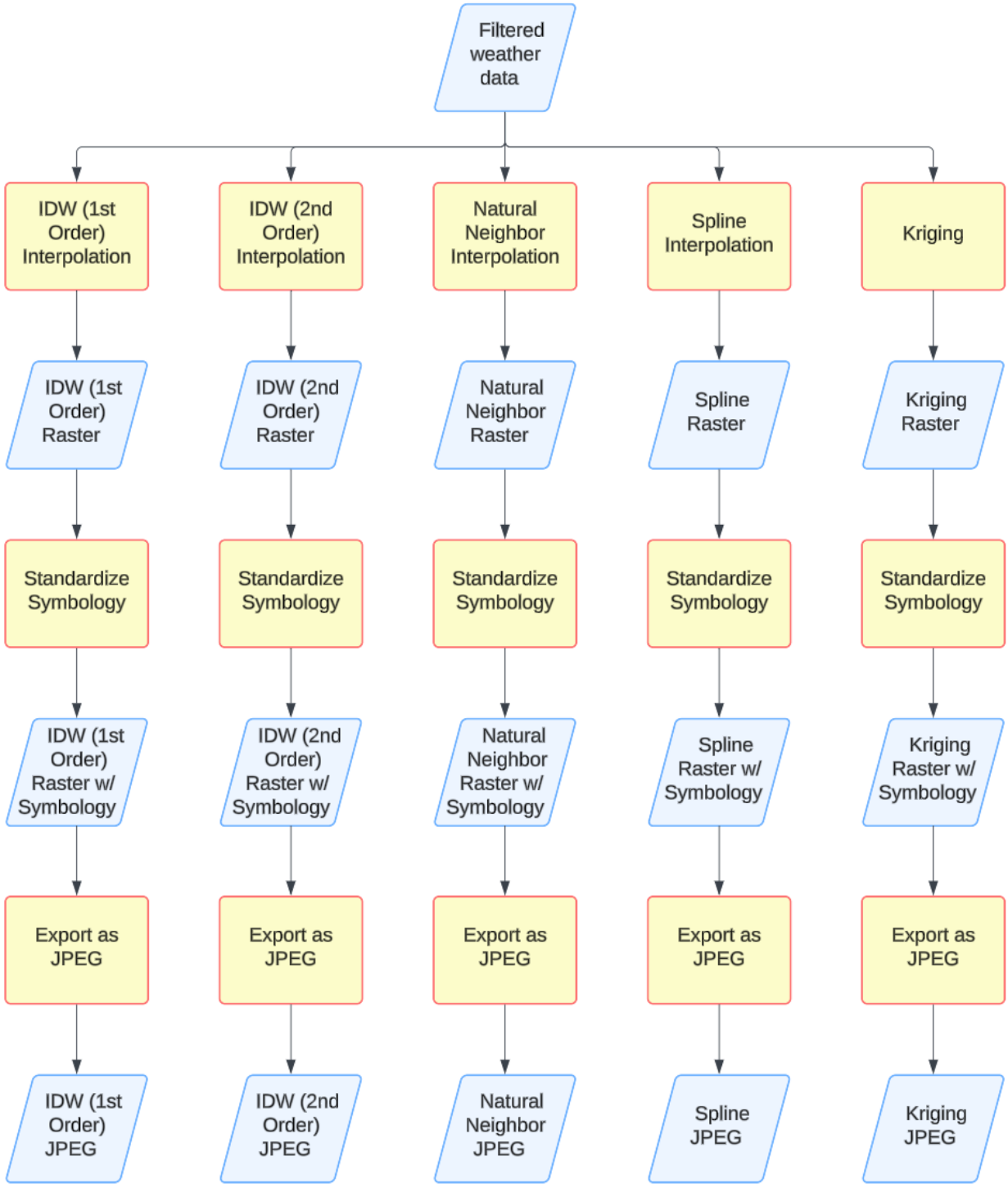


Figure 3: Data flow diagram for interpolating temperature, humidity, and pressure.

After filtering out data with high z-scores, five interpolation methods were used to create surfaces for temperature, humidity, and pressure. This was done for each of the 24 hours of the day. The color scheme of the output rasters was standardized to make them visually comparable before each raster was exported as an image.

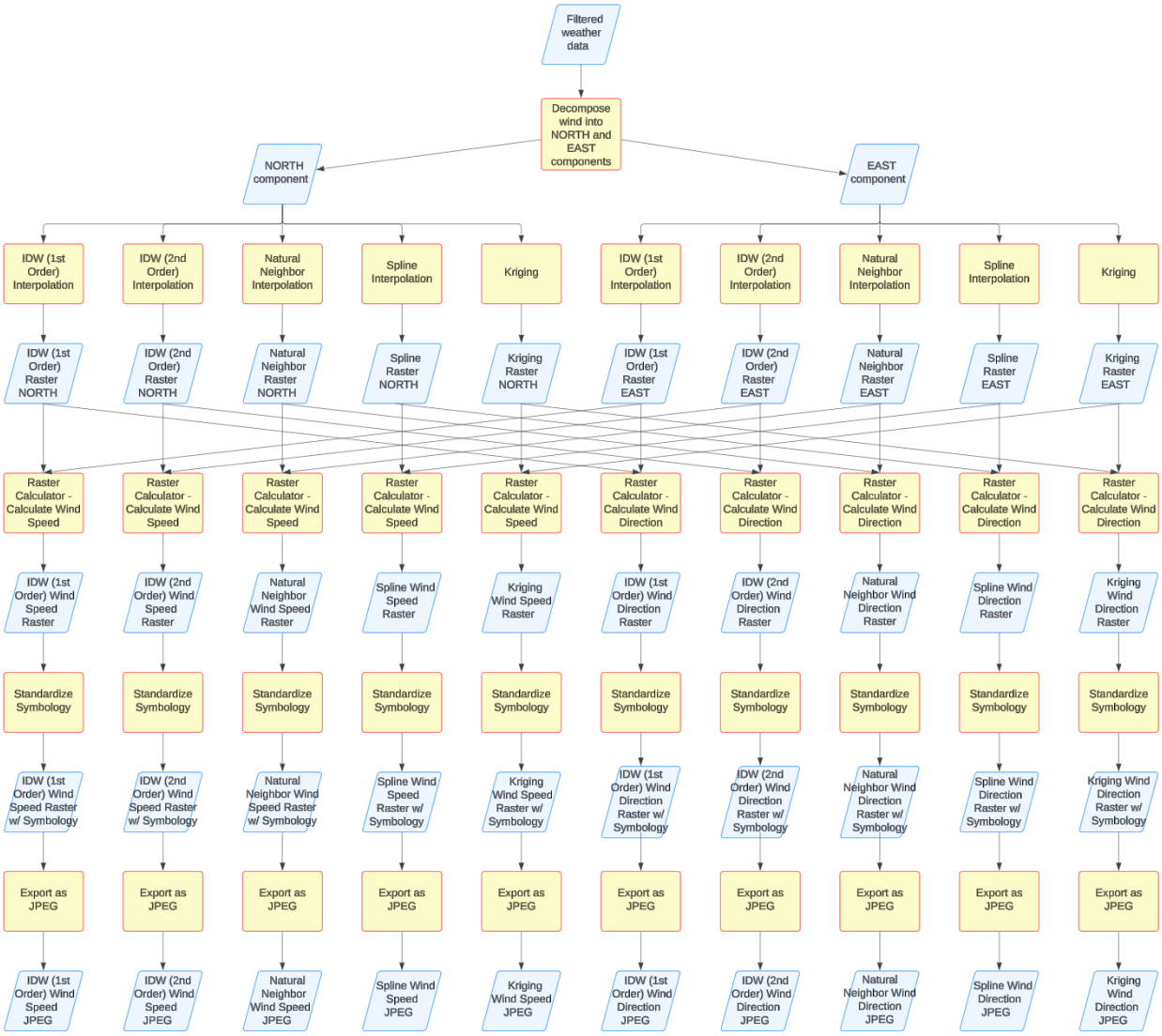


Figure 4: Data flow diagram for interpolating wind speed and direction.

Because wind is a vector quantity, creating an interpolation surface for it is more complex. The station data (which contains both wind speed and direction) was used to calculate the north-south and east-west components of wind. Interpolation surfaces were created for each component individually, and the raster calculator was used to combine the components into rasters of wind speed and direction. This was done for each of the 24 hours of the day. The symbology was standardized, and each raster was exported as an image.

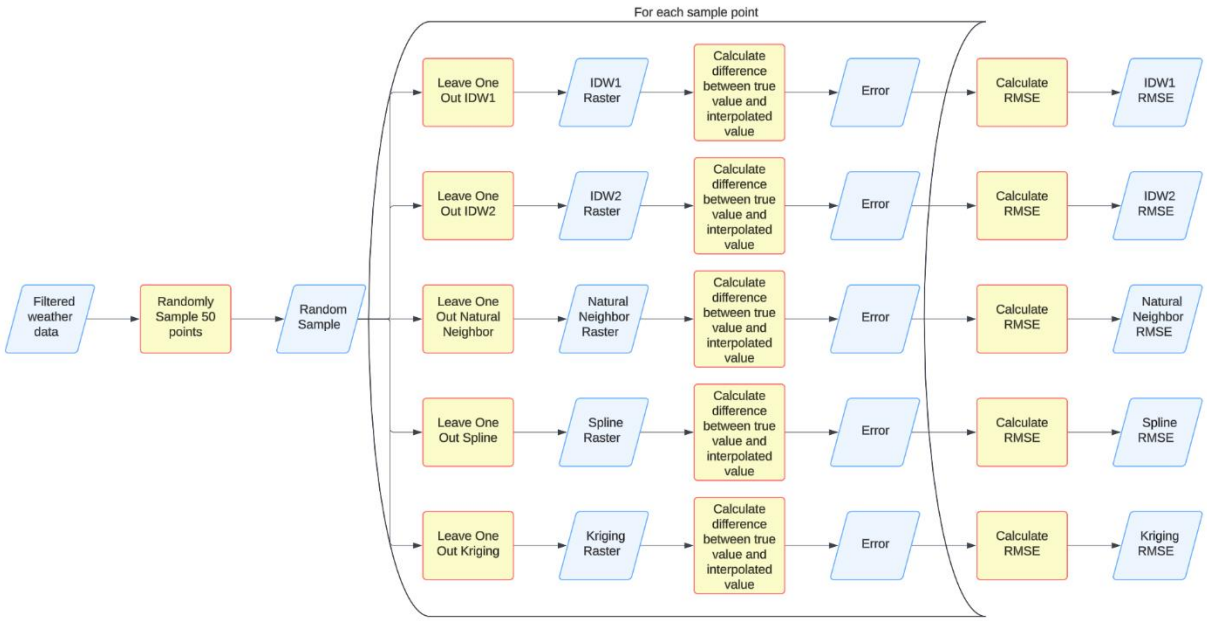


Figure 5: Data flow diagram for verifying results using leave-one-out cross validation.

To validate results and measure how accurate each interpolation method is, leave-one-out cross validation was conducted on a sample of 50 points for each interpolation method and each weather element. The error at each left out point was recorded and used to calculate RMSE.

Results

The following figures show the interpolated surfaces for each weather element for noon on March 25, 2024. See the Appendix for hourly small multiples for each element and each interpolation method.

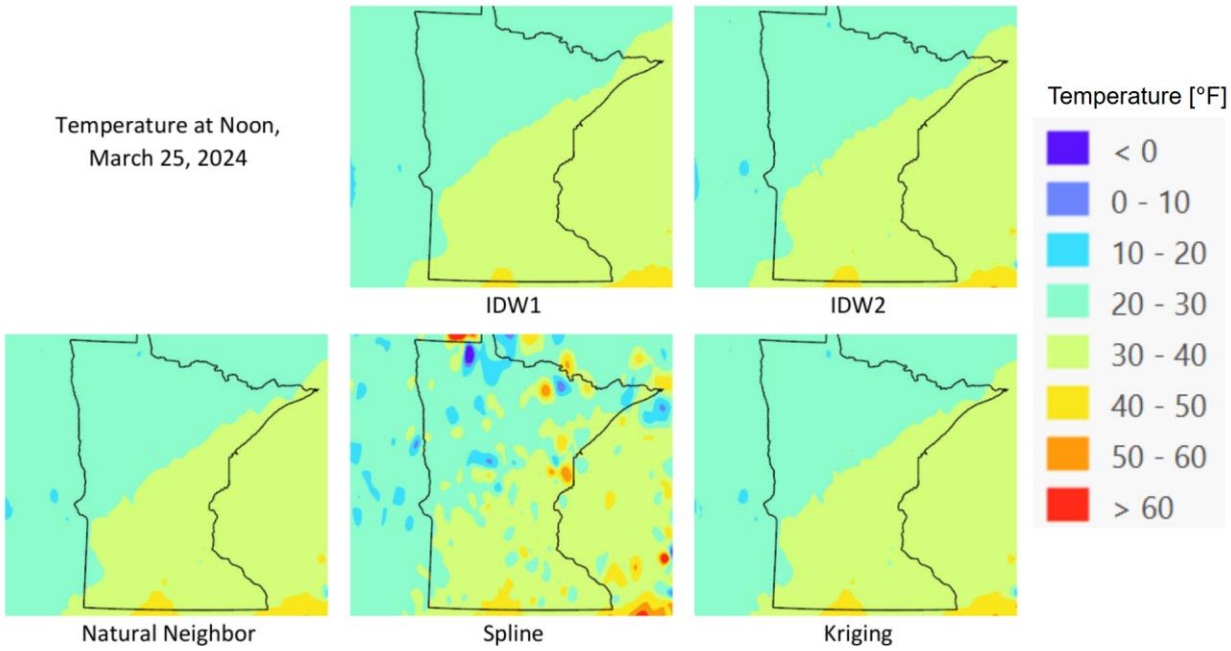


Figure 6: Comparison of methods for temperature.

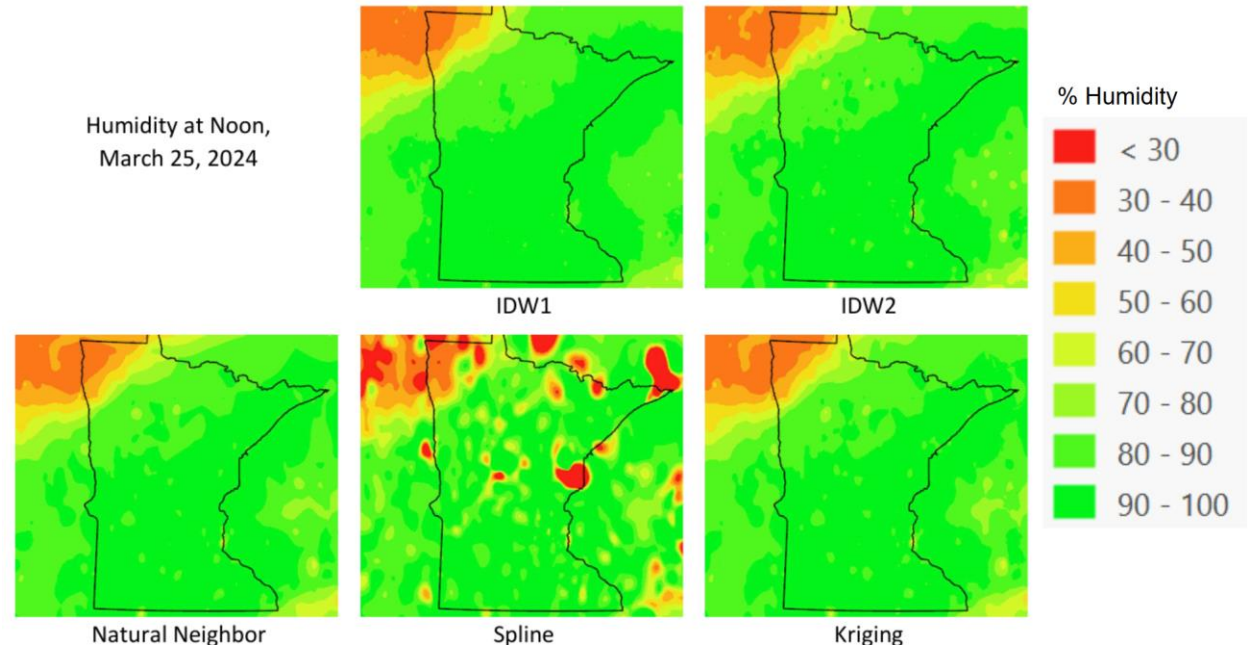


Figure 7: Comparison of methods for humidity.

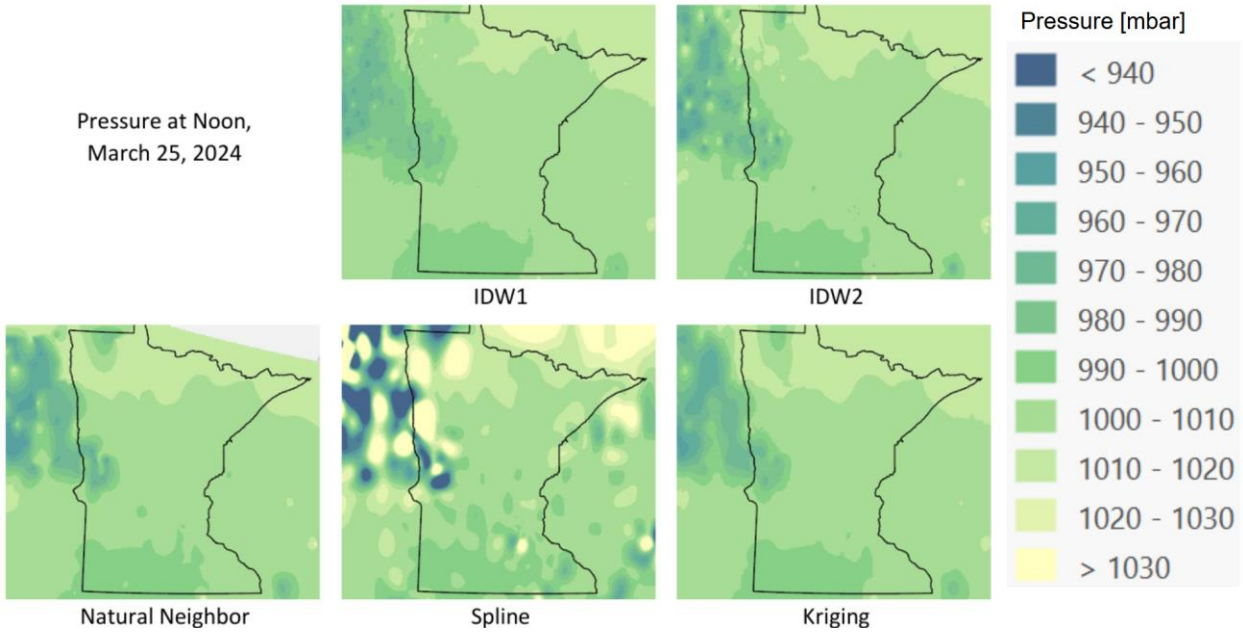


Figure 8: Comparison of methods for pressure.

