1. Range: 51.3 – 0.2 = **51.1**

Mean: **9.33**

Median: **4.95**

Standard Deviation: **11.12**

Interquartile Range: **6.82**

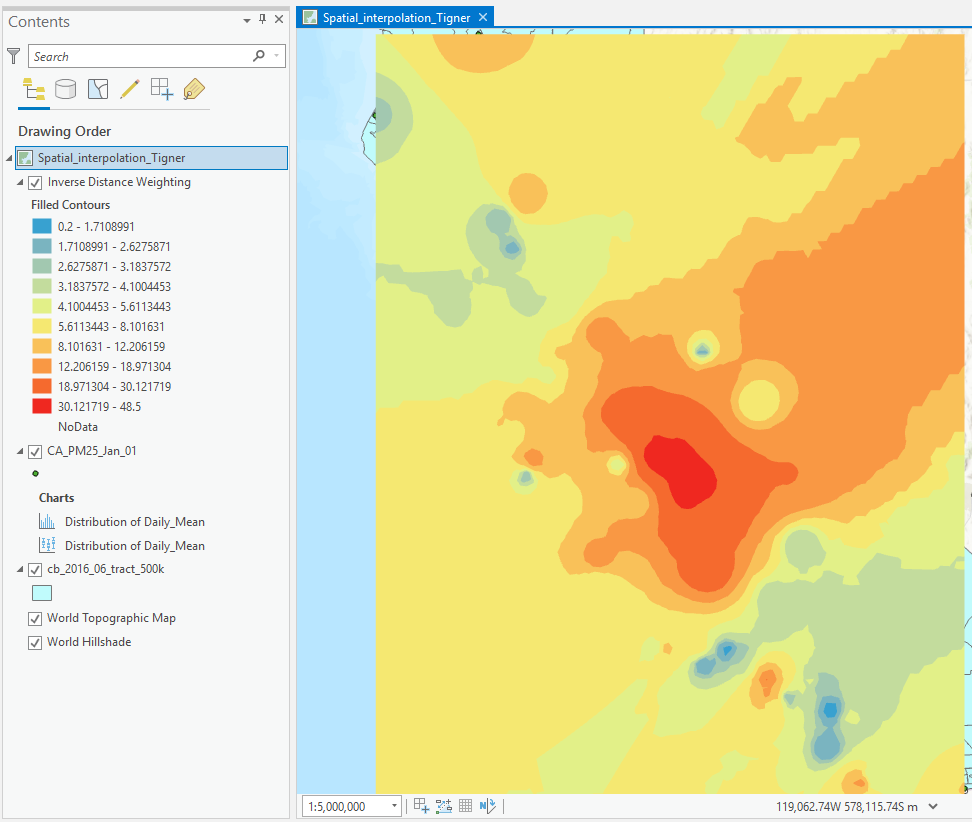
Skewness: **2.19**

Kurtosis: **7.36**

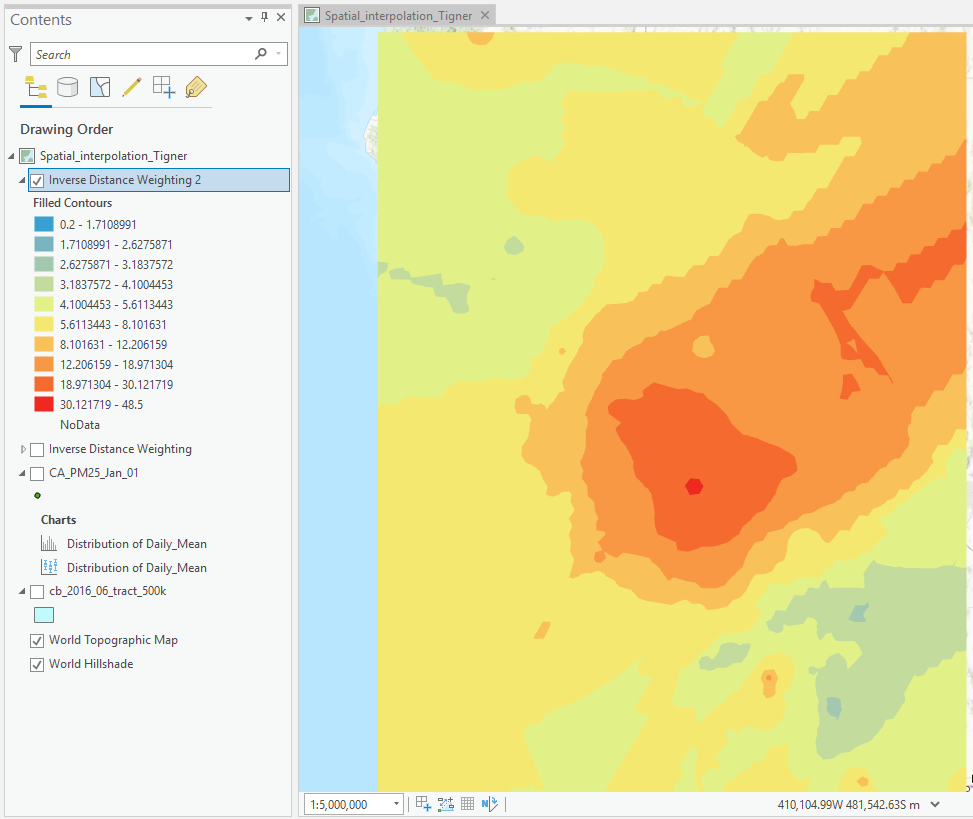
PM2.5 values **do not look normally distributed** since there is a high degree of kurtosis and skewness (to the left).

