The Payne Institute for Public Policy

COLORADOSCHOOL**OF MINES**
EARTH ● ENERGY ● ENVIRONMENT

Aggregation and Analysis of Methane Data in the DJ Basin, Colorado

William Daniels[*1], Dorit Hammerling[†1], and Morgan D. Bazilian[‡2]

[1]Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden CO
[2]Payne Institute for Public Policy, Colorado School of Mines, Golden CO

May 5, 2020

Abstract

This paper briefly considers emissions data in the Denver-Julesburg (DJ) Basin. We focus on methane data from the TROPOMI instrument on board the Copernicus Sentinel-5 Precursor satellite. We have aggregated the data into a variety of temporal packets and performed an initial exploratory analysis. This review will help inform ongoing and future air pollution monitoring efforts. These efforts rely on data gathered from a wide array of monitoring techniques, including ground-level sensors, drones, and planes. Being able to better incorporate satellite data into these efforts will offer a more complete emissions profile, which can be used to inform both operations and regulations.

Keywords: emissions data, methane, satellite data, Denver-Julesburg Basin, TROPOMI

[*]wdaniels@mines.edu
[†]hammerling@mines.edu
[‡]mbazilian@mines.edu

Contents

The Payne Institute *for* Public Policy

1 Introduction

The Denver-Julesburg (DJ) Basin is a major oil and gas producing region centered in eastern Colorado that extends into southeast Wyoming, western Nebraska, and western Kansas. Monitoring emissions in this region is an important aspect of responsible oil and gas development. Emissions typically come from flares, venting, and fugitives, and they primarily consist of carbon dioxide, methane, and volatile organic compounds (VOCs). The Digital Canopy project seeks to integrate data from multiple levels of emission detection technology to effectively monitor the DJ Basin. Specifically, the Digital Canopy project has the following objectives:

1. **Collect** pertinent emissions data in the DJ basin.

2. **Analyze** and compare different data sources.

3. **Compile** the data sources into an accessible database.

Collect: Ideally, there will be three levels of emissions data: on-site sensors, drone and plane-based monitoring systems, and satellite observations. Combining these three levels will give us a variety of spatial and temporal resolutions with which to capture both high and low frequency trends in the emissions data for different spatial areas. Additionally, we will consider looking beyond the emissions themselves to secondary data sources that could potentially be related to the causes of emissions. One option for this type of data are the VIIRS DNB temporal profiles. VIIRS is an instrument on board the Suomi National Polar-orbiting Partnership (NPP) satellite platform, and it provides global satellite observations spanning the visible and infrared wavelengths. In particular, the VIIRS day/night band (DNB) is able to observe nighttime lights with better spatial and temporal resolutions than previous instruments. We hope to use VIIRS DNB data to extract flaring information that could potentially be correlated with emissions data. In fact, a dataset specifically for flares has already been created, which contains 12,000 flaring sites in 2015 alone [1].

Analyze: We will compare the different data sources discussed above to see if they are compatible and can be combined into a larger database. As a first step, we will focus on satellite data from the TROPOMI instrument. We have selected the TROPOMI instrument because it provides the best quality and most recent high-resolution satellite data for atmospheric gasses such as methane. Furthermore, other notable studies have used TROPOMI methane data to investigate oil and gas emissions. For instance, see [2]. In this exploratory stage, we will analyze the TROPOMI satellite data and gain a better understanding of how it can contribute to the Digital Canopy project.

Compile: As part of the Digital Canopy project, all pertinent data will be packaged in easily accessible formats and ultimately compiled into a larger database.

In this report, we focus on the aggregation and analysis of TROPOMI satellite data over the DJ Basin. This data could potentially be integrated into the larger Digital Canopy project described above. We further narrow our analysis by focusing solely on methane, a significant component of oil and gas emissions. The remainder of this report is laid out as follows. In Section 2, we describe the TROPOMI data and the spatial region we have chosen to study. In Section 3, we visualize the TROPOMI data spatially over our chosen region, and in Section 4, we visualize it temporally. In Section 5, we perform an initial exploratory analysis into the TROPOMI methane data. In Section 6, we conclude our findings.