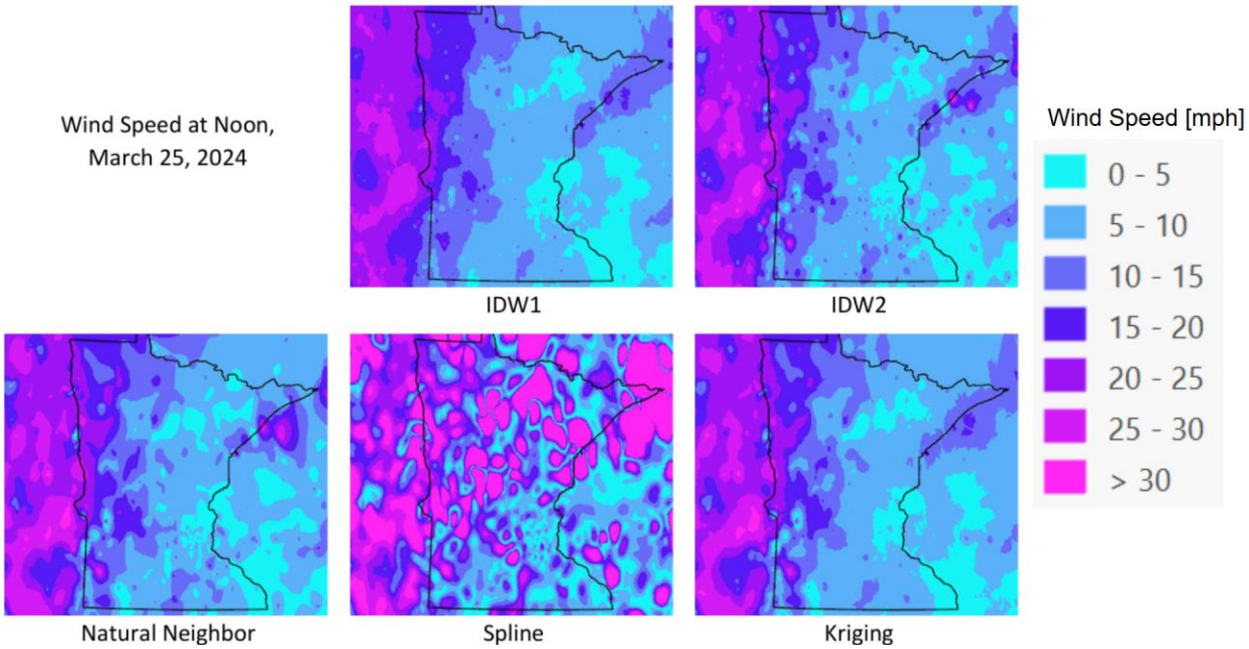


Figure 9: Comparison of methods for wind speed.

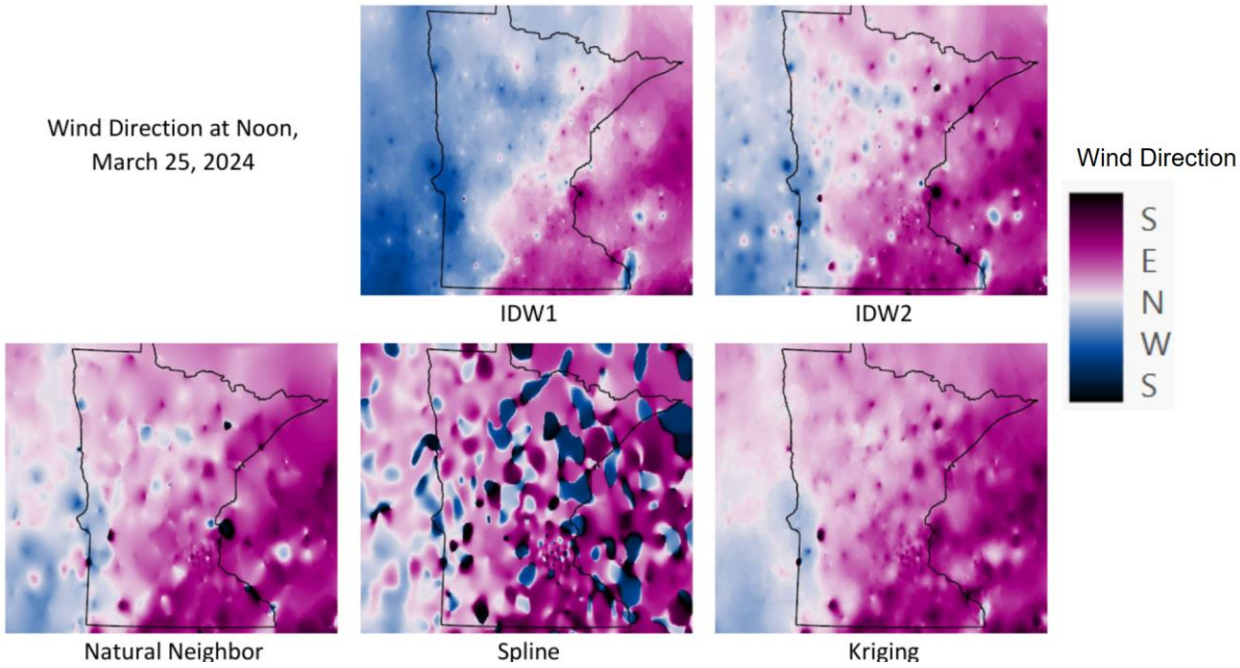


Figure 10: Comparison of methods for wind direction.

Results Verification

Table 3: RMSE for each interpolation method and weather element.

	Temperature (°F)	Humidity (%)	Pressure (mbar)	Wind (N-S) (mph)	Wind (E-W) (mph)
1st Order IDW	3.54	9.22	**7.99**	5.50	**6.53**
2nd Order IDW	2.65	7.93	8.48	5.42	6.71
Natural Neighbor	2.74	**6.90**	8.81	**5.37**	6.99
Spline	**4.70**	**23.39**	**15.56**	**9.32**	**9.87**
Kriging	**2.50**	7.35	8.36	5.43	6.87

A leave-one-out cross validation scheme was used to analyze the accuracy of the results. For each temperature, humidity, pressure, north-south wind, and east-west wind, 50 points were chosen at random for the 12:00 hour to be left out. The value at these points were compared to the interpolated value and RMSE was calculated for each method.

Spline interpolation performed the worst for every weather element. The other methods performed similarly. The best method was not consistent across weather elements. Interpolation of the east-west component of wind consistently performed worse than the north-south component.

Table 4: RMSE for each interpolation method and weather element, only considering NOAA data.

	Temperature (°F)	Humidity (%)	Pressure (mbar)	Wind (N-S) (mph)	Wind (E-W) (mph)
1st Order IDW	2.99	7.33	1.63	**4.91**	**5.72**
2nd Order IDW	2.73	6.64	1.57	5.94	5.59
Natural Neighbor	**2.23**	**6.05**	**1.33**	5.48	5.83
Spline	**3.44**	**13.73**	**3.34**	**7.53**	**8.32**
Kriging	2.29	9.92	1.41	5.38	5.57

Leave-one-out cross validation was also conducted only considering NOAA data. RMSE was consistently better for this case, and is especially better for pressure, where the effects of using multiple data sources was most apparent. Natural neighbor performed the best for all one-dimensional quantities while first-order IDW performed the best for wind. Interpolation of the east-west wind component was still consistently worse than the north-south component.

Discussion and Conclusion

For the three one-dimensional elements (temperature, humidity, and pressure), the results of first-order IDW, second-order IDW, natural neighbor, and Kriging look very similar. Spline interpolation shows the same general patterns, but extreme values much more commonly appear. The effects of using two different data sources are most apparent when looking at pressure. While the interpolated surfaces are mostly smooth, they become much more spotted in North Dakota, where there is data from both sources.

Wind speed and wind direction show more differences between first-order IDW, second-order IDW, natural neighbor, and Kriging. This is most apparent for wind direction, where first-order IDW appears quite different from the rest. Spline interpolation appears very noisy for both wind speed and direction, and the general patterns that can be seen in the one-dimensional quantities are much less apparent. Spline interpolation again commonly shows extreme values.

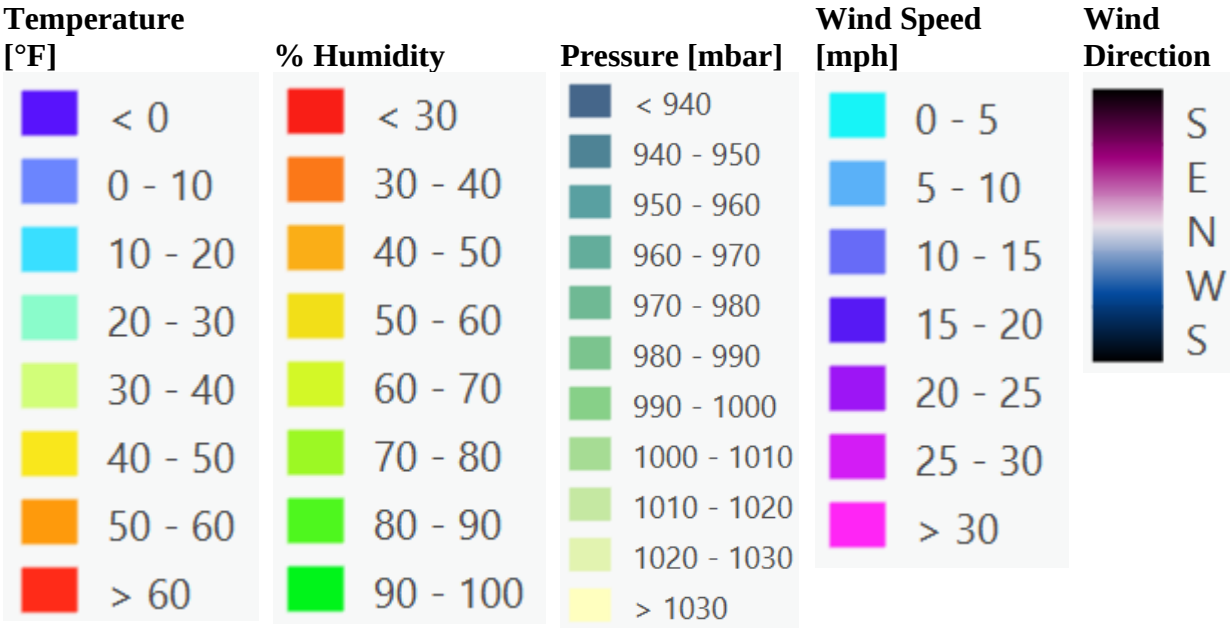
Spline interpolation should not be used for interpolating weather data. All other methods examined in this project produce comparable results. It is unclear which method would be the best choice overall.

Self-score

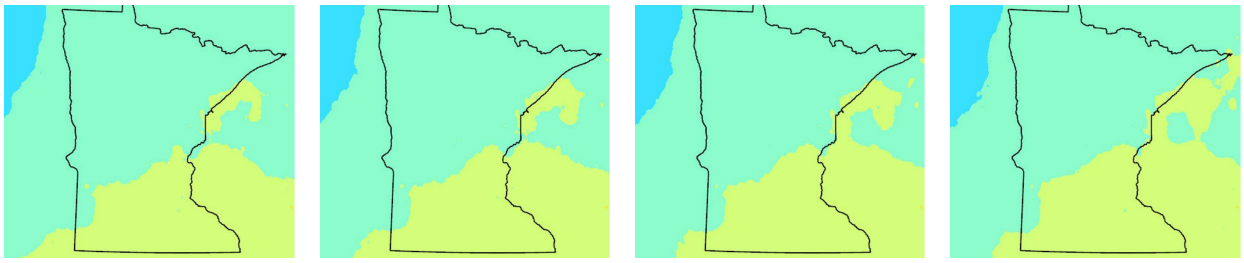
Category	**Description**	**Points Possible**	**Score**
Structural Elements	All elements of a lab report are included **(2 points each)**: Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score	28	**26**
Clarity of Content	Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory-level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**.	24	**24**
Reproducibility	Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified.	28	**26**
Verification	Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**.	20	**19**
		100	**95**

Appendix

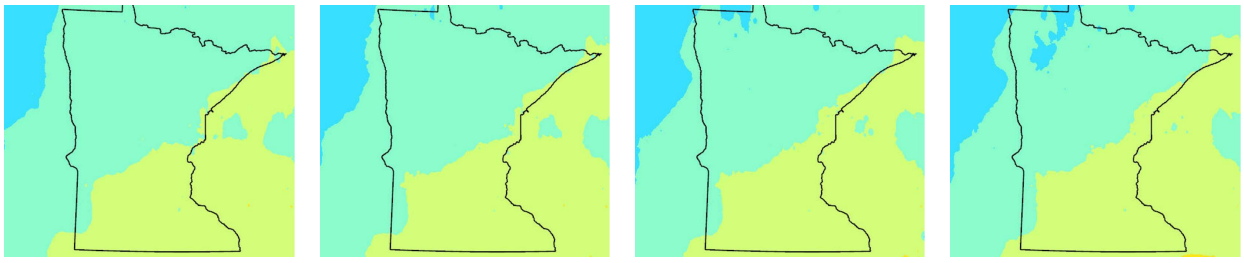
Legends for Small Multiples:



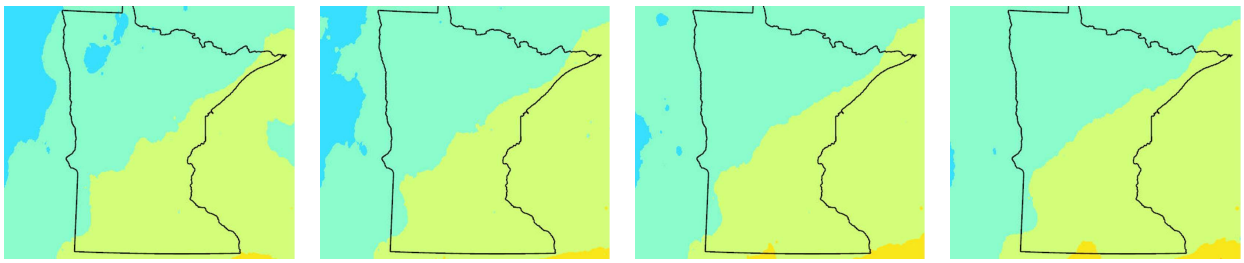
Temperature - First-order IDW



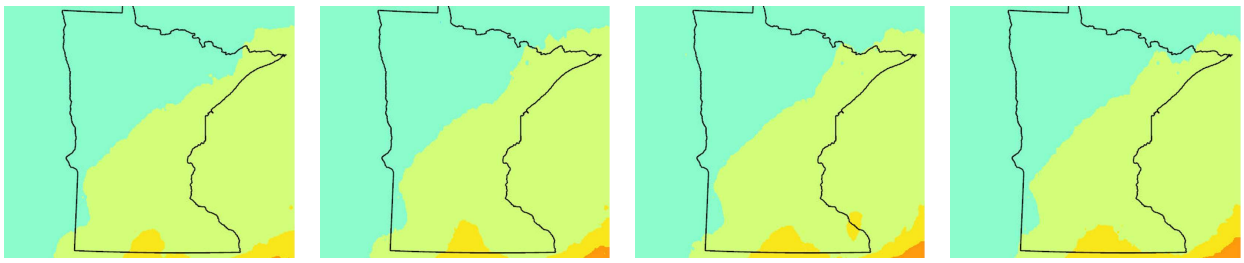
01:00 02:00 03:00 04:00



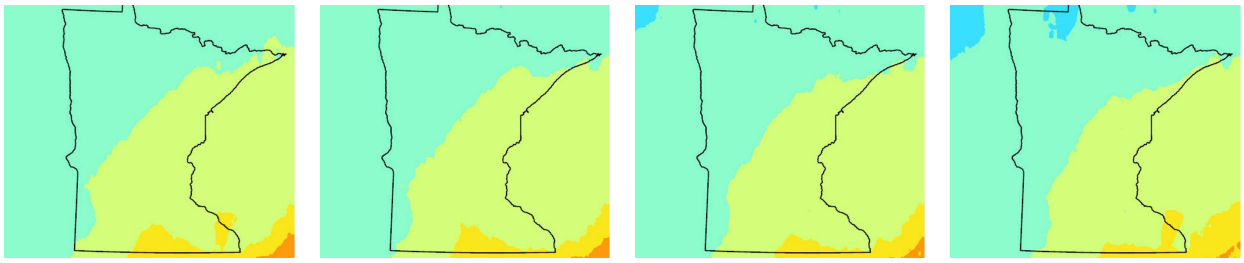
05:00 06:00 07:00 08:00



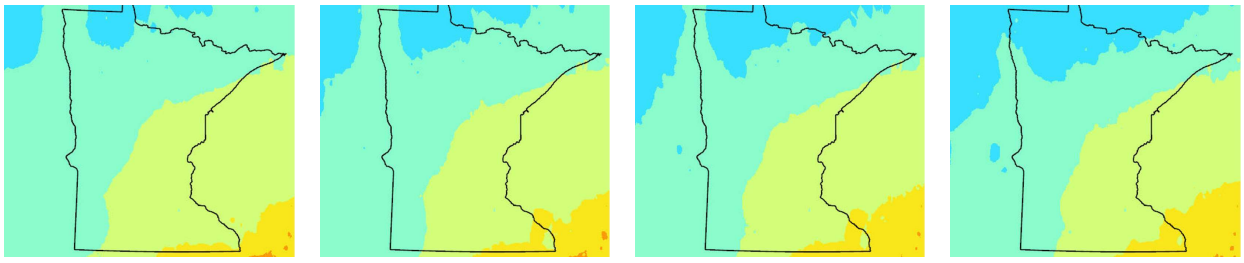
09:00 10:00 11:00 12:00



13:00 14:00 15:00 16:00

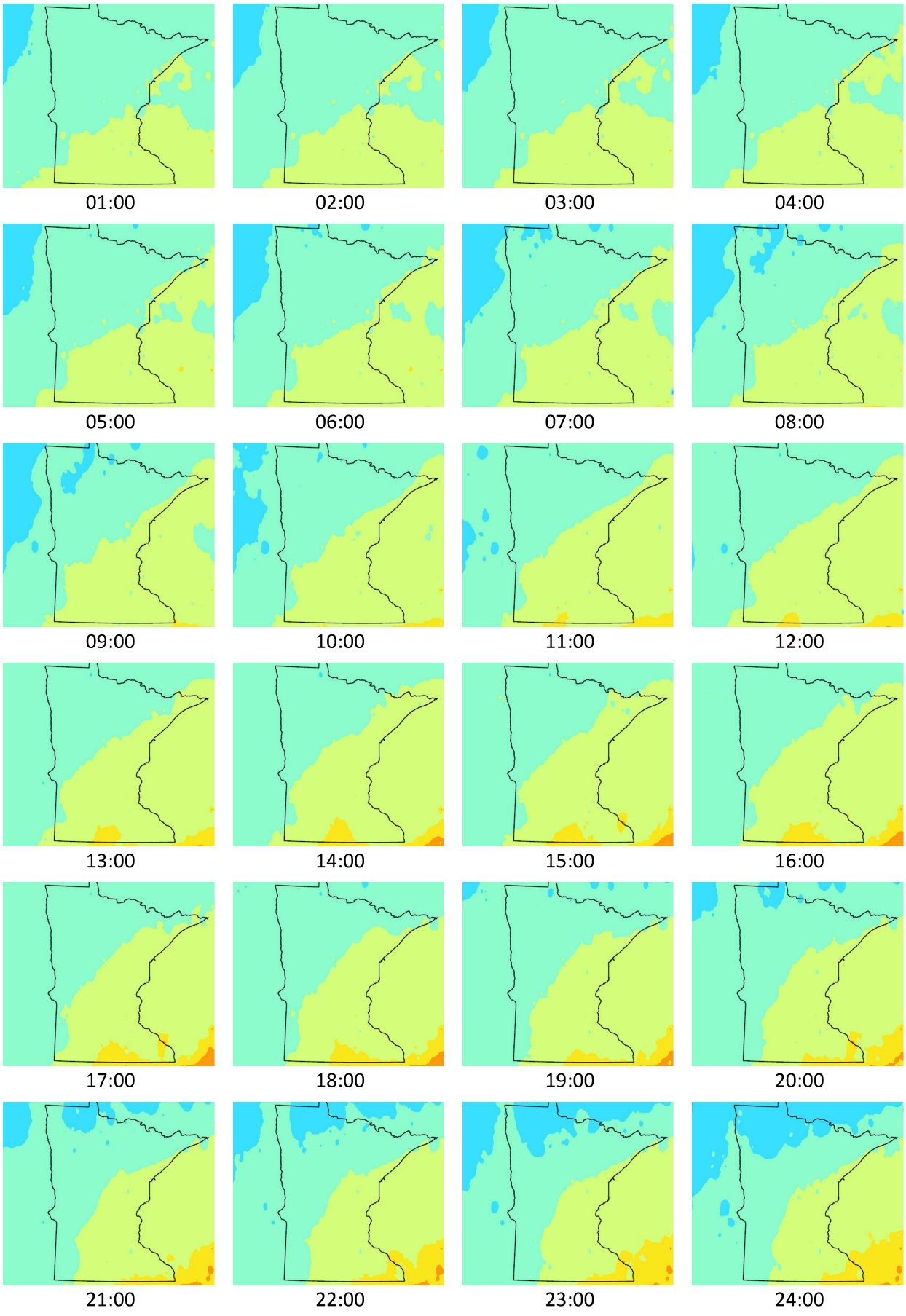


17:00 18:00 19:00 20:00

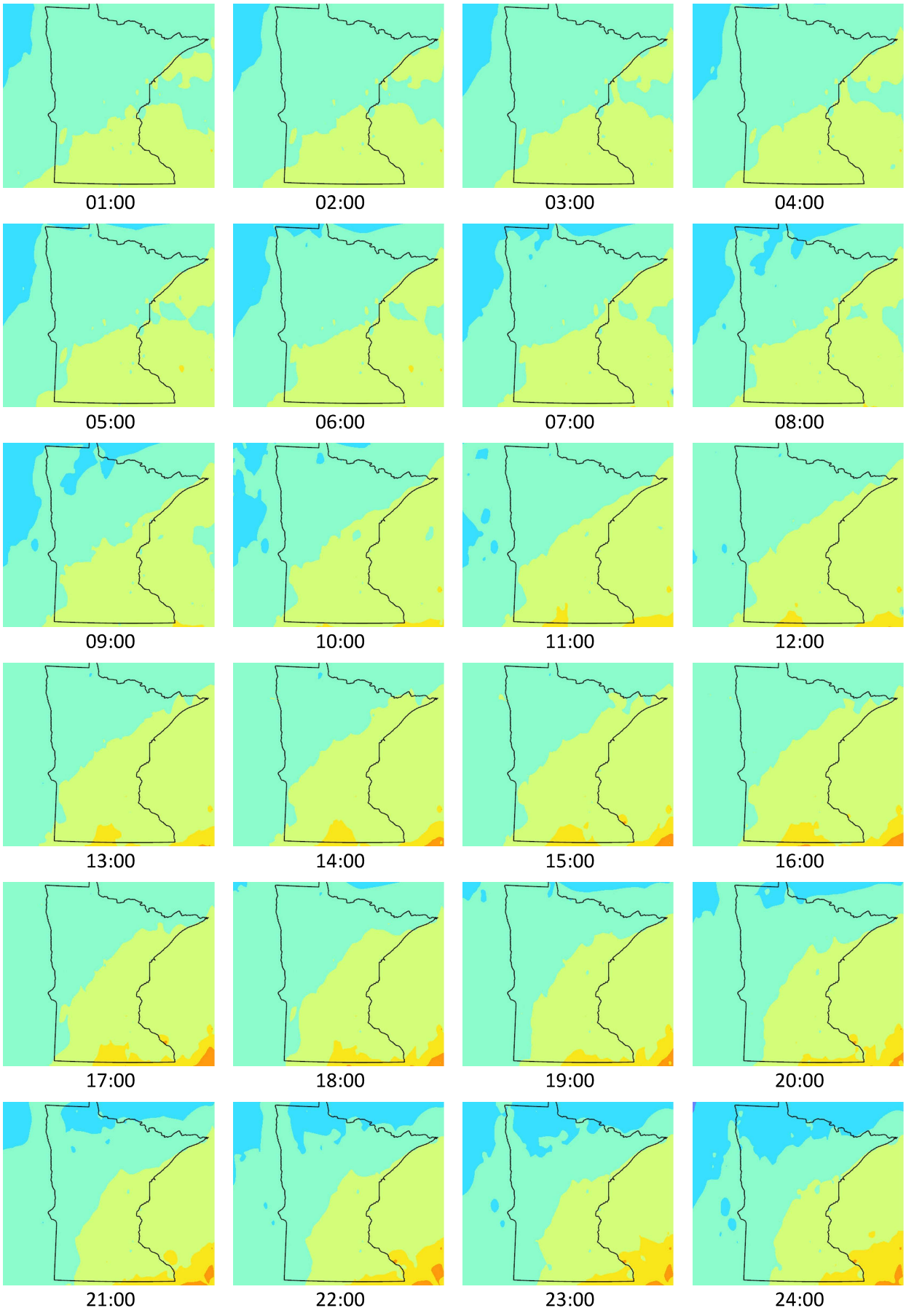


21:00 22:00 23:00 24:00

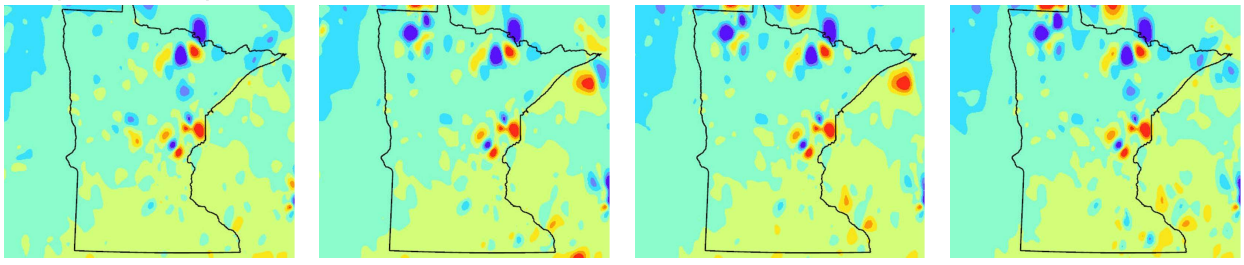
Temperature - Second-order IDW



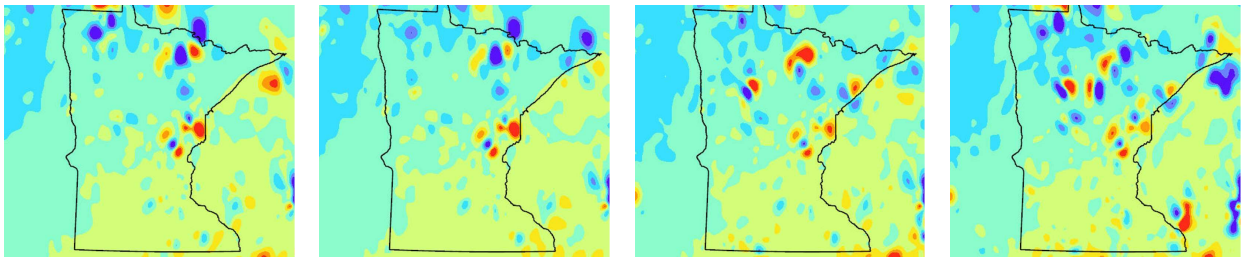
Temperature - Natural Neighbor



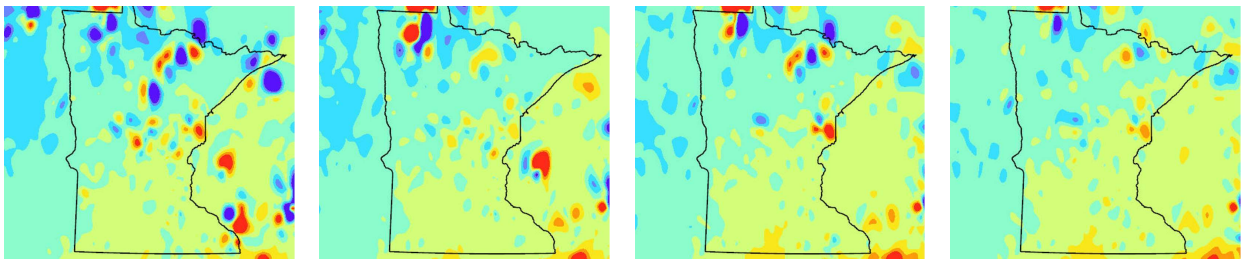
Temperature - Spline



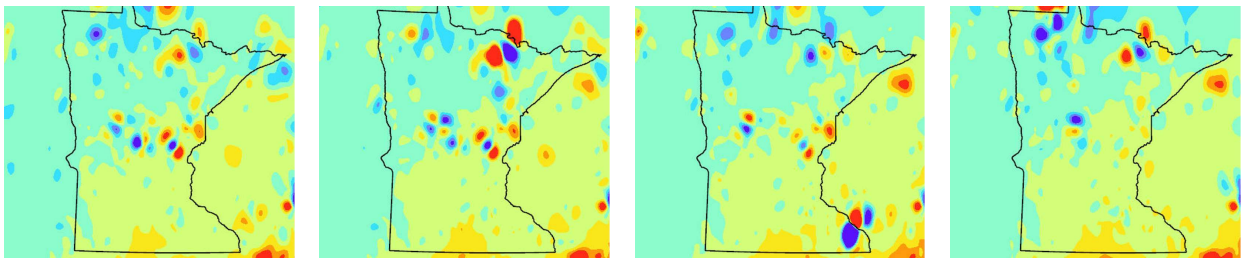
01:00 02:00 03:00 04:00



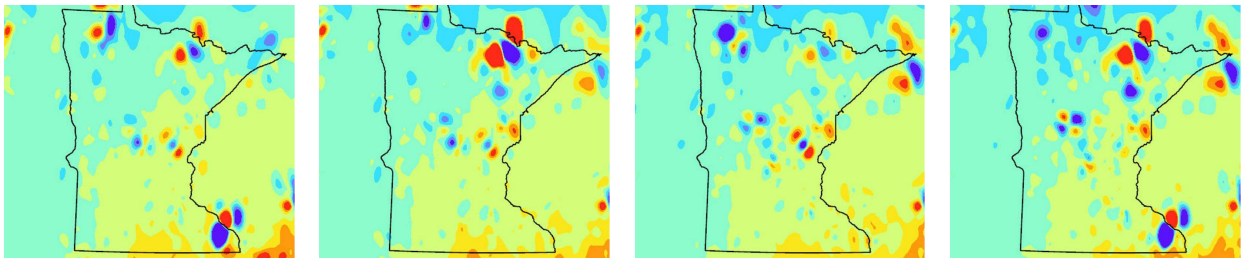
05:00 06:00 07:00 08:00



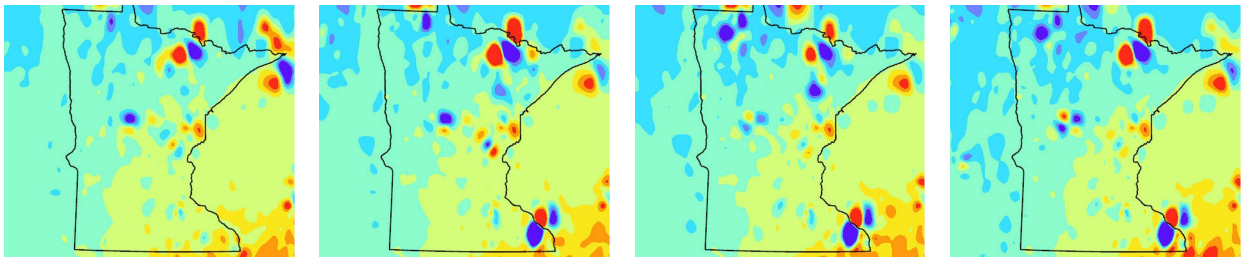
09:00 10:00 11:00 12:00



13:00 14:00 15:00 16:00

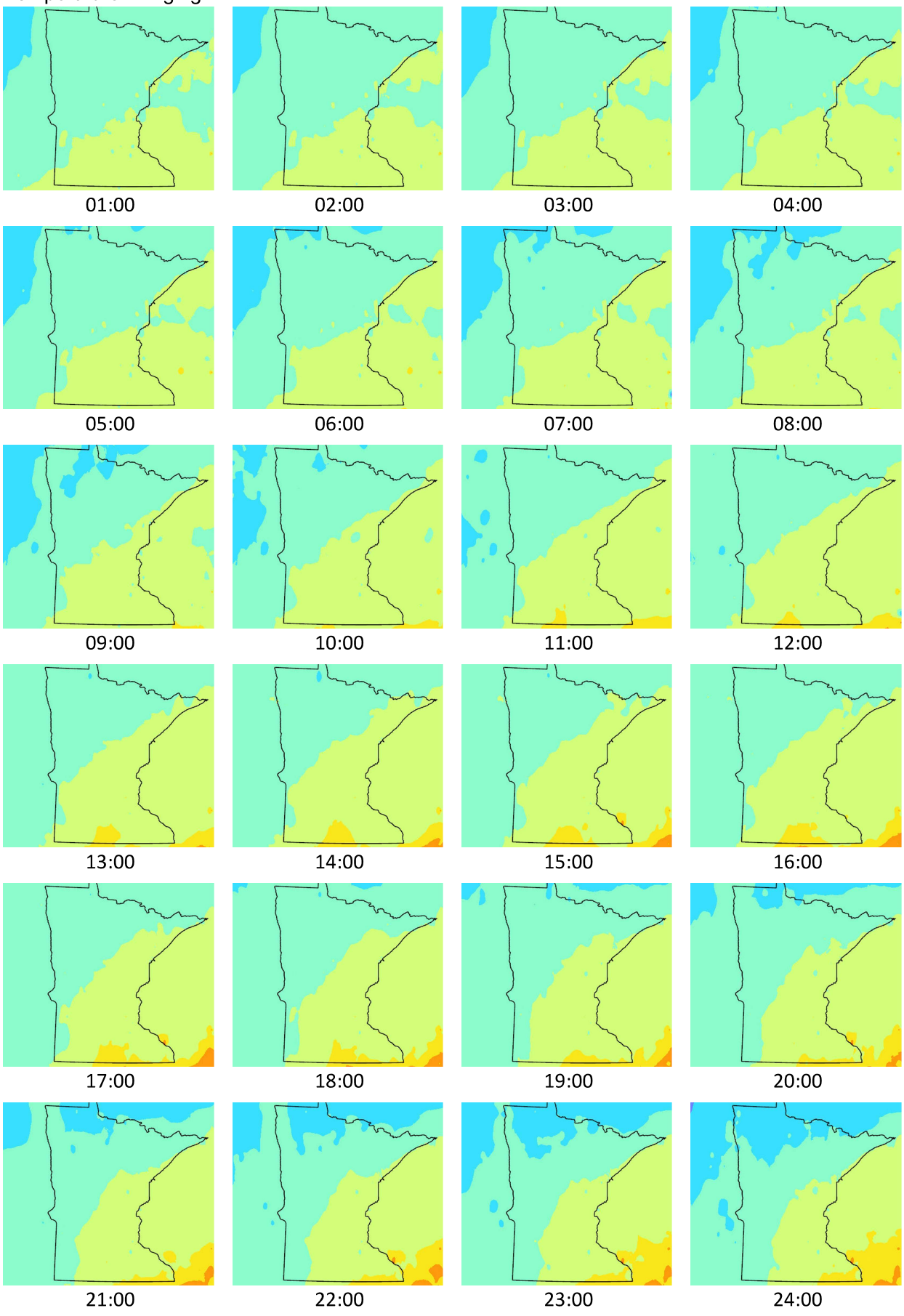


17:00 18:00 19:00 20:00

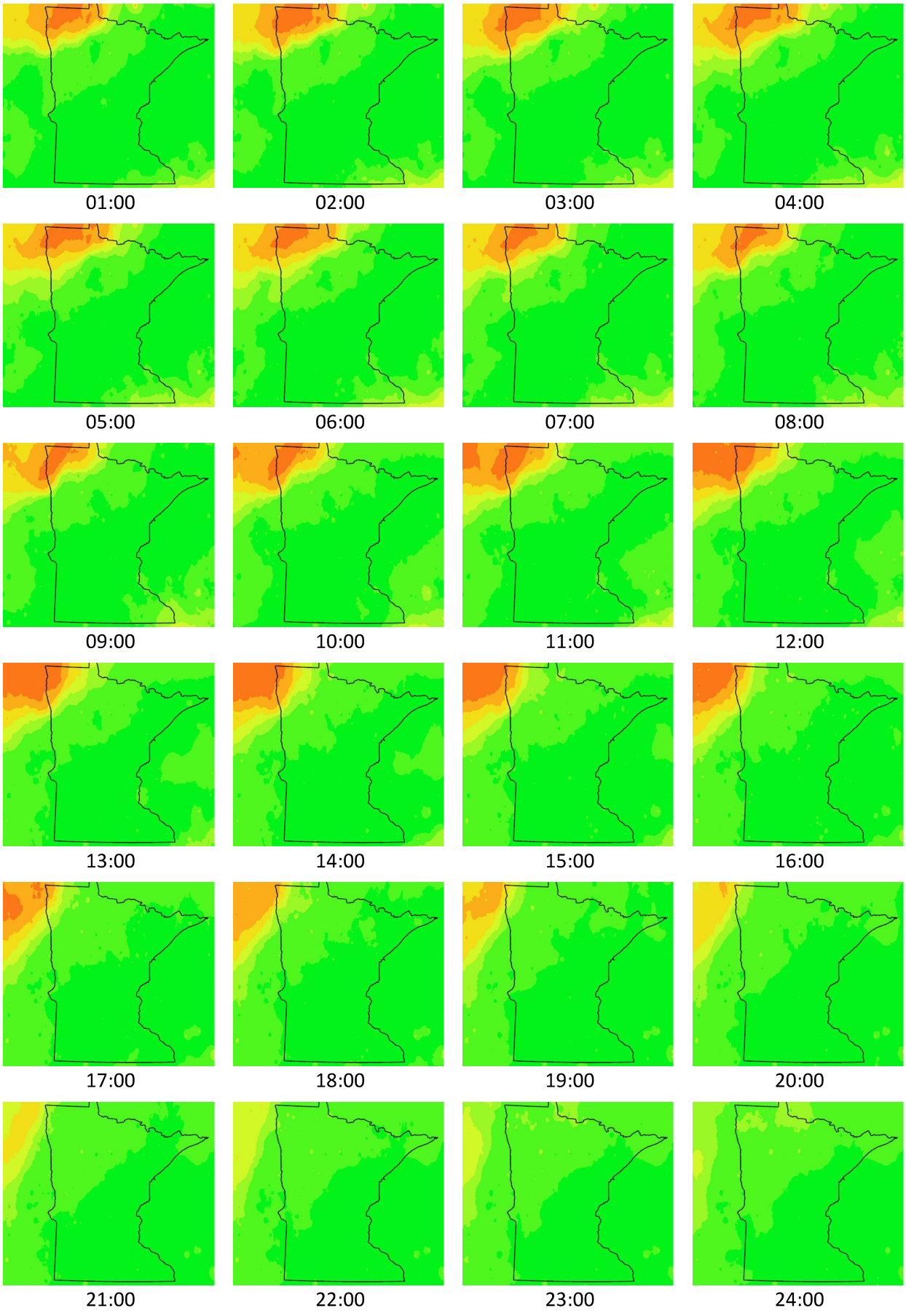


21:00 22:00 23:00 24:00

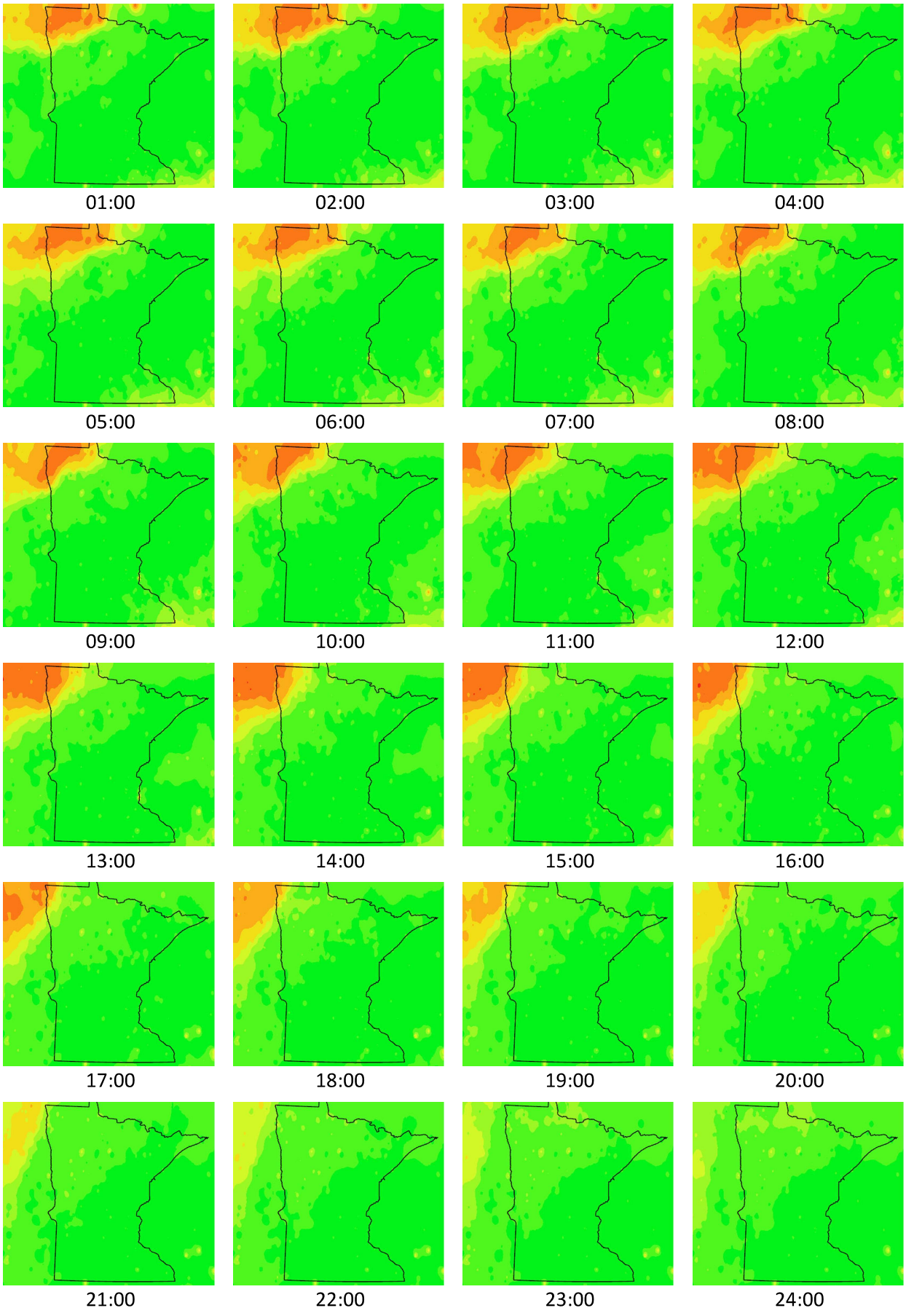
Temperature - Kriging



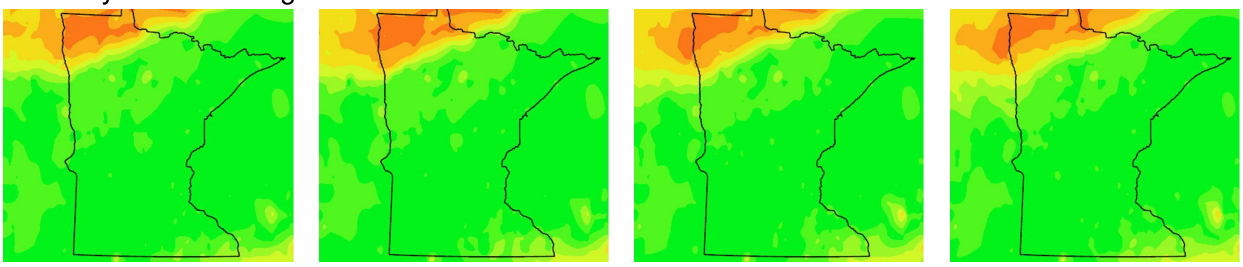
Humidity - First-order IDW



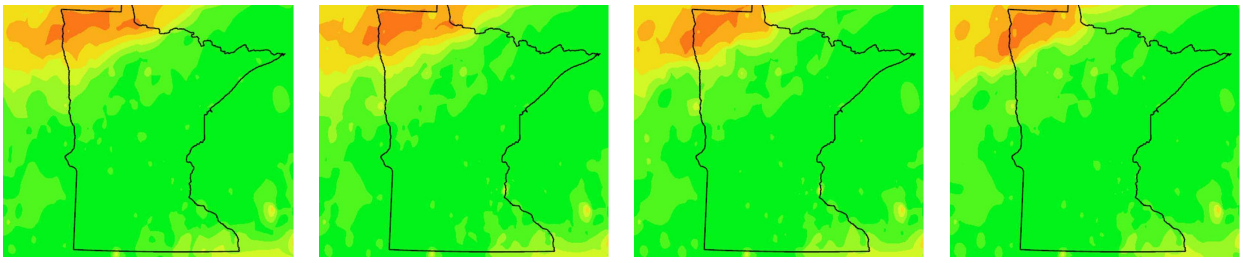
Humidity - Second-order IDW



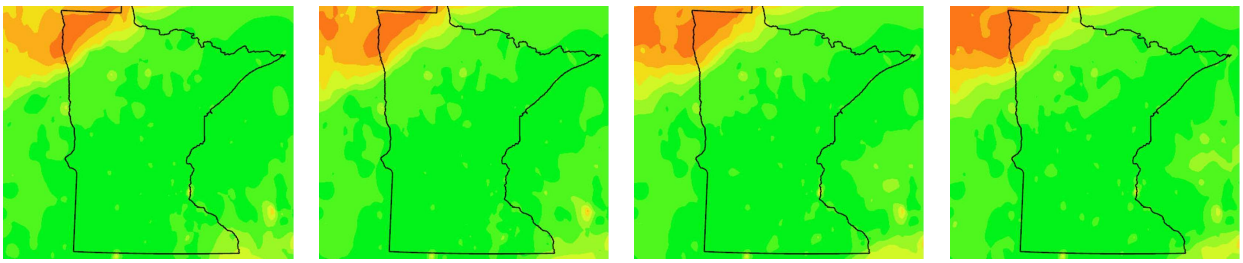
Humidity - Natural Neighbor



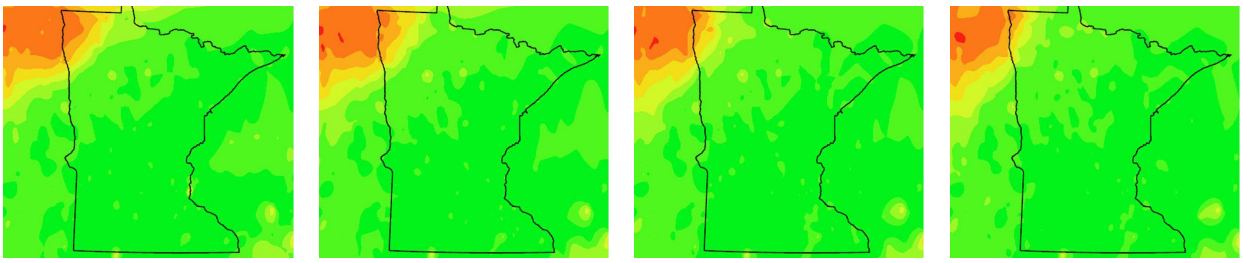
01:00 02:00 03:00 04:00



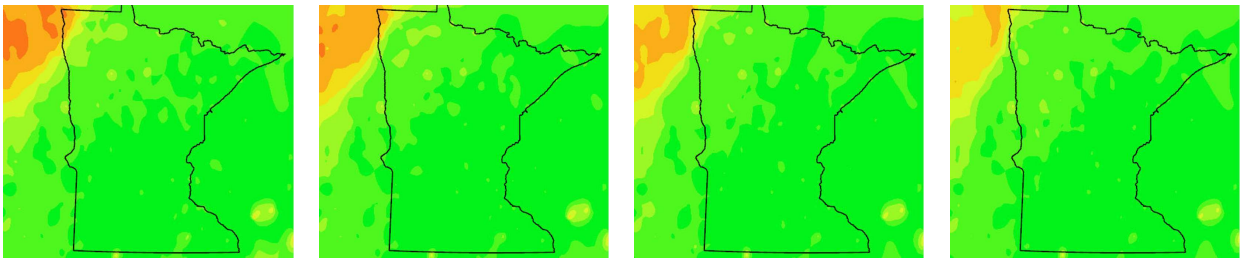
05:00 06:00 07:00 08:00



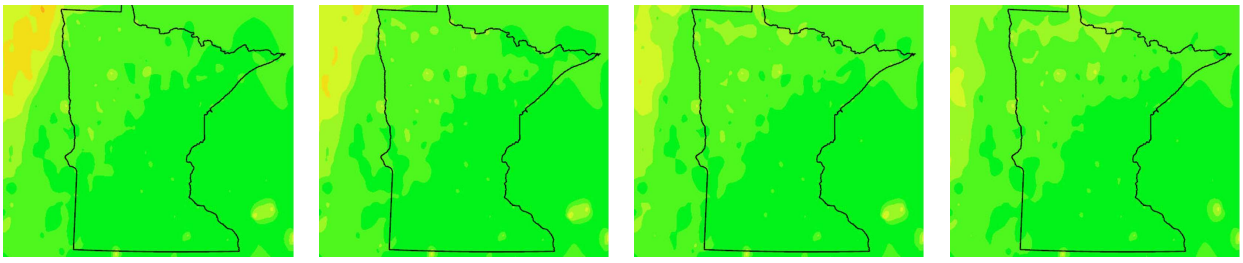
09:00 10:00 11:00 12:00



13:00 14:00 15:00 16:00

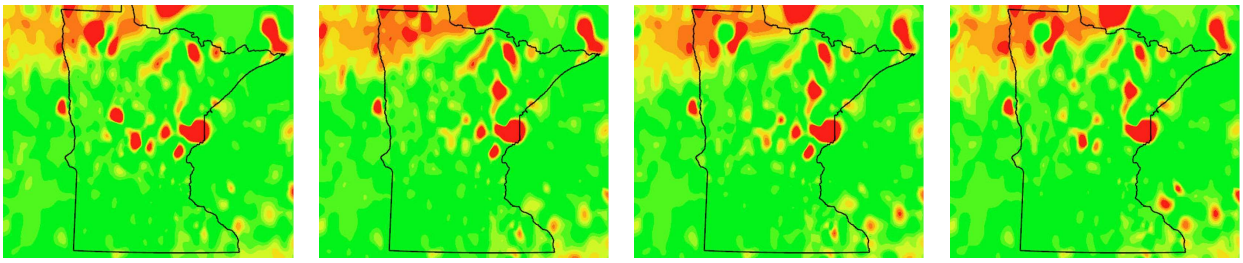


17:00 18:00 19:00 20:00

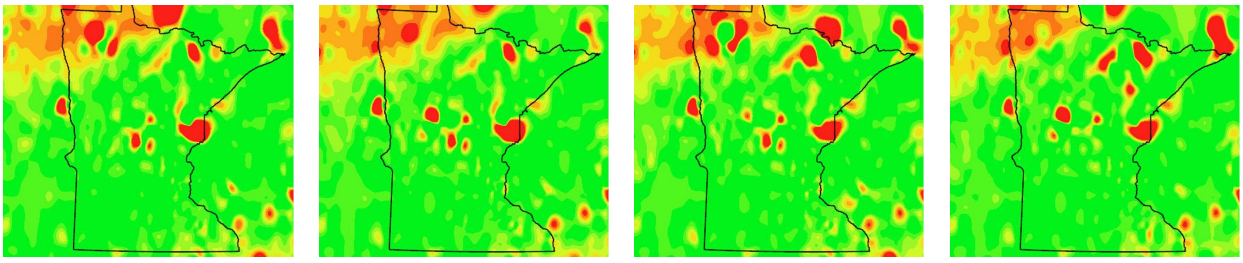


21:00 22:00 23:00 24:00