1. After performing a logarithmic transformation, the values appear to be **normally distributed**, since the skewness is near 0 and the kurtosis is near 3.
2. IDW model’s Mean Error: **0.43**

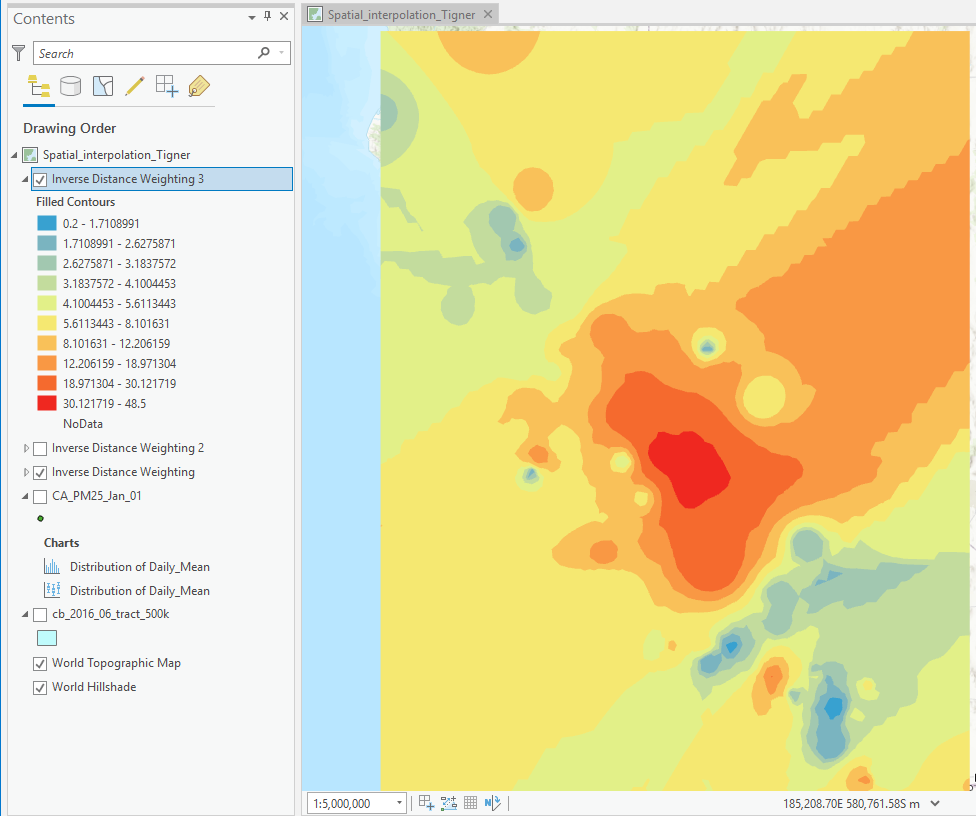
IDW model’s Root-Mean-Square error: **7.13**



Areas with the highest values are San Jose, San Diego, and Fresno.