2 Methane Data

2.1 TROPOMI

The TROPOspheric Monitoring Instrument (TROPOMI) on board the Copernicus Sentinel-5 Precursor satellite measures key atmospheric constituents including O_3, NO_2, SO_2, CO, and CH_4. It has relatively high spatial and temporal resolutions. Spatially, it has a resolution of 7 km $\times$ 3.5 km, meaning that 7 km $\times$ 3.5 km is the smallest feature that TROPOMI can distinctly make out. Anything smaller cannot be directly resolved and will be indistinguishable from other sources in the 7 km $\times$ 3.5 km area. Temporally, it passes over every geospatial point around one time a day. This means that a new TROPOMI measurement is available about once per day per location. TROPOMI was launched in October 2017 and has been publicly releasing data since May 2018 [3]. Therefore, a little under two years of data are available for analysis.

2.2 Region of Interest

We have downloaded TROPOMI methane measurements over the DJ basin from May 2018 through February 2020. Specifically, we have focused on the following area: -107 to -100 degrees in longitude and 38 to 43 degrees in latitude. Figure 1 shows the continental United States and our region of interest highlighted in green.
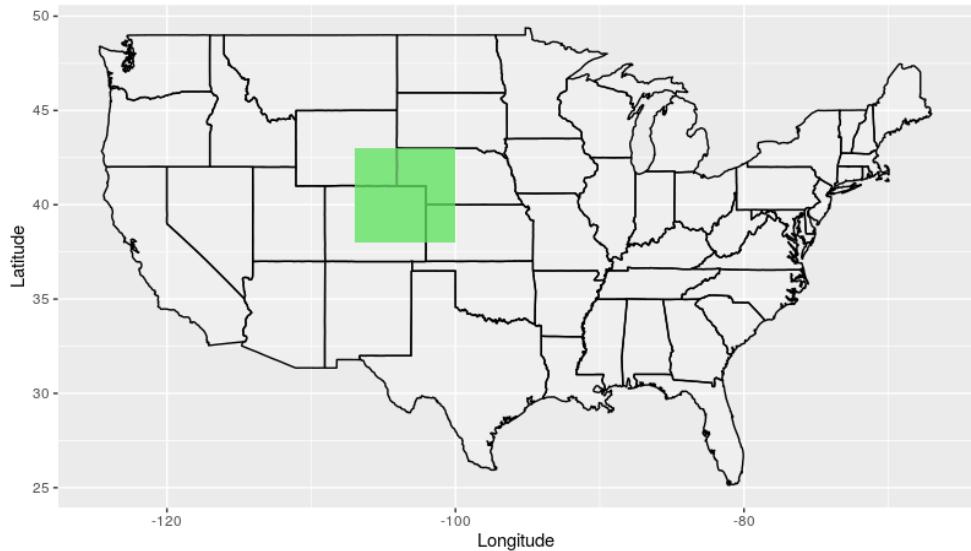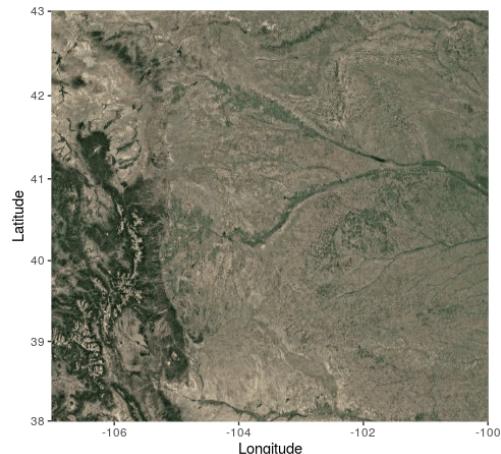

Figure 1: The green box indicates the region in which we have collected TROPOMI methane data.

Figure 2 shows our region of interest in more detail. Subfigure (a) shows roads and state boundaries within the region. Subfigure (b) is a satellite image of the region. Figure 3 shows our region of interest with some pertinent information overlaid. The orange region is the DJ basin, and the red region is the Niobara play. Bayswater Production and Exploration wells are shown in black. Well locations were obtained from the Colorado Oil and Gas Conservation Commission (COGCC) website [4].

(a) Road map and state lines.

(b) Satellite image.