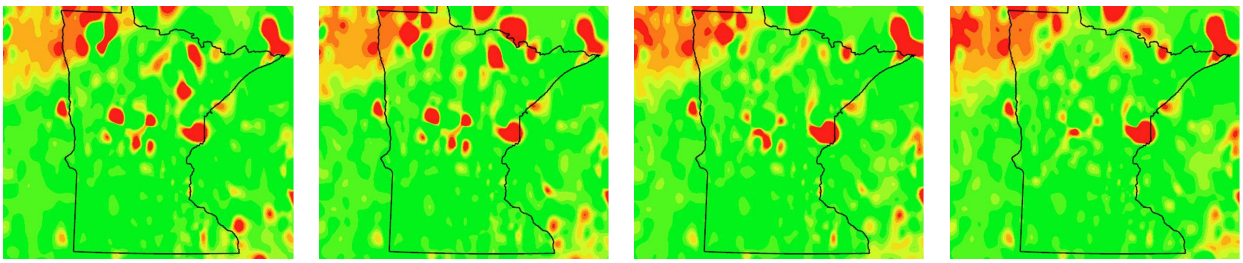
Humidity - Spline



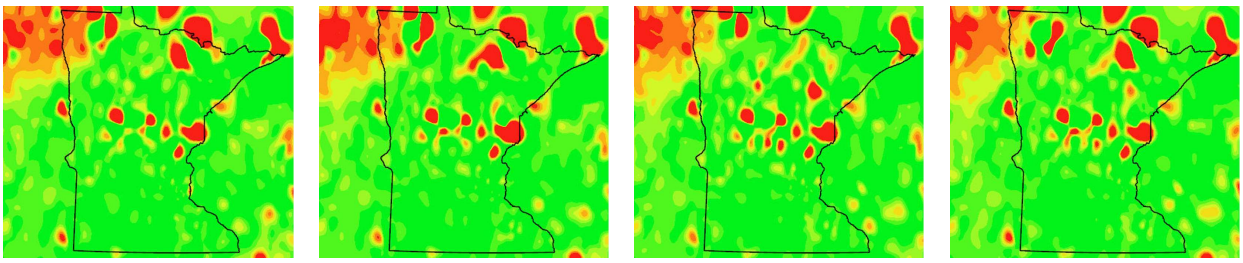
01:00 02:00 03:00 04:00



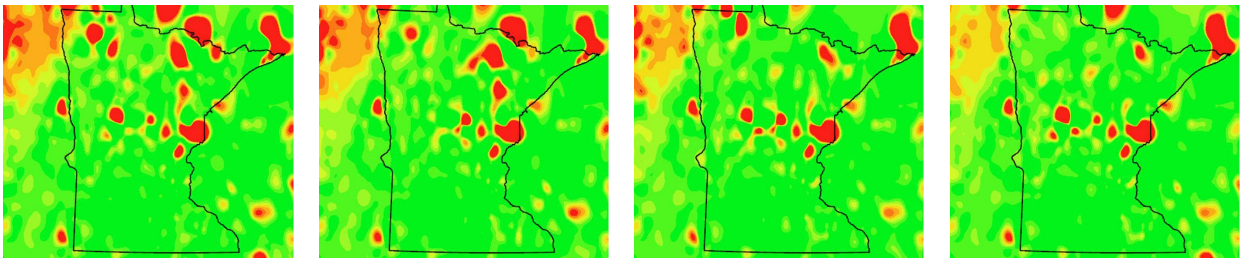
05:00 06:00 07:00 08:00



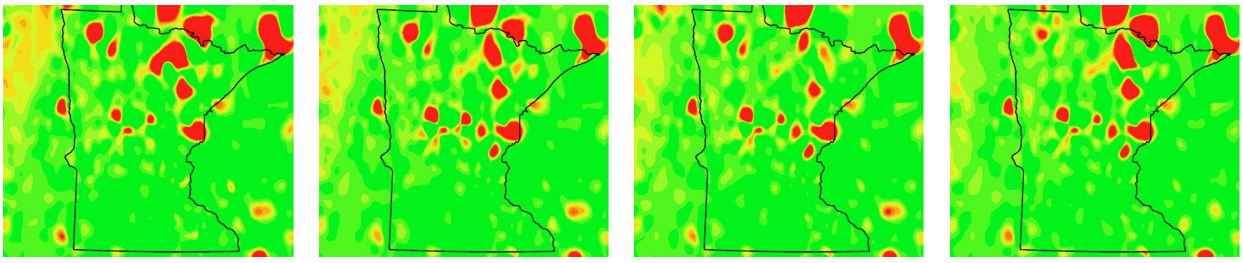
09:00 10:00 11:00 12:00



13:00 14:00 15:00 16:00

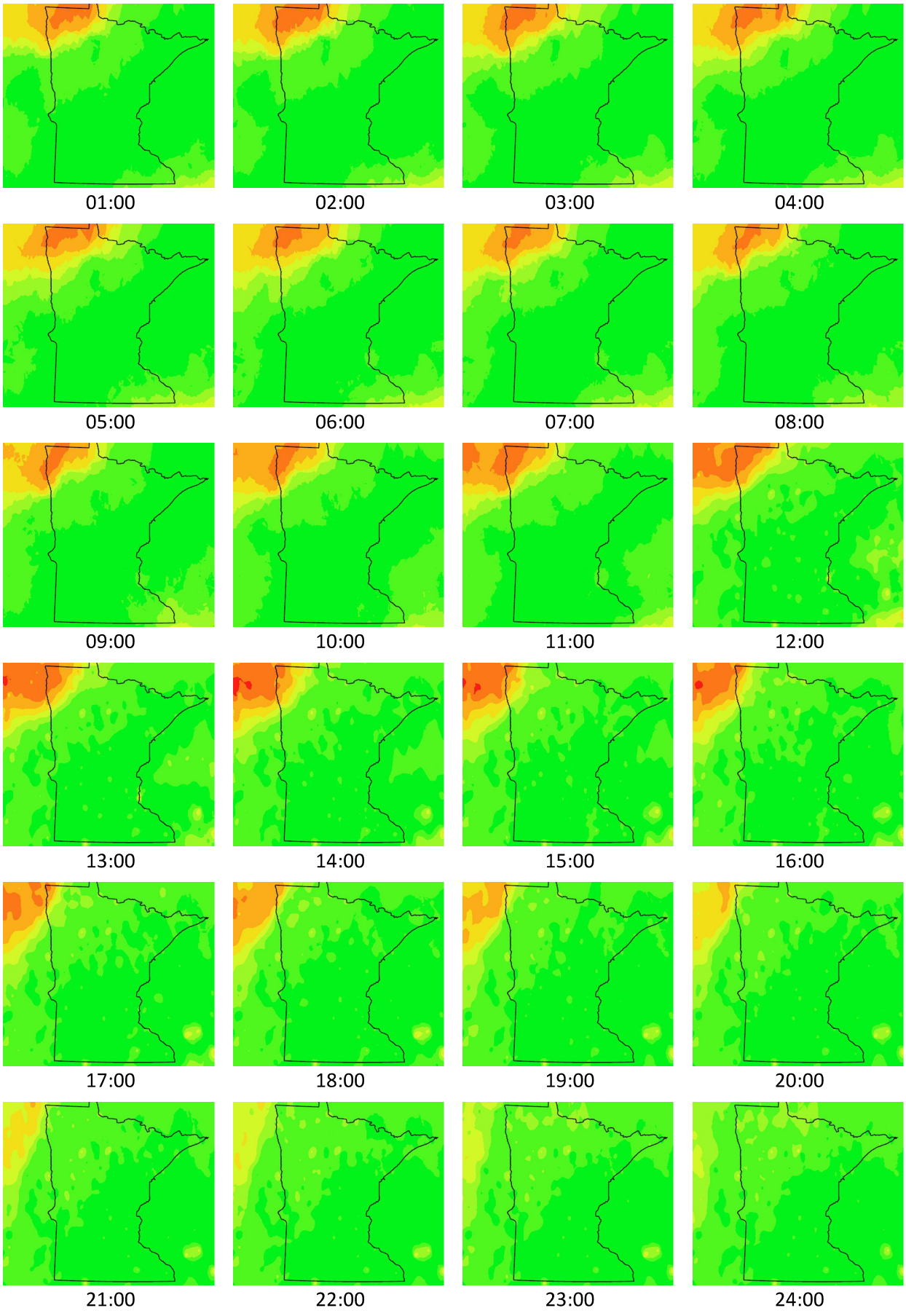


17:00 18:00 19:00 20:00

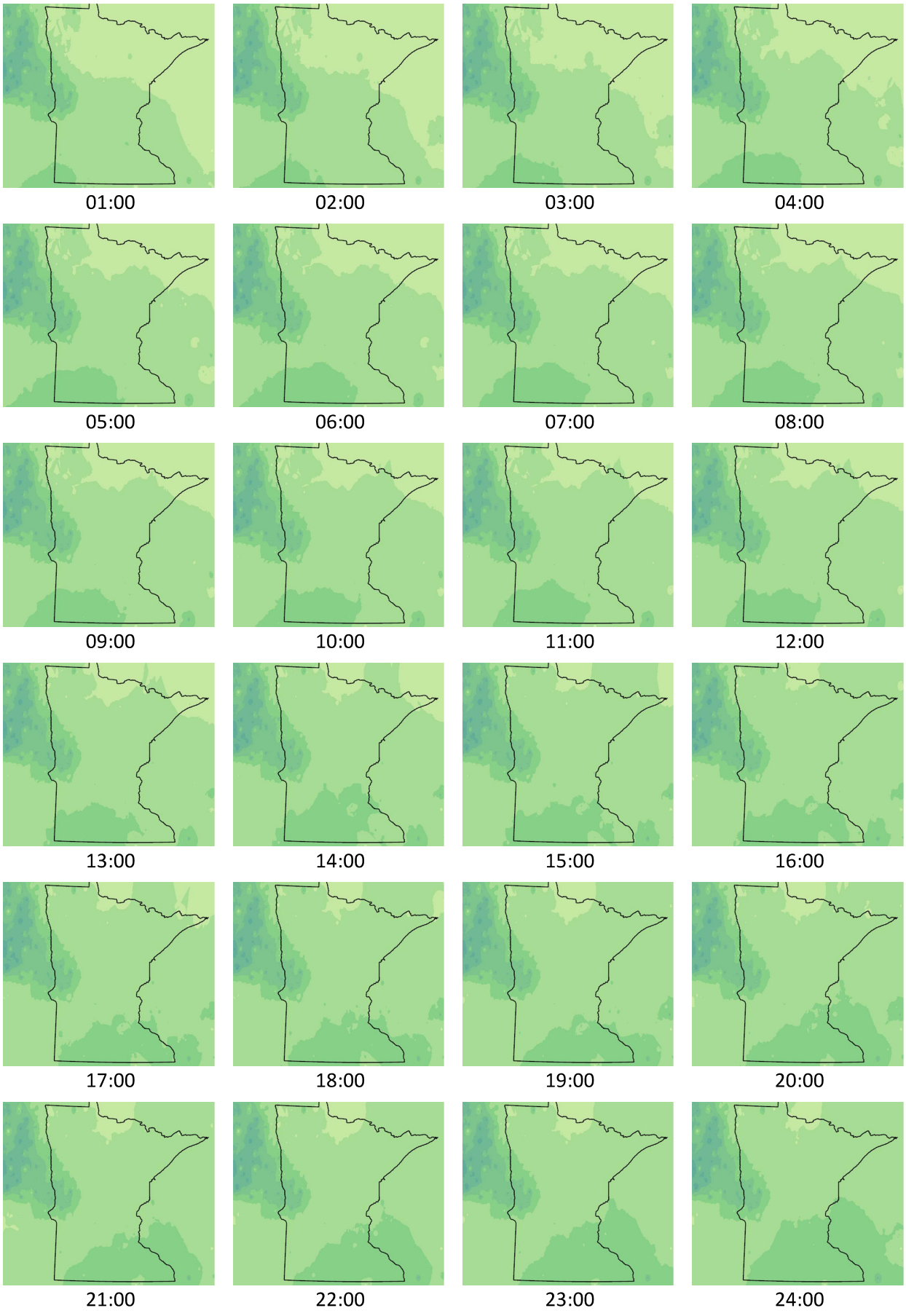


21:00 22:00 23:00 24:00

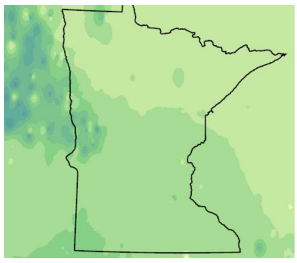
Humidity - Kriging

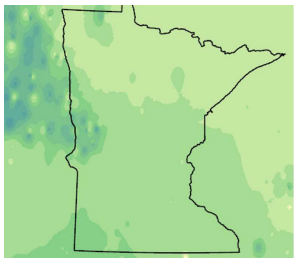


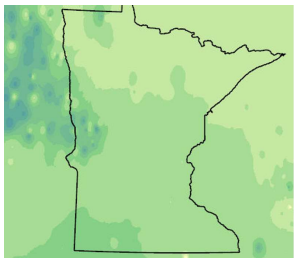
Pressure - First-order IDW

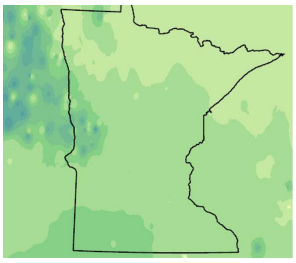


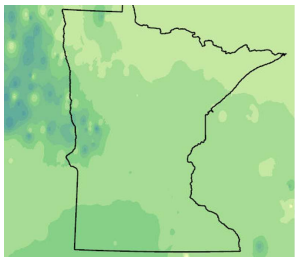
Pressure - Second-order IDW

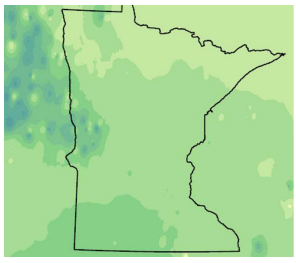

01:00

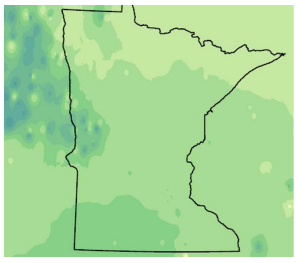

02:00

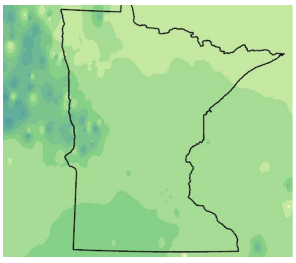

03:00

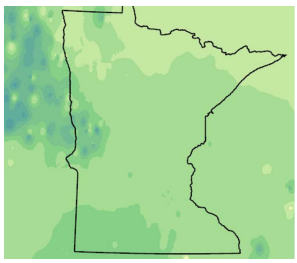

04:00

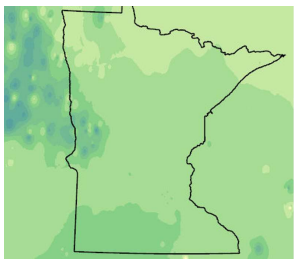

05:00

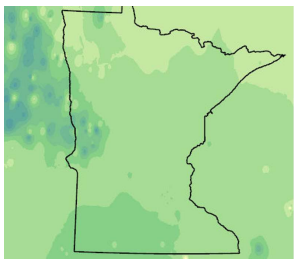

06:00

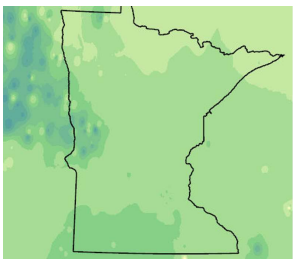

07:00

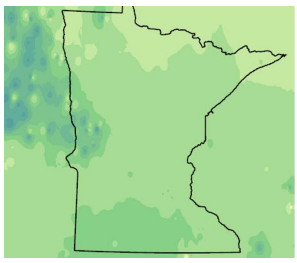

08:00

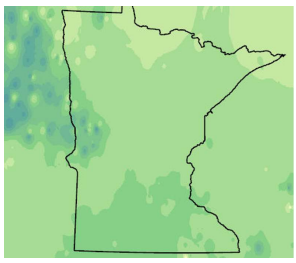

09:00

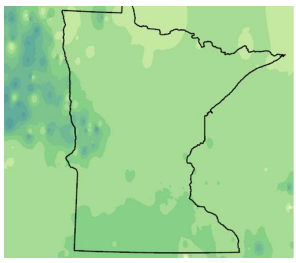

10:00

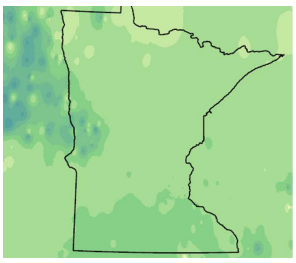

11:00

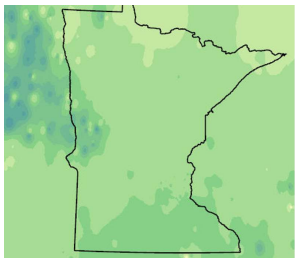

12:00

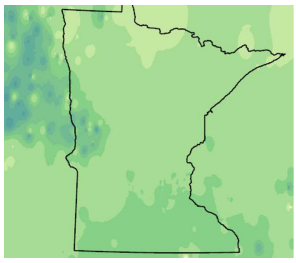

13:00

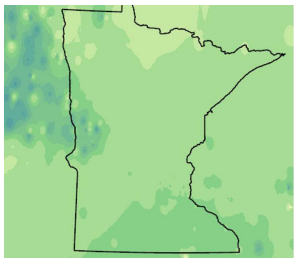

14:00

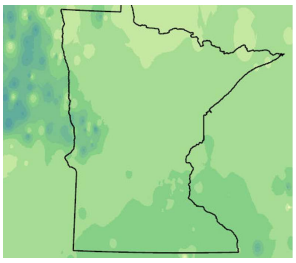

15:00

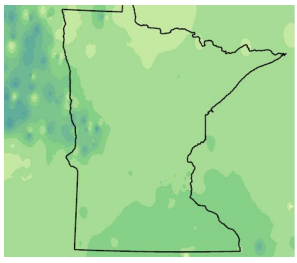

16:00

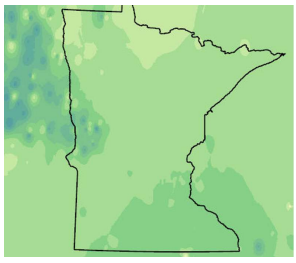

17:00

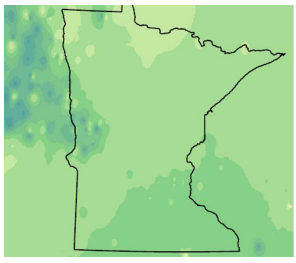

18:00

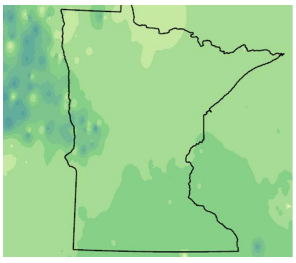

19:00


20:00

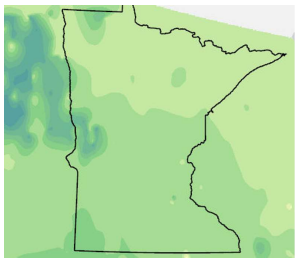