The high values are more consolidated and less distributed across the area, since the distance decay is weaker with power=1. There is a weaker localized effect.



The results of this IDW are similar to that in Q4, except the areas of high and low values are larger.

1. The higher the power, the stronger the distance decay effect is, explaining the difference between Q4 and Q5. When the number of neighbors is increased, unknown values change, since they are assigned the weighted average of neighboring points (the amount of which has increased), explaining the difference between Q4 and Q6.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Q4 | Q5 | Q6 |
| Mean Error | 0.42 | 0.05 | 0.57 |
| Root-Mean-Squared Error | 7.12 | 7.10 | 7.30 |

The IDW parameters used in **Q5** (power=1) produced the lowest mean error and root-mean-squared error and therefore provided the best interpolation result.

1. Using Kriging with the default parameters, we get:

Nugget: **0.6**

Major Range: **316,702.44**

Partial Sill: **0.85**

Sill: 0.6 + 0.85 = **1.45**

1. Mean Error: **0.12**

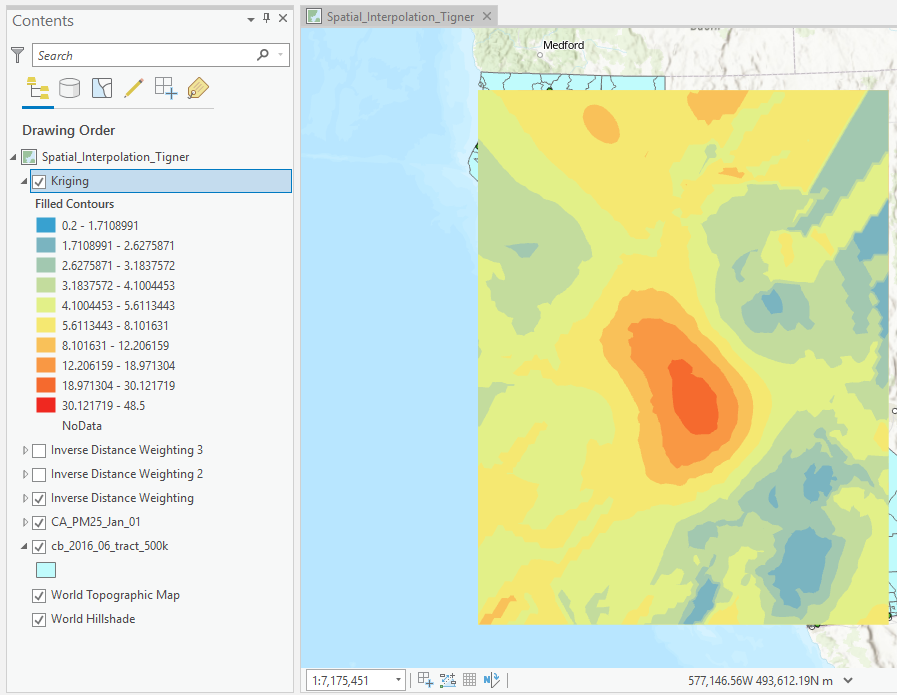
Root-mean-Square Error: **7.39**

Mean Standardized Error: **-0.06**

Root-Mean-Square Standardized Error: **0.96**

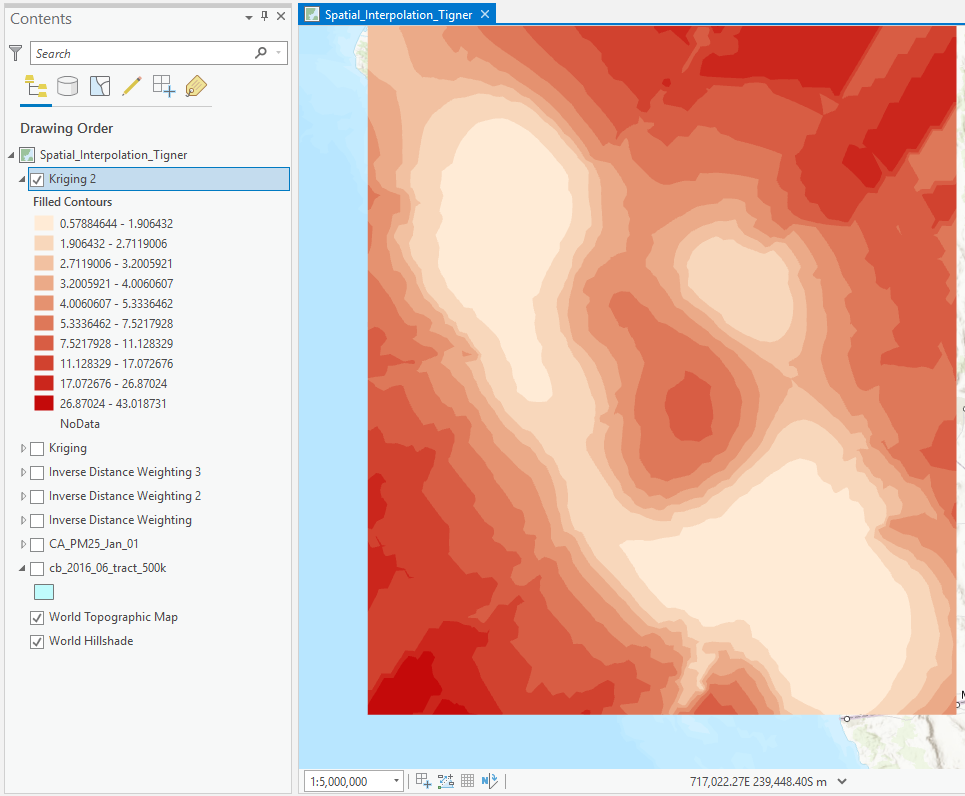
Average Standard Error: **10.82**





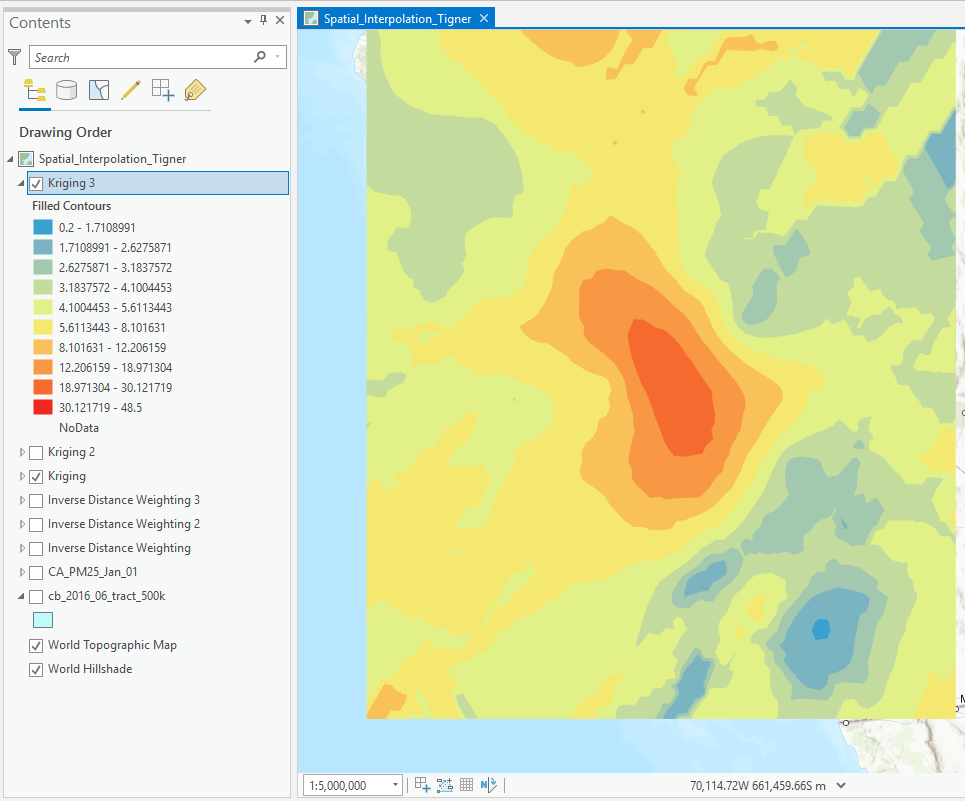
The highest values are in the San Joaquin Valley near Fresno, with areas of low values in the northwest and east.