Figure 2: Zoomed in plots of our region of interest. TROPOMI methane data has been collected in this region from May 2018 through February 2020.
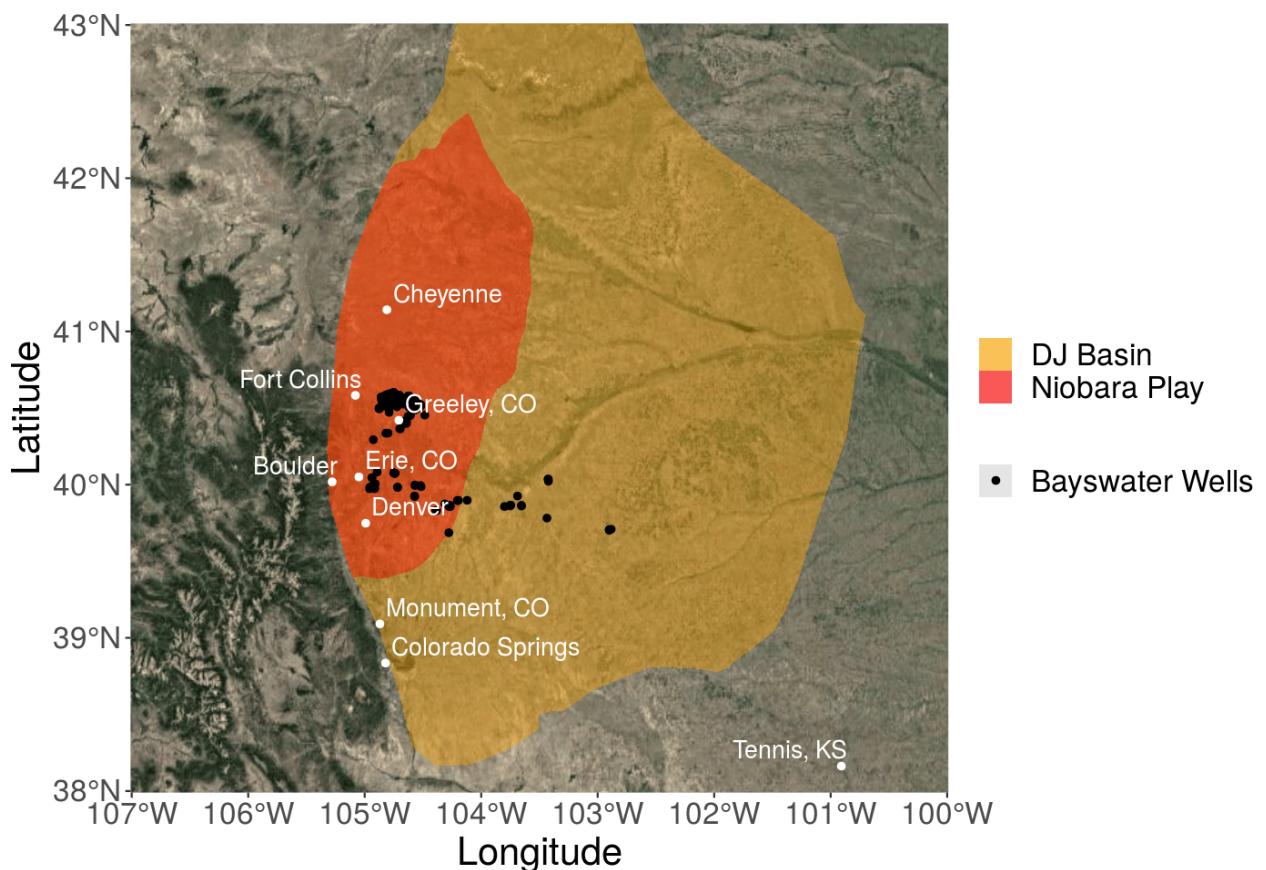

Figure 3: Zoomed in plot of our region of interest. Plotted in orange is the DJ basin, and plotted in red is the Niobara play. Bayswater Production and Exploration wells are shown in black.

2.3 Data Format

We have aggregated the TROPOMI methane data into gridboxes. Figure 4 is a visual representation of this gridding system, with each red square representing a gridbox. All methane measurements falling into a given gridbox will be averaged on whatever timescale is being used. For instance, if performing weekly analysis, all data measured during a given week and falling within a given gridbox will be averaged. The resolution of the grid can be easily changed. For instance, Figure 4 shows a 40×40 grid, which is the standard size used in this report. This grid size results in gridboxes that are roughly 12 miles wide by 9 miles tall. Grid size is defined as RES×RES, where RES is the number of gridboxes per row and column. Currently, columns and rows of the grid always contain the same number of gridboxes. The result of this grid structure is RES^2 averages for each time period being studied.

We use this grid structure because individual TROPOMI measurements can be quite sparse, making analysis challenging. By creating a grid of averaged measurements, we can more easily capture spatial trends over our region. Furthermore, individual measurements can be erroneous due to cloud cover or surface albedo. Working with a grid of averages smoothes out these erroneous data points, making analysis more reliable.