21:00

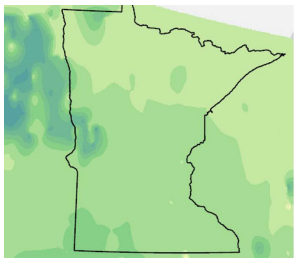

22:00

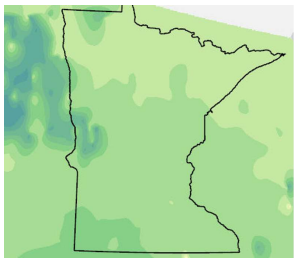

23:00

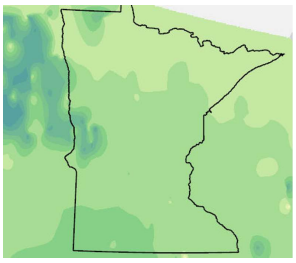

24:00

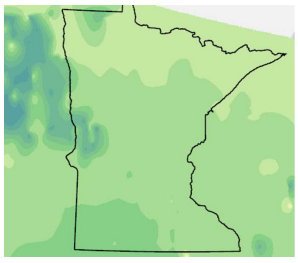
Pressure - Natural Neighbor

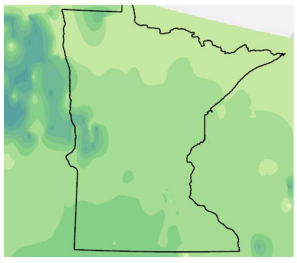

01:00

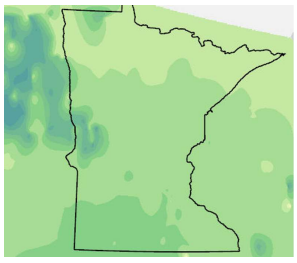

02:00

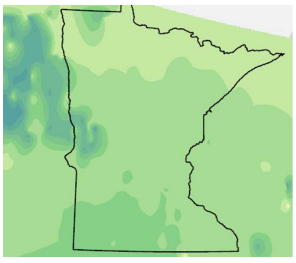

03:00

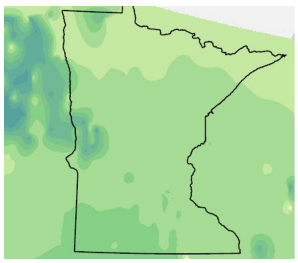

04:00

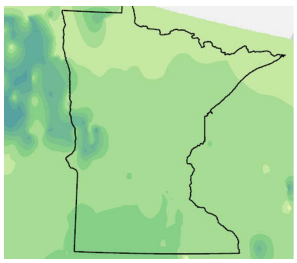

05:00

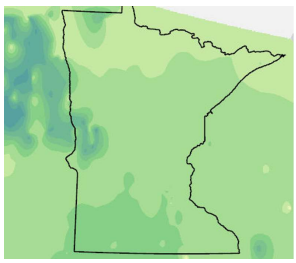

06:00

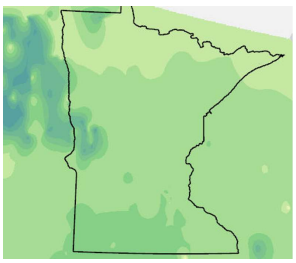

07:00

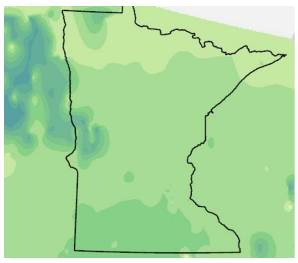

08:00

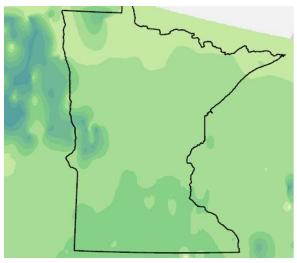

09:00

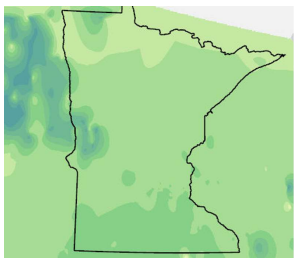

10:00

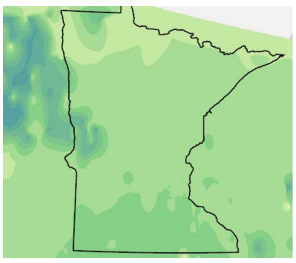

11:00

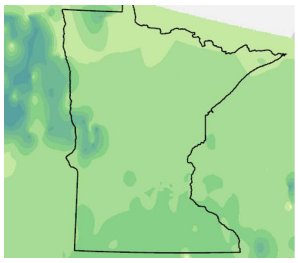

12:00

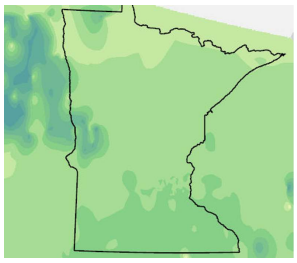

13:00

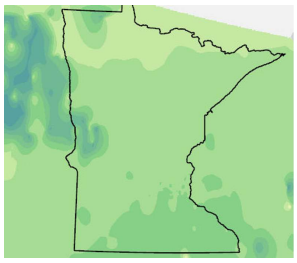

14:00

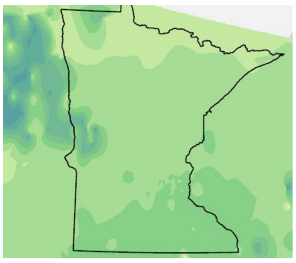

15:00

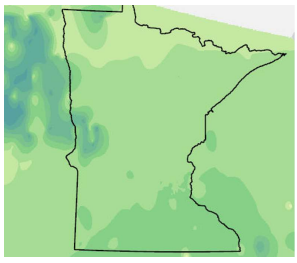

16:00

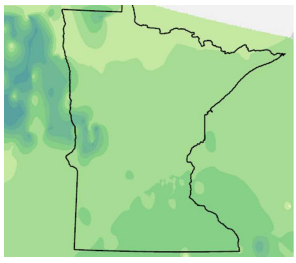

17:00

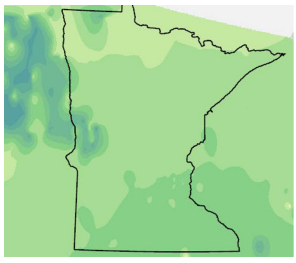

18:00

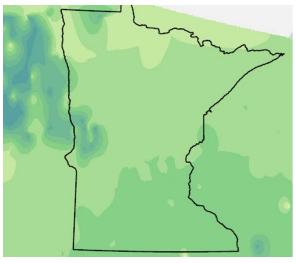

19:00


20:00

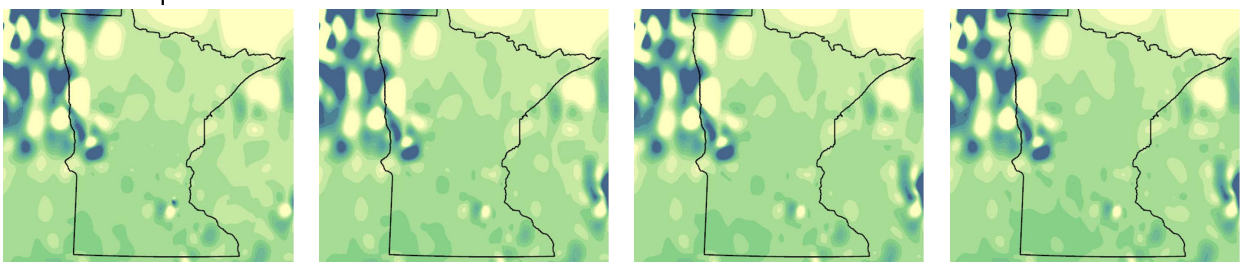

21:00


22:00

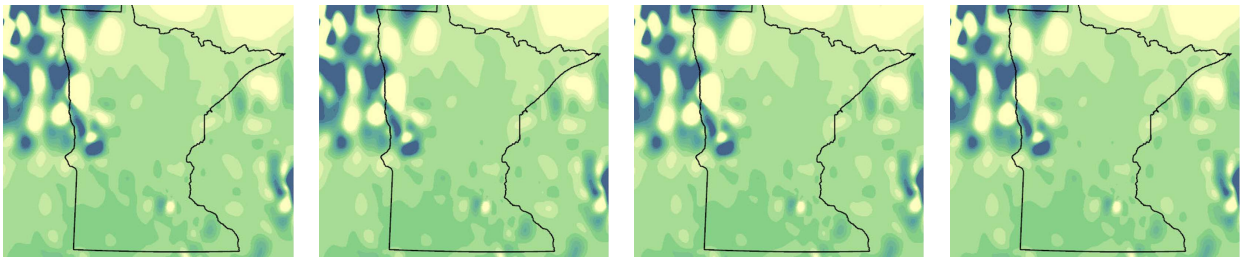

23:00


24:00

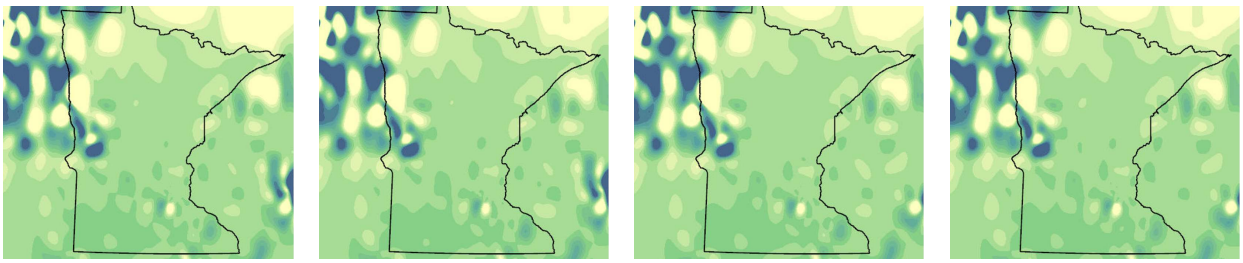
Pressure - Spline



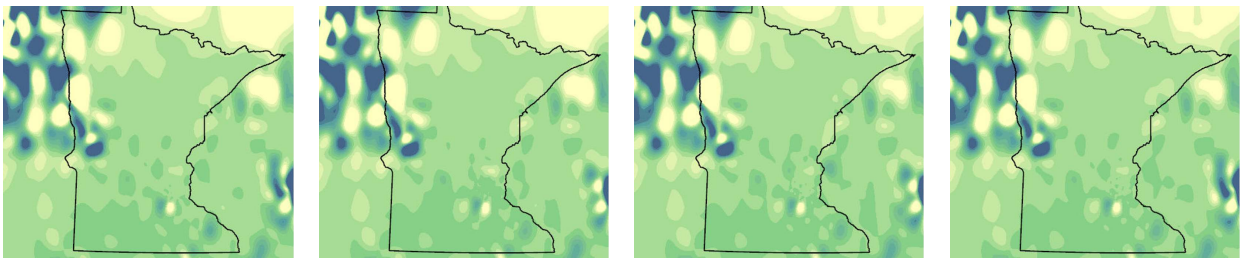
01:00 02:00 03:00 04:00



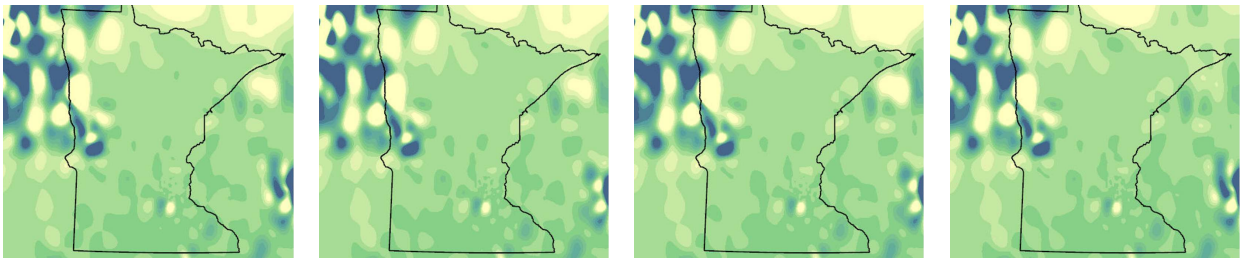
05:00 06:00 07:00 08:00



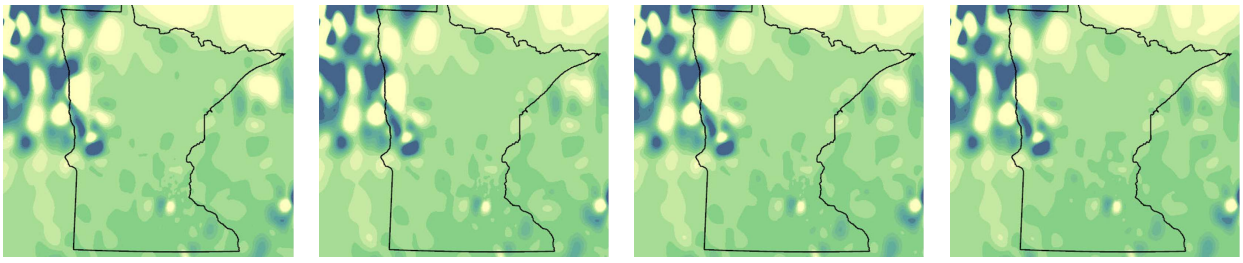
09:00 10:00 11:00 12:00



13:00 14:00 15:00 16:00

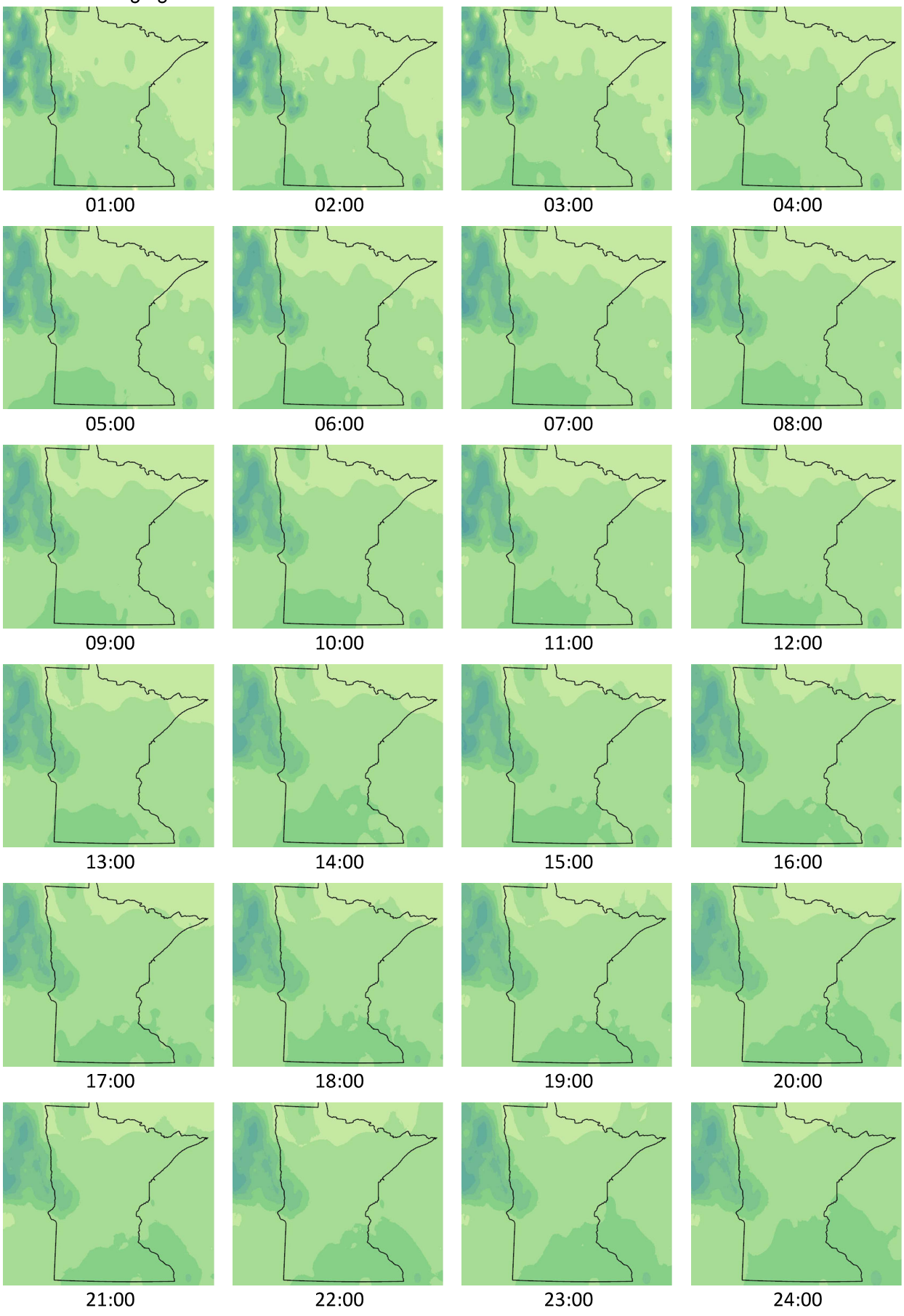


17:00 18:00 19:00 20:00

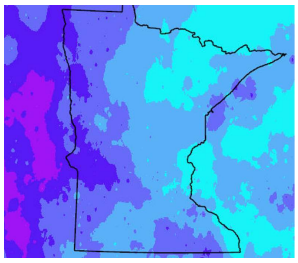