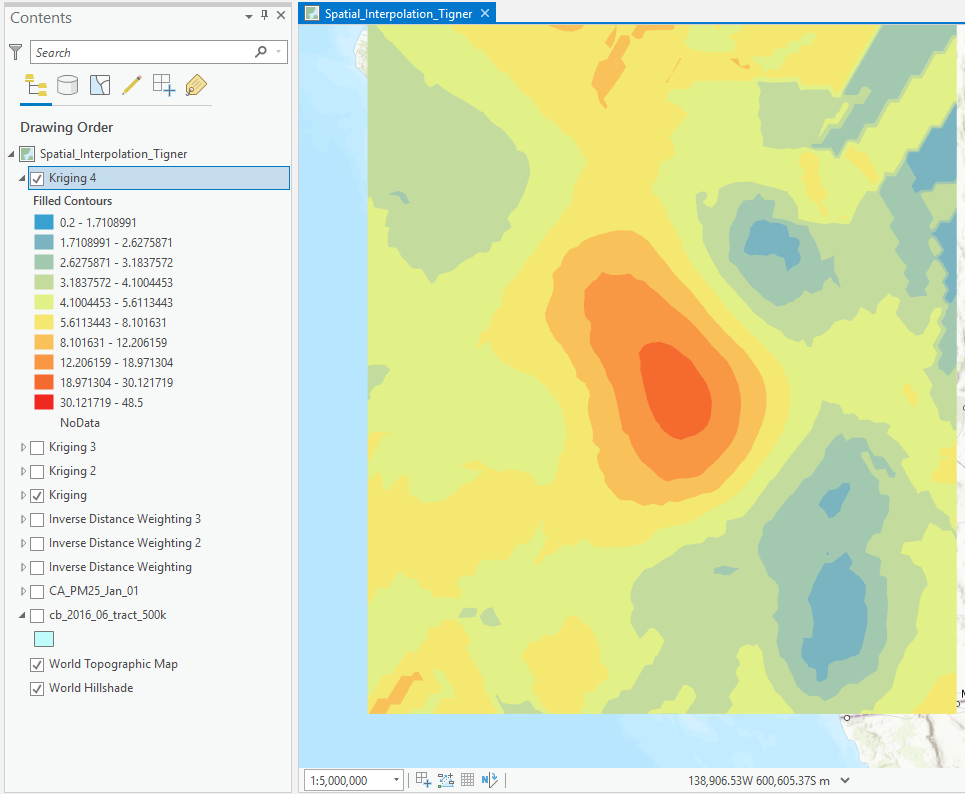
The predicted standard error seems to be highest in the southwest northeast. This is because these regions cover the ocean and Nevada, respectively, where no points of data exist.





When compared to Q10 the above results have larger areas of high values (in the center).





Increasing the maximum number of neighbors from 5 to 10 seemed to smooth out the dispersion of high and low values but did not have a major impact otherwise.



|  |  |  |  |
| --- | --- | --- | --- |
|  | Q10 | Q12 | Q13 |
| Mean Error | 0.12 | 0.13 | 0.07 |
| Root-mean-Square Error | 7.39 | 7.17 | 7.46 |
| Mean Standardized Error | -0.06 | -0.07 | -0.07 |
| Root-Mean-Square Standardized Error | 0.96 | 0.97 | 0.99 |
| Average Standard Error | 10.81 | 10.87 | 10.88 |

The prediction map generated in Q10 is the best, as shown above.

1. IDW and Kriging both produced continuous surfaces over the study region. IDW resulted in interpolations showing high values in the center of the study region and moving (while decreasing) in the northeast. Kriging produced high values in the center of the study region with lower values in the east and northwest. Both showed the highest values in the center of the study area, near Fresno. Both IDW and Kriging interpolations do not cover the entirety of California. Another disadvantage of both is they both interpolate over regions such as the Ocean and Nevada, where no points exist whatsoever (meaning high likelihood of error in these areas). IDW assigns unknown values as a weighted average of surrounding points, while Kriging creates a general model that describes how the sample values vary with distance and direction. Because of these definitions, IDW is likely better suited for local/small scale analyses while Kriging is best at overall trends.