Note that our region is not a square, as it is 7 degrees wide and only 5 degrees tall. Our current gridding system creates rows and columns with an equal number of gridboxes, which means that the gridboxes are not square either. In this report, each gridbox, like the region itself, is a rectangle with a 7:5 width to height ratio. However, this could be easily changed if a different aspect ratio is desired.
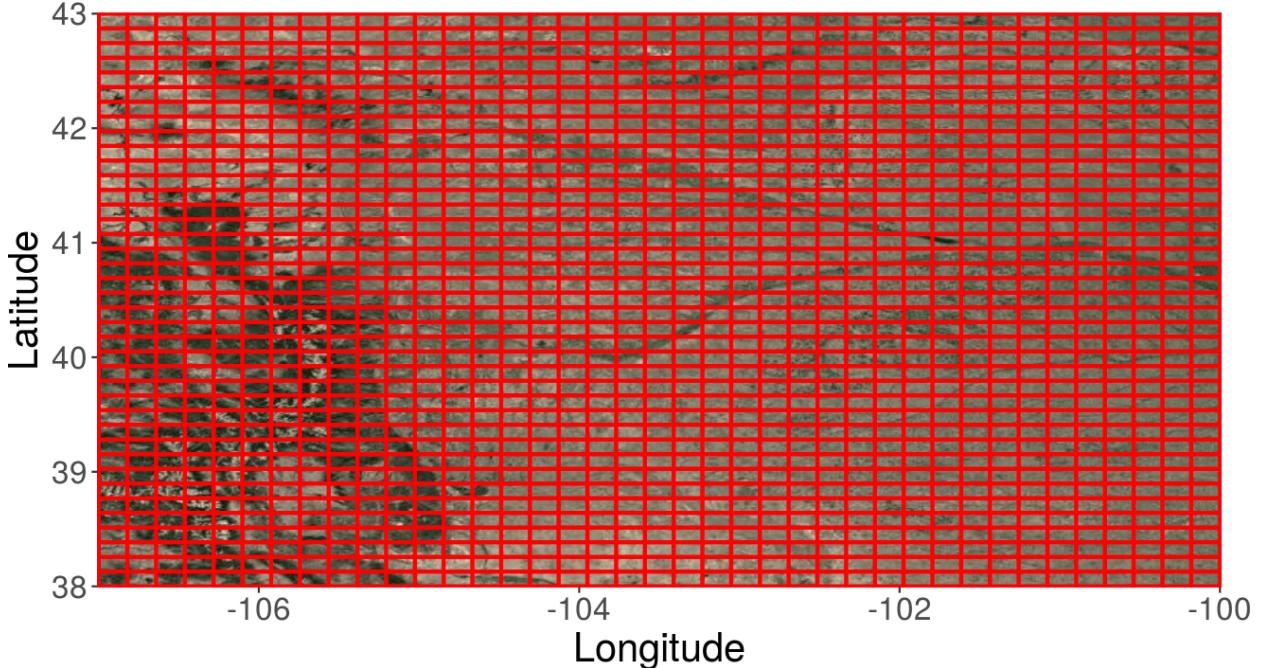

Figure 4: Satellite image of our region of interest with a 40×40 grid overlaid.

3 Spatial Plots

In this section, we present spatial visualizations of the TROPOMI data. In Subsection 3.1, we show averages over the entire time span of the data: May 2018 through February 2020. In Subsection 3.2, we break the data up into smaller timescales.

3.1 Full Time Span

Figure 5 shows the average methane mixing ratio over the entire time span of the data: May 2018 through February 2020. The outer black outline shows the DJ Basin, and the inner black outline shows the Niobara Play. A selection of cities and towns are plotted as well. The heatmap of methane concentrations is slightly opaque, and a satellite image showing topography is plotted underneath.