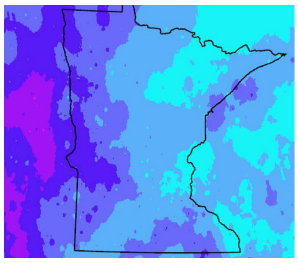
21:00 22:00 23:00 24:00

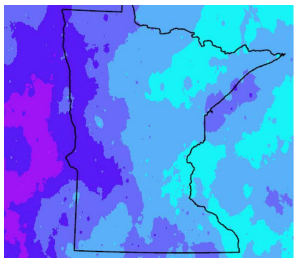
Pressure - Kriging

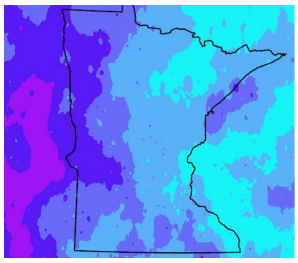


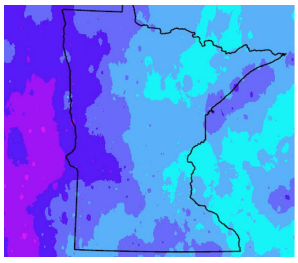
Wind Speed - First-order IDW

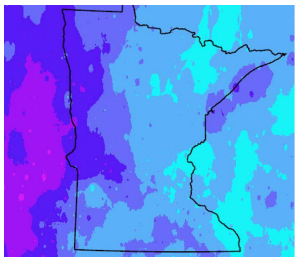

01:00

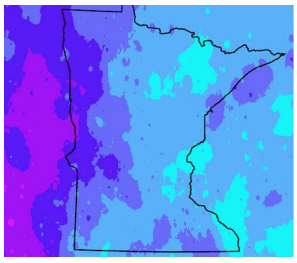

02:00

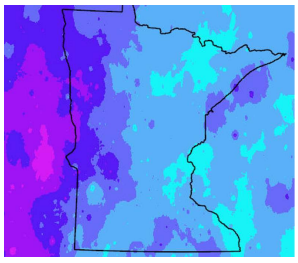

03:00

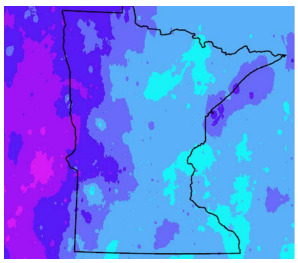

04:00

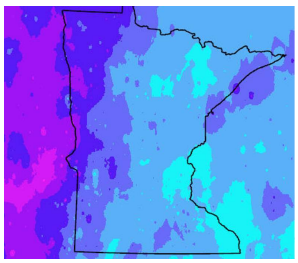

05:00

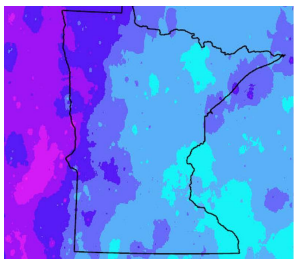

06:00

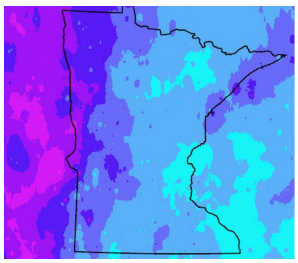

07:00

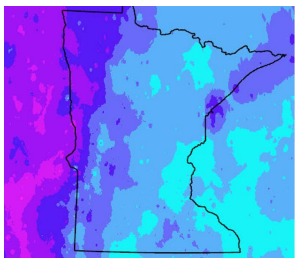

08:00

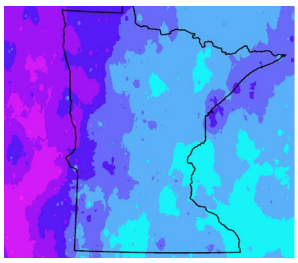

09:00

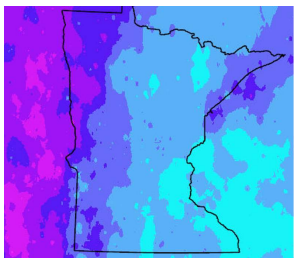

10:00

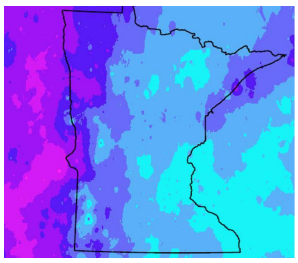

11:00

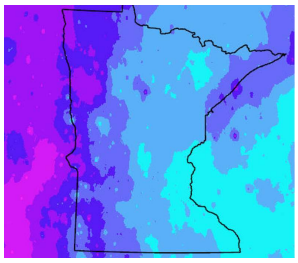

12:00

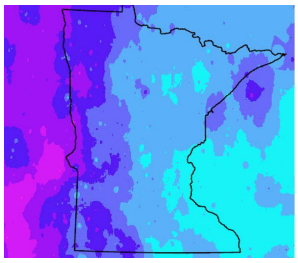

13:00

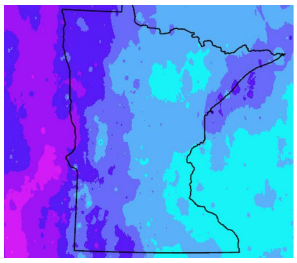

14:00

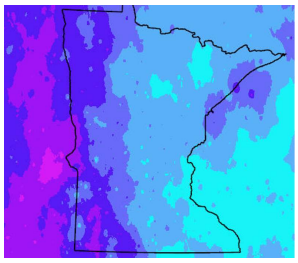

15:00

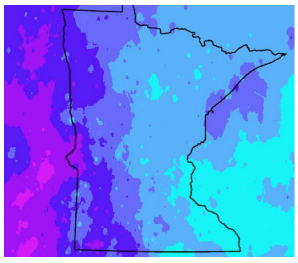

16:00

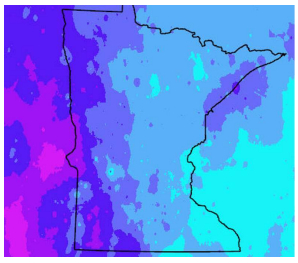

17:00

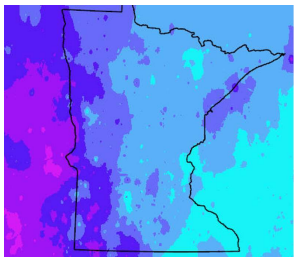

18:00

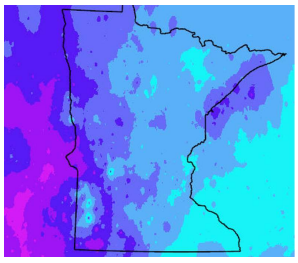

19:00


20:00

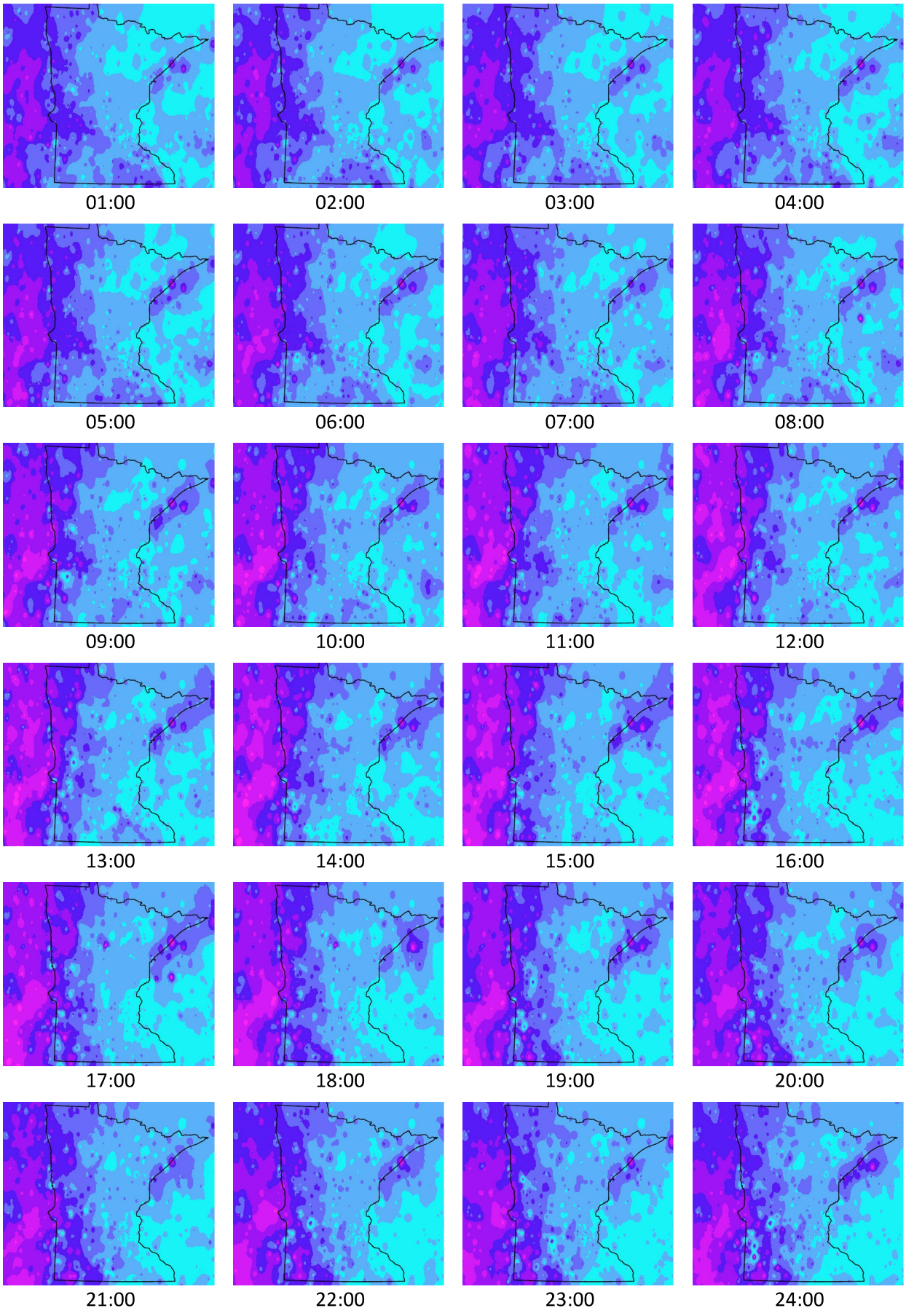

21:00


22:00

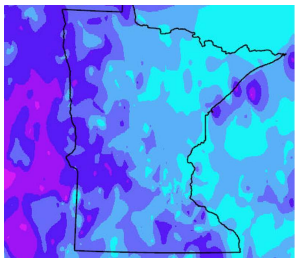

23:00

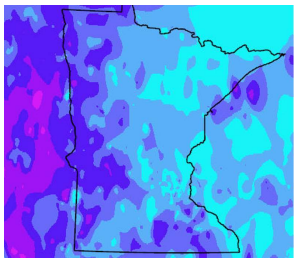

24:00

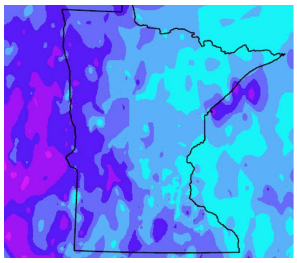
Wind Speed - Second-order IDW

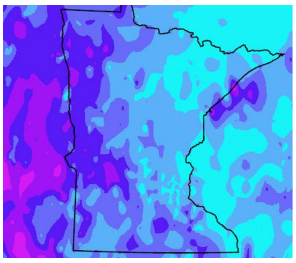


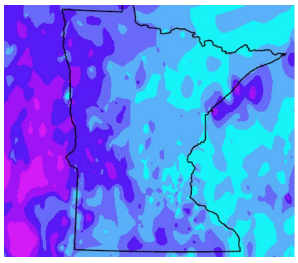
Wind Speed - Natural Neighbor

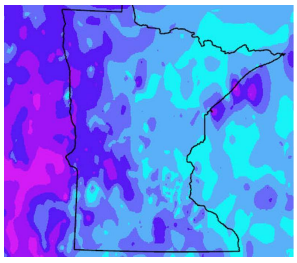

01:00

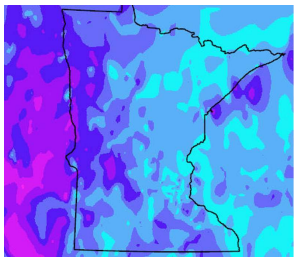

02:00

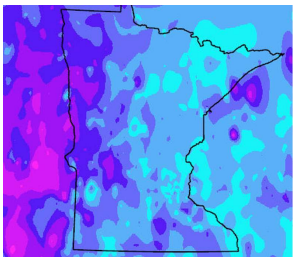