We can see that there are higher methane concentrations in the southeast corner of our region, possibly due to larger background methane concentrations in Oklahoma and Kansas. Data in the southwest corner of our region is quite sparse. As a result, there are many gridboxes in this area with no data. These gridboxes are not plotted, clearly revealing the topographic map underneath. The data sparsity in this area is most likely due to the presence of mountains, as increased albedo from snow could make satellite retrieval harder. There are also elevated methane concentrations in the area north of Denver, east of Boulder, and southeast of Fort Collins. These concentrations could be due to agriculture or oil and gas developments, both of which are very prevalent in Weld County [5, 4]. Finally, there appears to be a "right caret" pattern of elevated methane concentrations starting northeast of Cheyenne and east of Denver. This "right caret" directly follows the path of the North and South Platte Rivers. Elevated methane concentrations in this area could be due to agriculture or the presence of wetlands, both of which are known to be sources of methane emissions [6, 7].
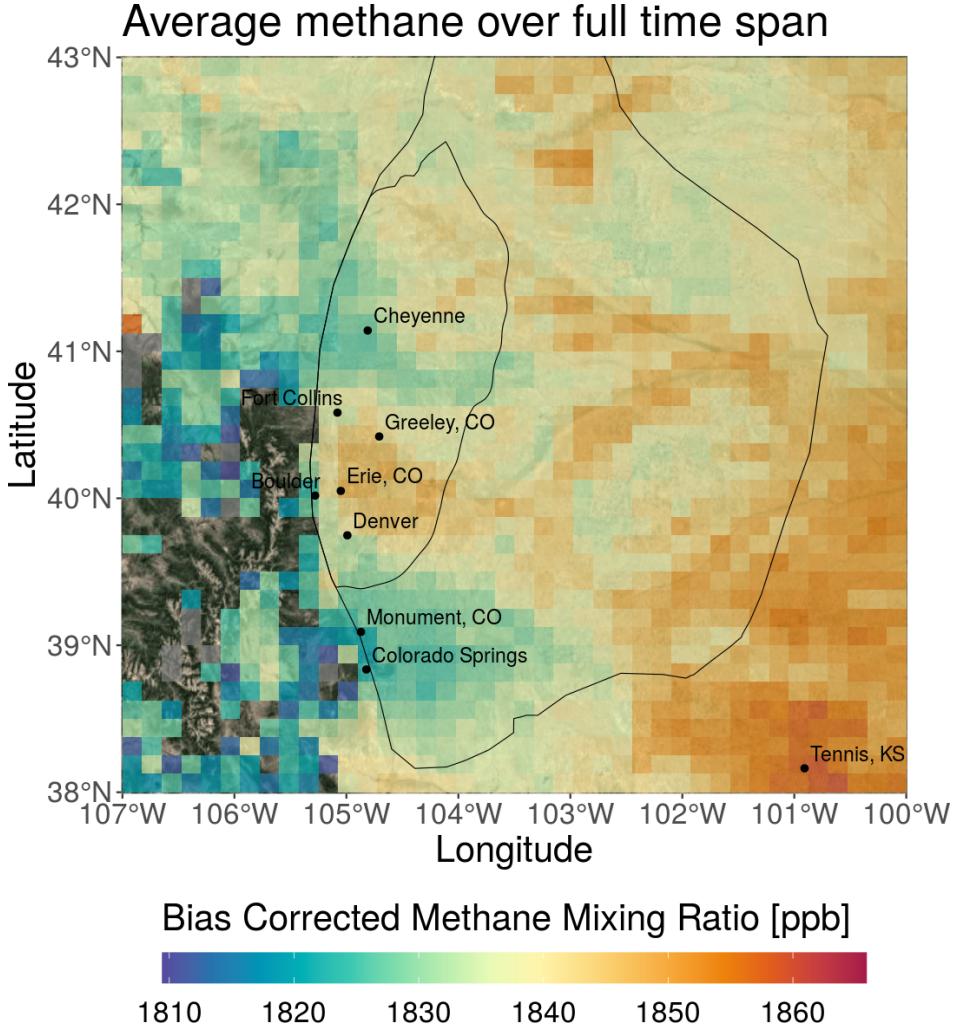

Figure 5: Average methane mixing ratio over the entire 22 month time span of the TROPOMI data. The outer boundary shows the DJ Basin, and the inner boundary shows the Niobara Play.

Figure 6 shows the standard deviation of the methane mixing ratio over the entire time span of the data. Methane concentrations in gridboxes with larger standard deviations fluctuate more than in gridboxes with smaller standard deviations.