03:00

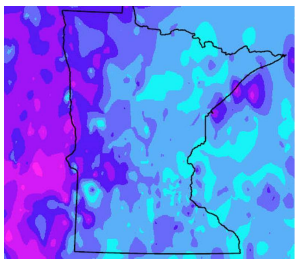

04:00

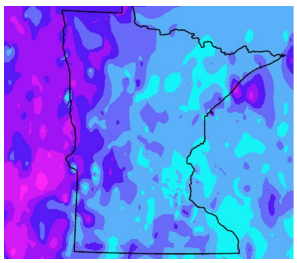

05:00

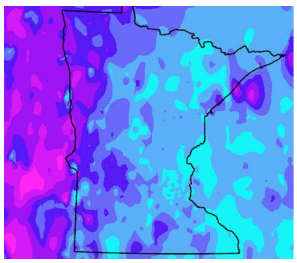

06:00

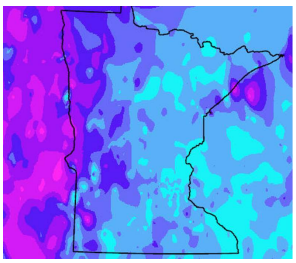

07:00

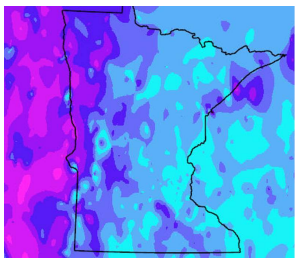

08:00

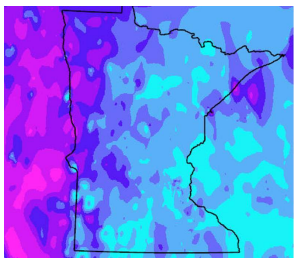

09:00

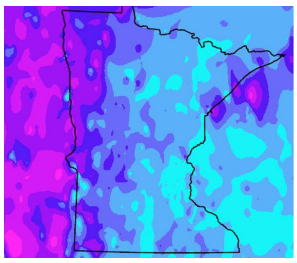

10:00

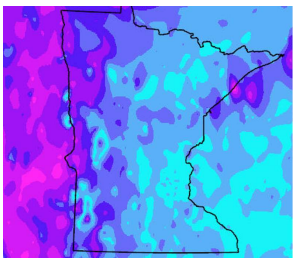

11:00

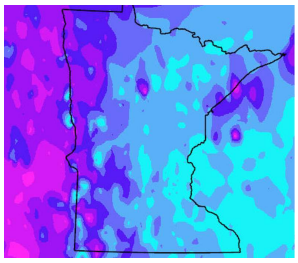

12:00

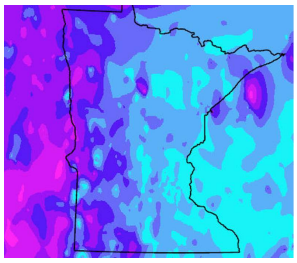

13:00

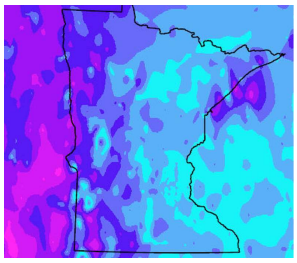

14:00

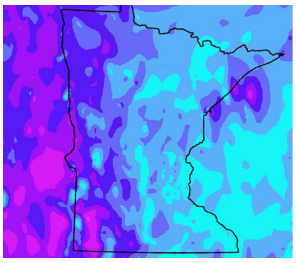

15:00

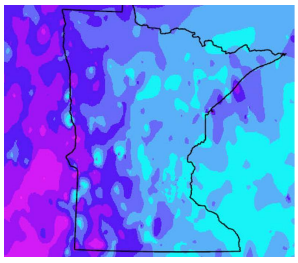

16:00

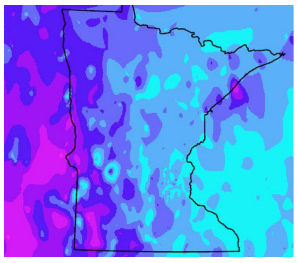

17:00

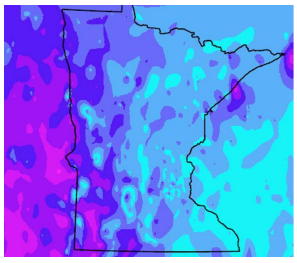

18:00

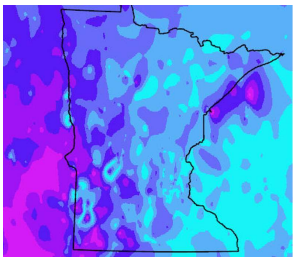

19:00


20:00

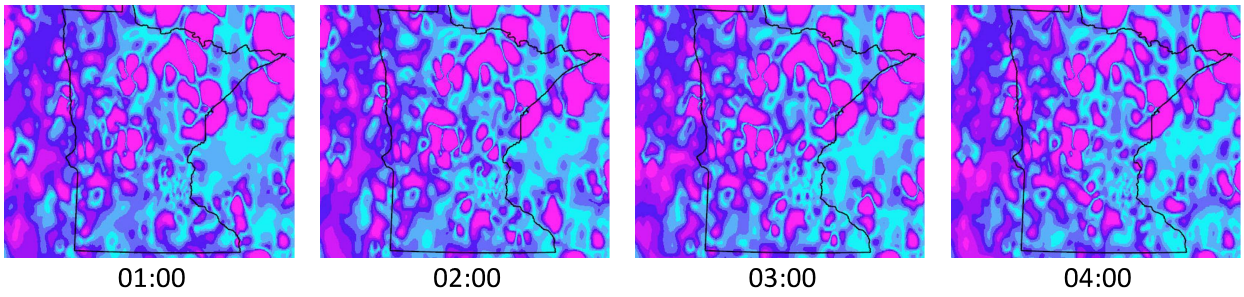

21:00

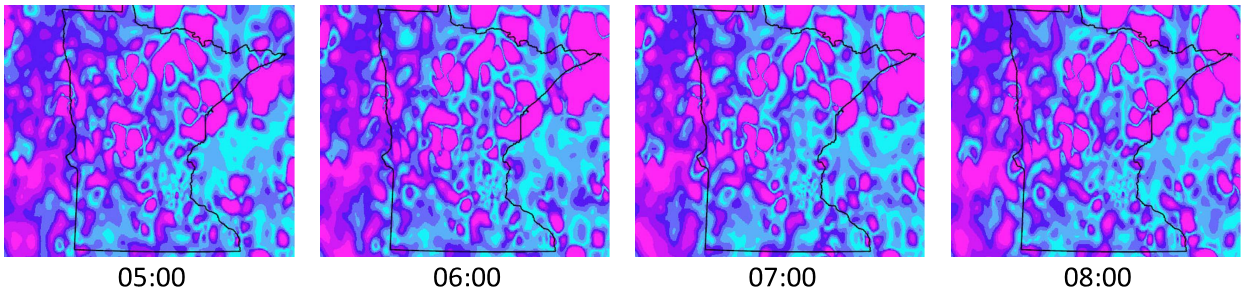

22:00

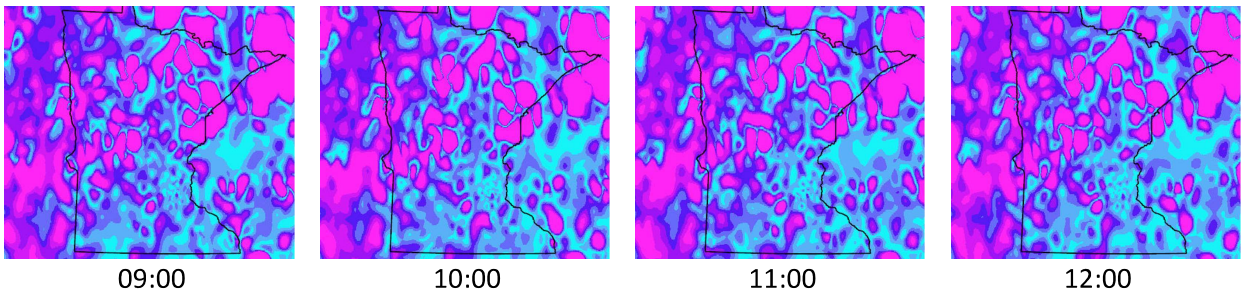

23:00

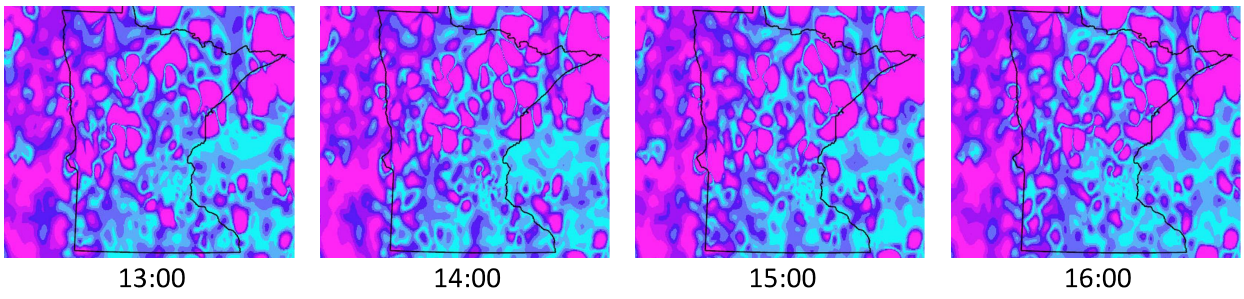

24:00

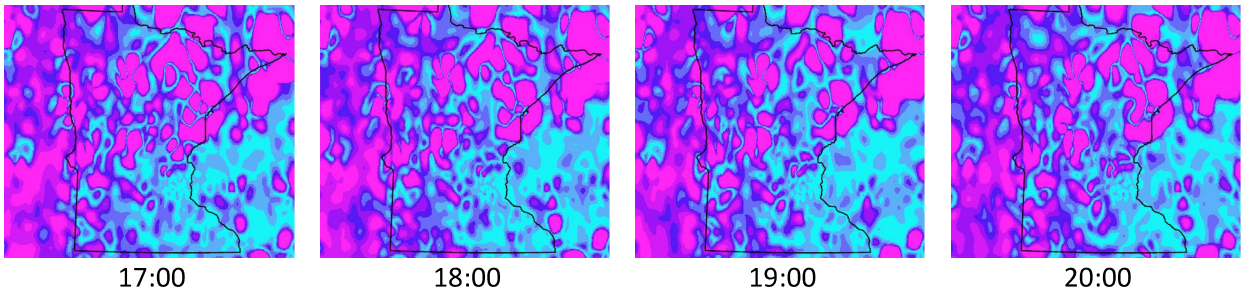
Wind Speed - Spline

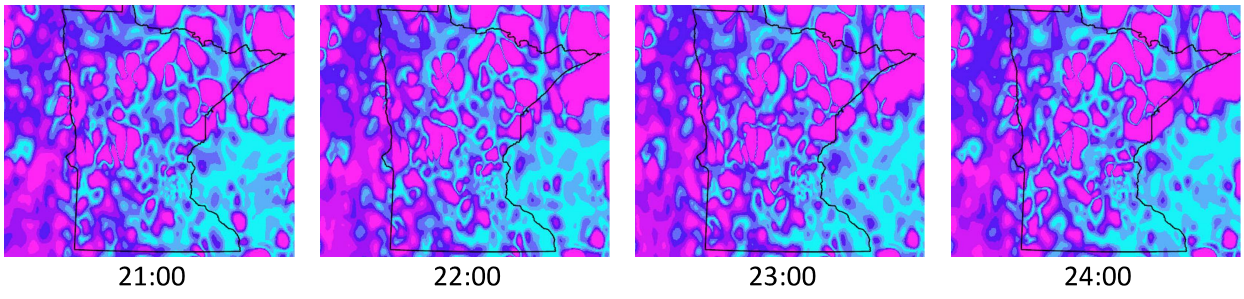




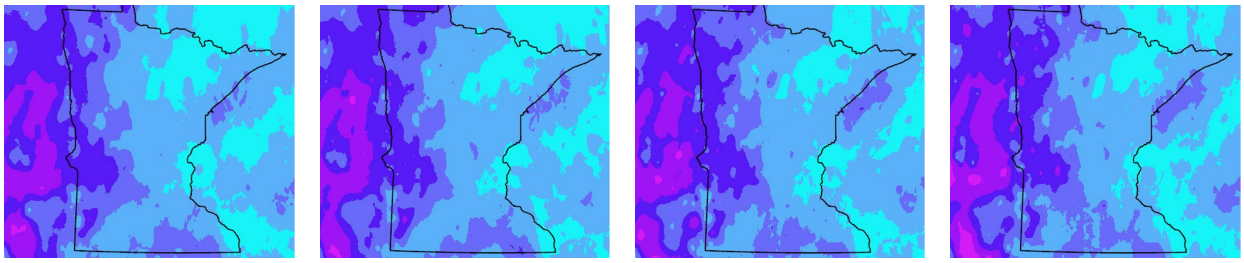


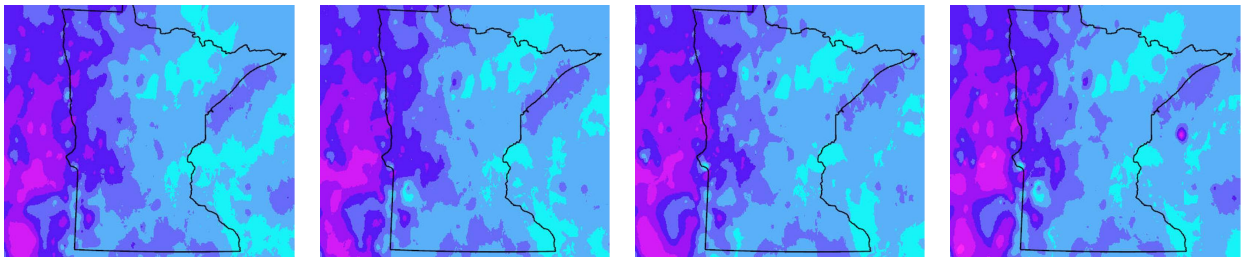


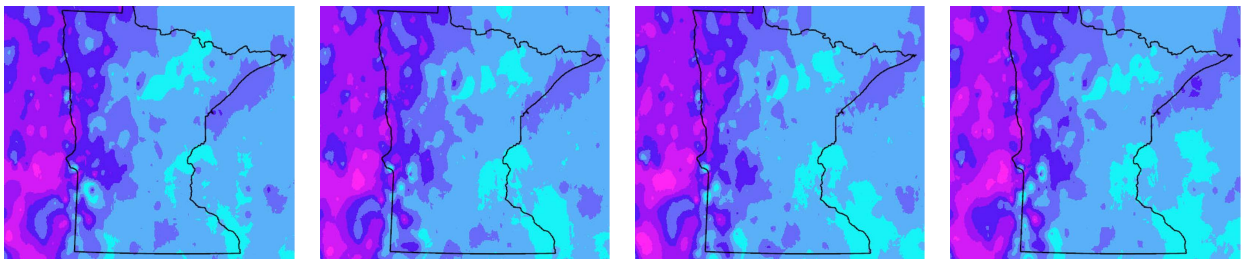


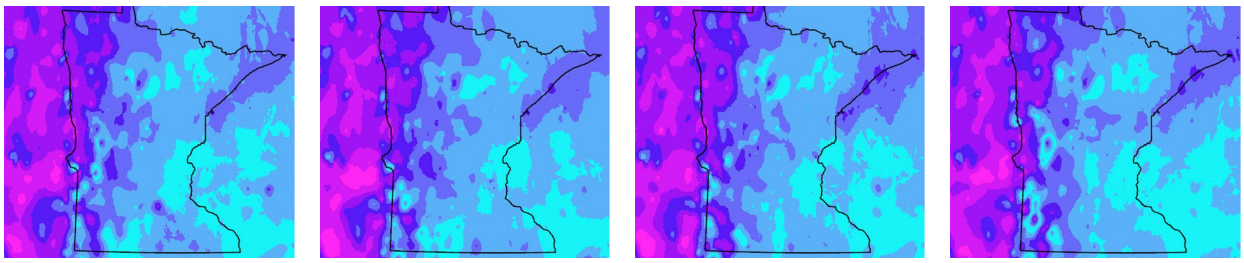


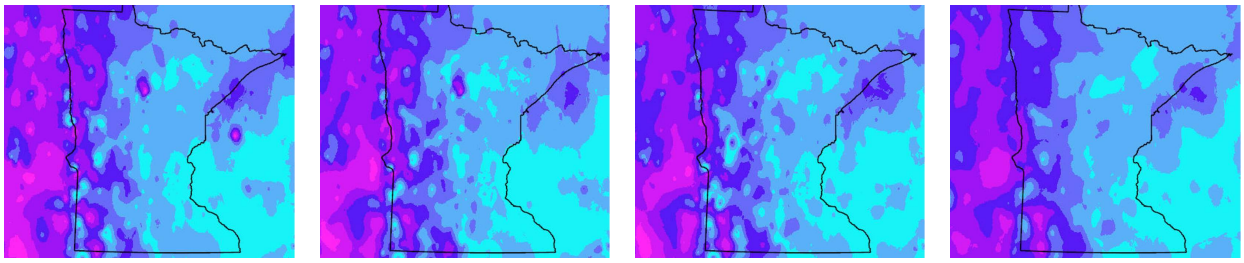
Wind Speed - Kriging

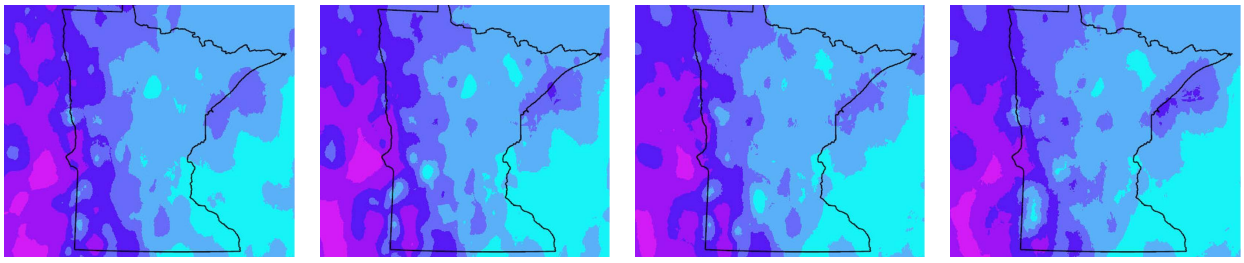

01:00 02:00 03:00 04:00


05:00 06:00 07:00 08:00

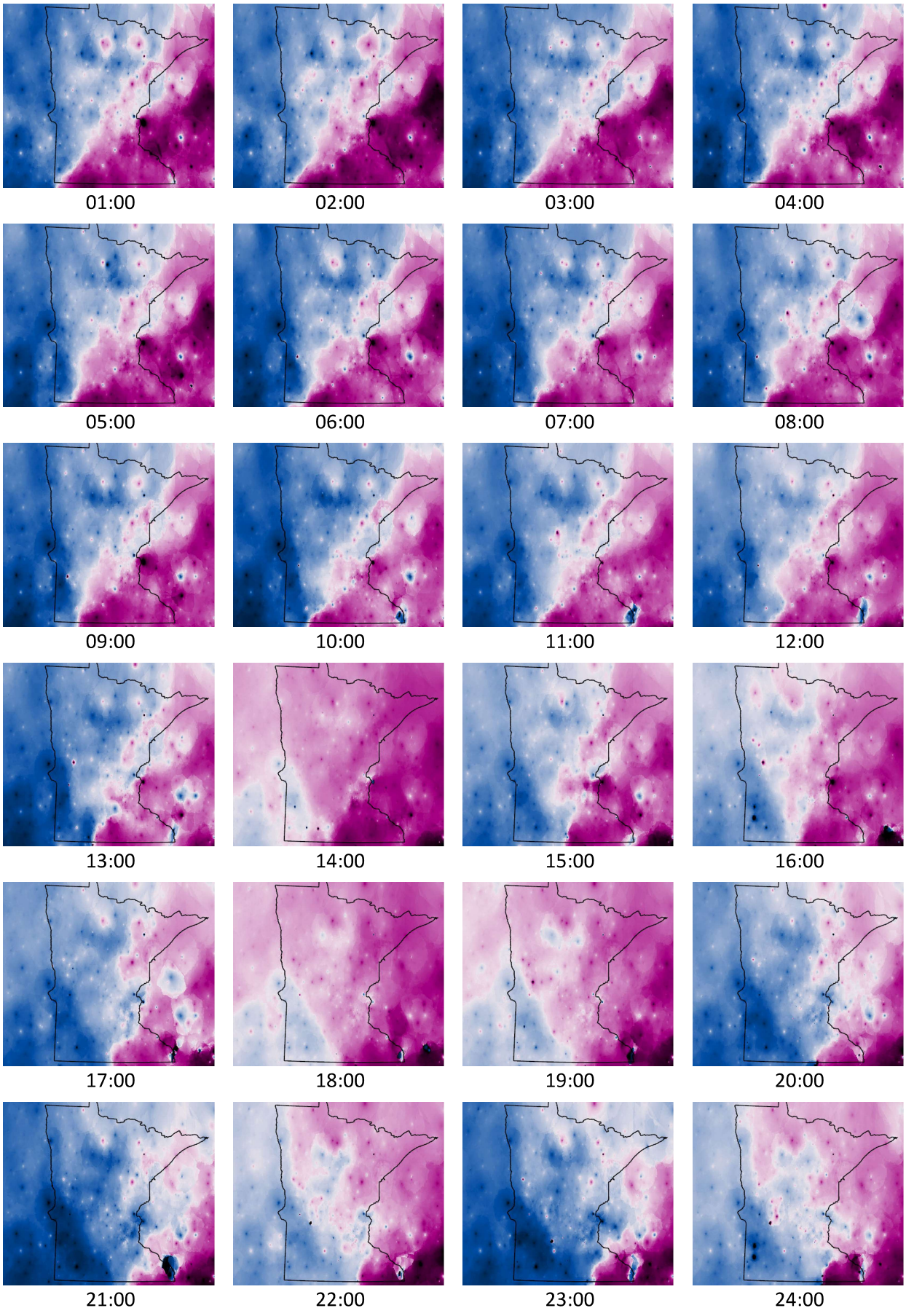

09:00 10:00 11:00 12:00


13:00 14:00 15:00 16:00

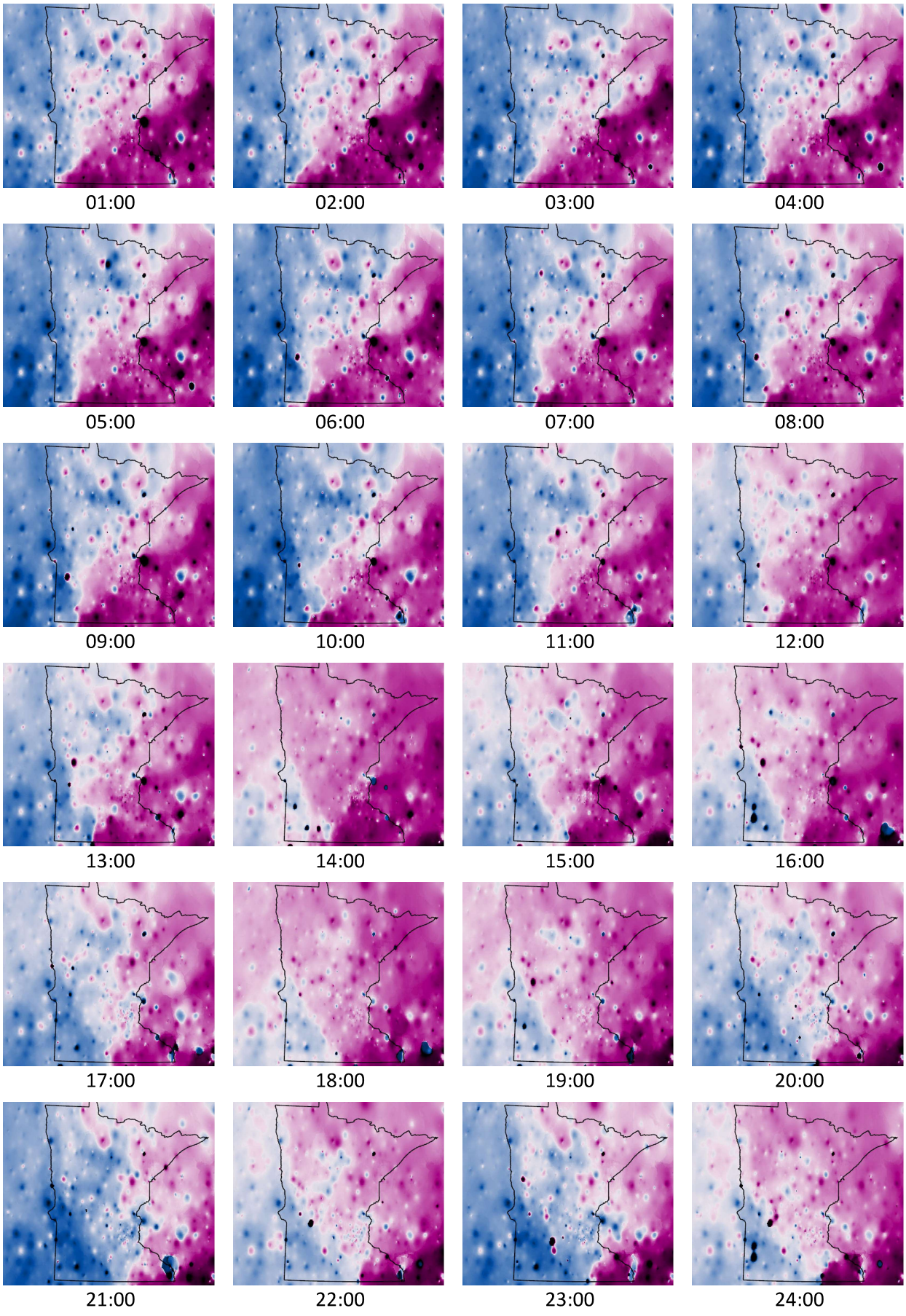

17:00 18:00 19:00 20:00


21:00 22:00 23:00 24:00

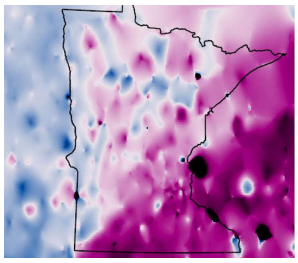
Wind Direction - First-order IDW

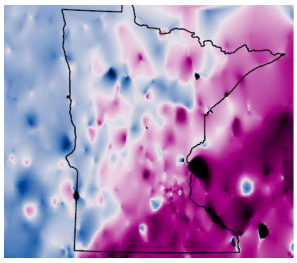


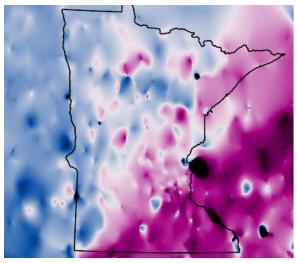
Wind Direction - Second-order IDW

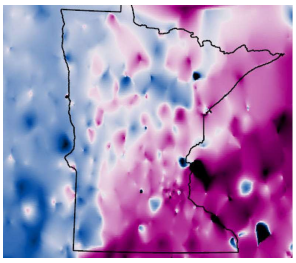


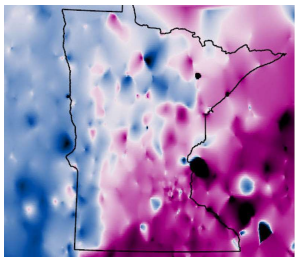
Wind Direction - Natural Neighbor

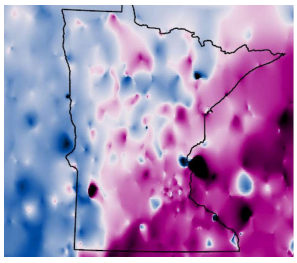

01:00

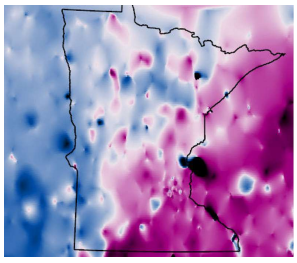

02:00

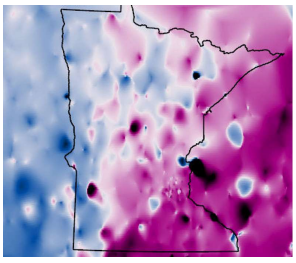

03:00

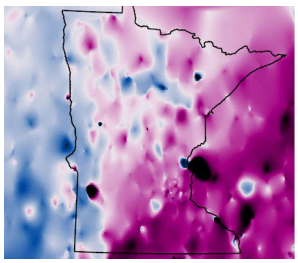

04:00

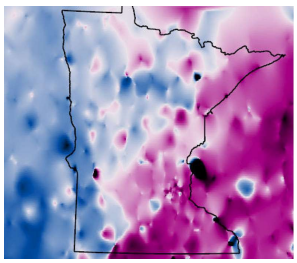

05:00

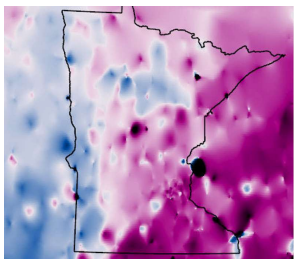

06:00

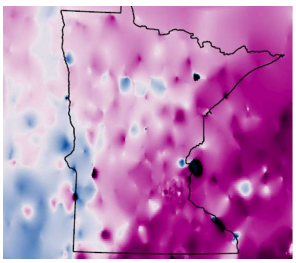

07:00

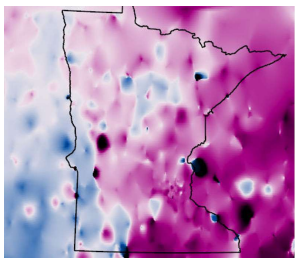

08:00

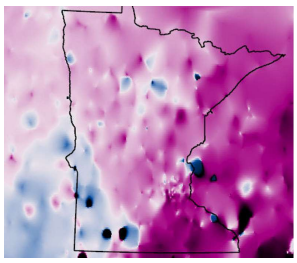

09:00

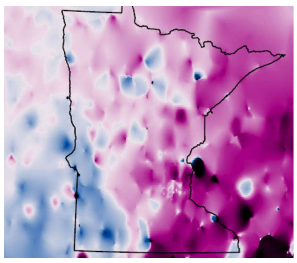

10:00

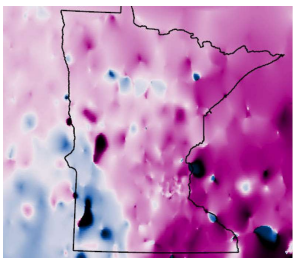

11:00

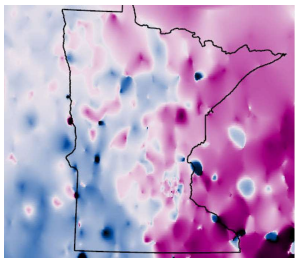

12:00

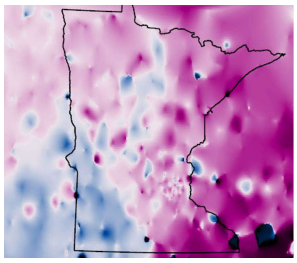

13:00

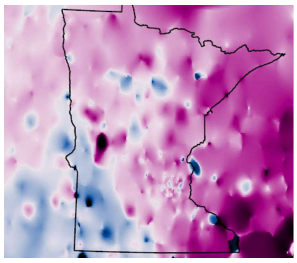

14:00

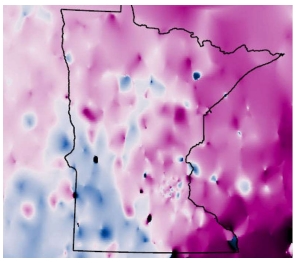

15:00

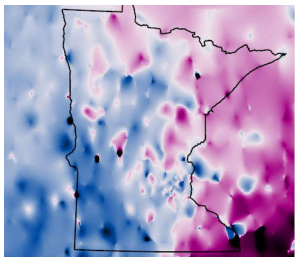

16:00

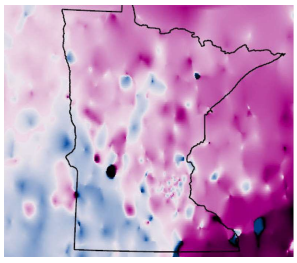

17:00

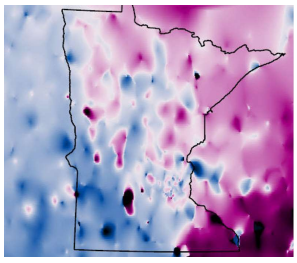

18:00

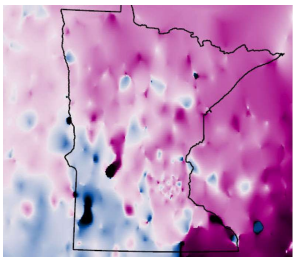

19:00


20:00

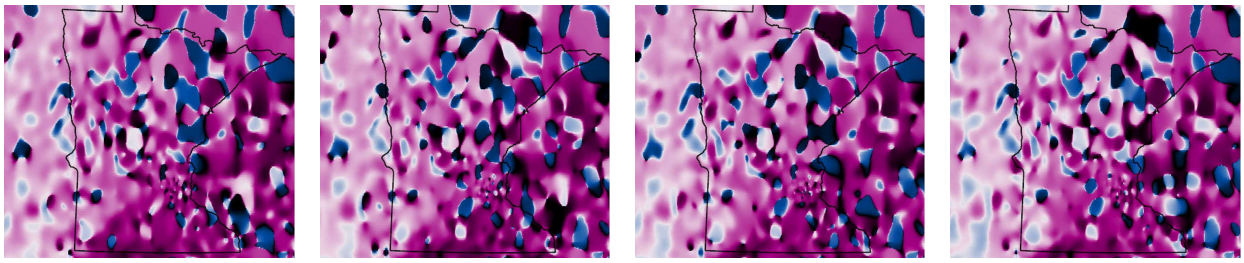

21:00

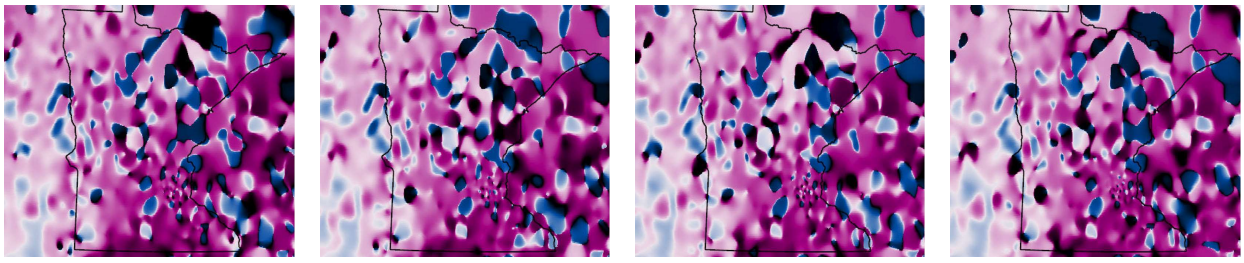

22:00

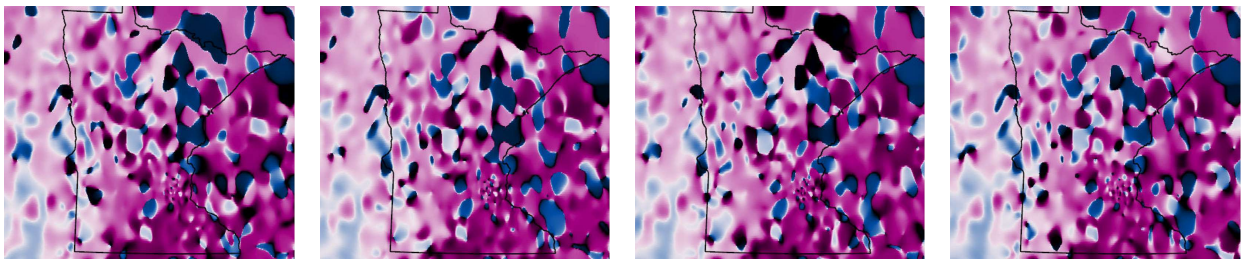

23:00

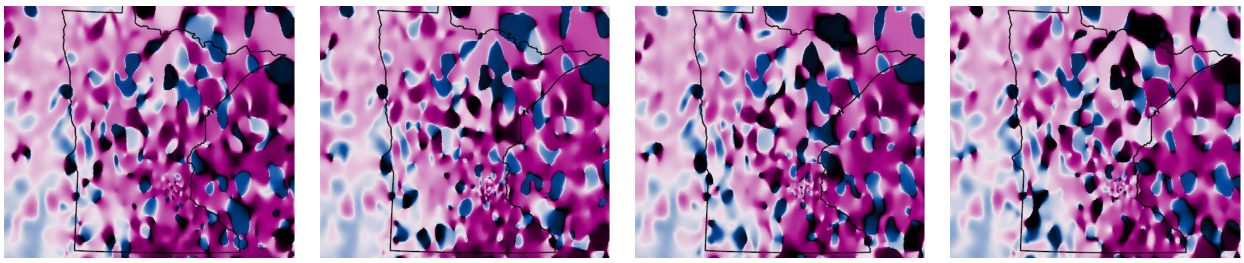

24:00

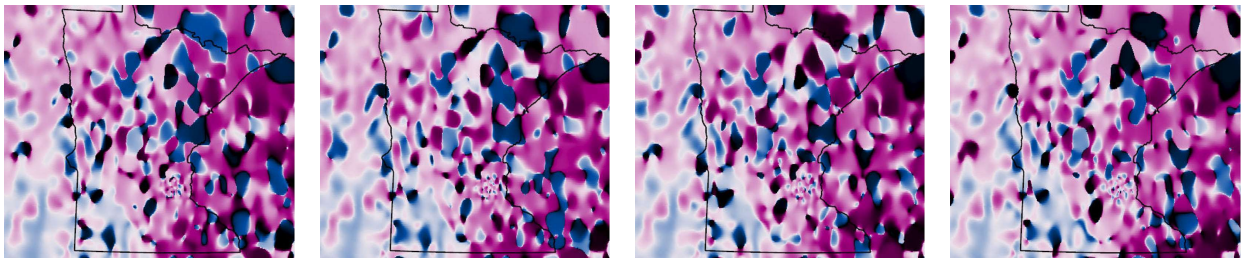
Wind Direction - Spline

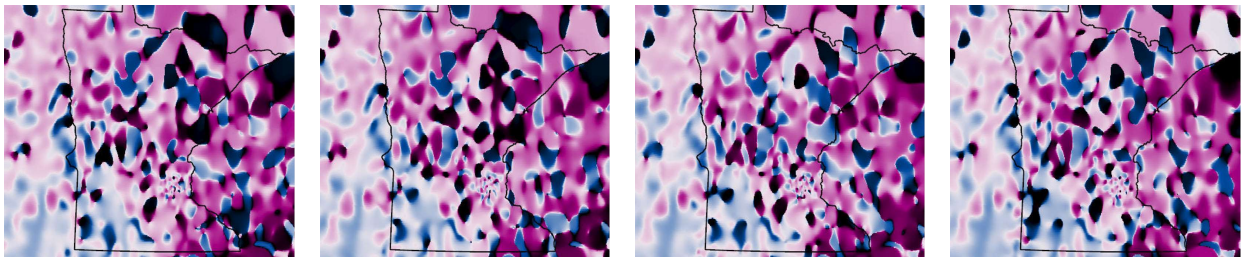

01:00 02:00 03:00 04:00


05:00 06:00 07:00 08:00


09:00 10:00 11:00 12:00


13:00 14:00 15:00 16:00


17:00 18:00 19:00 20:00


21:00 22:00 23:00 24:00

Wind Direction - Kriging