We see that there are two areas of more variable data: one on the east side of the DJ basin and one northwest of Cheyenne. Again, the data is quite sparse over the mountains in the southwest corner of our region. The standard deviation in the bottom right corner of the region is relatively low, indicating that the higher concentrations of methane found here are relatively consistent. The standard deviation is also relatively low along the front range and in the Niobara Play, again indicating that methane concentrations are consistent in this area.
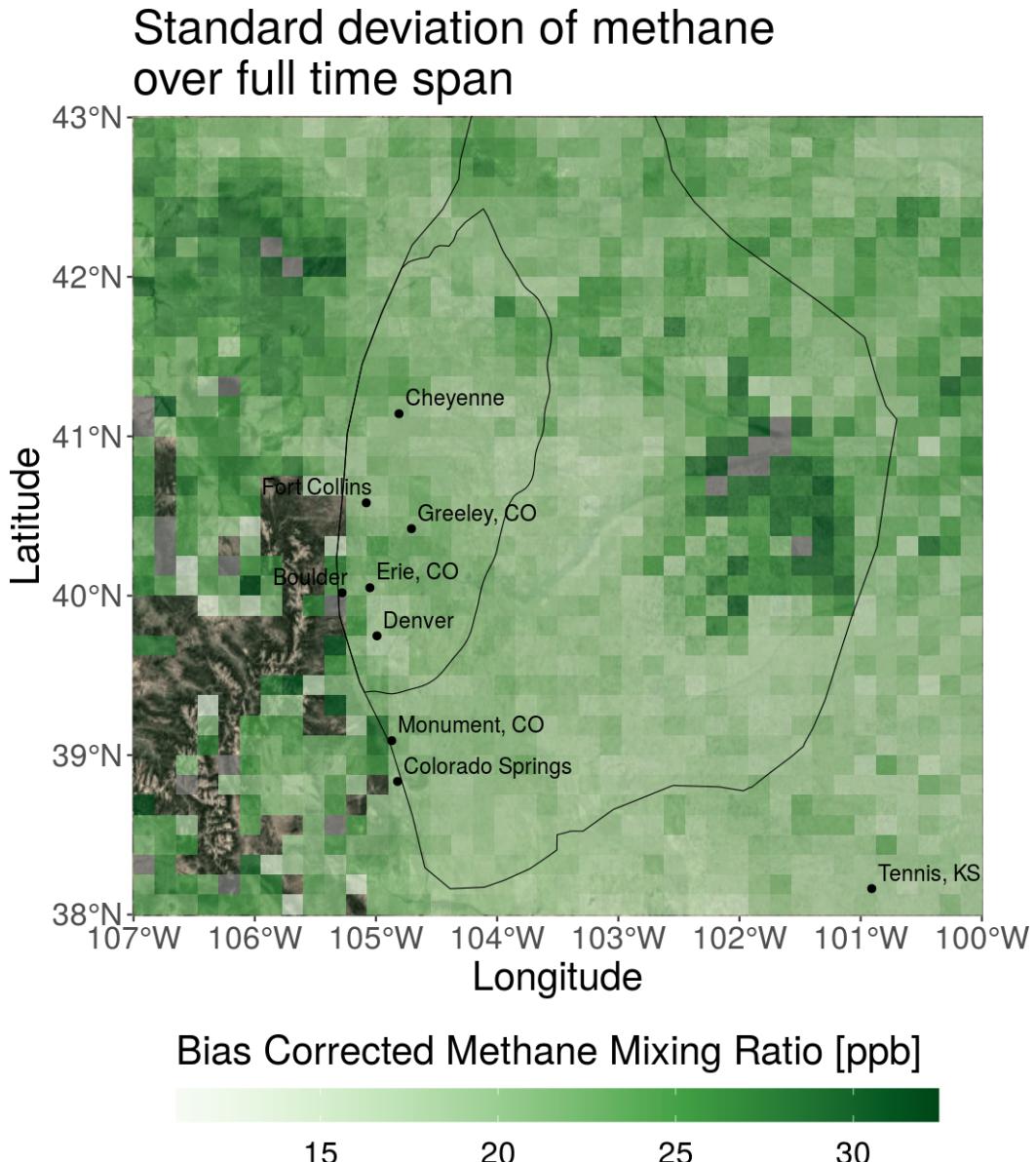
Figure 6: Standard deviation of methane mixing ratio over the entire 22 month time span of the TROPOMI data. The outer boundary shows the DJ Basin, and the inner boundary shows the Niobara Play.

Figure 7 shows the number of data points in each gridbox over the entire time span of the TROPOMI data. Large values correspond to areas with a large amount of available data, and low values correspond to areas with little available data. Statistical analysis is more reliable with more data, so results in areas with more observations are more reliable.

We again see that the mountainous region in the southwest contains very little data. Analysis here will be relatively unreliable. The rest of the region has a considerable amount of data. The areas south of Colorado Springs, east of Denver and Fort Collins, and northeast of Cheyenne have a large amount of available data in particular. A more detailed analysis of data sparsity can be found in Subsection 3.2.
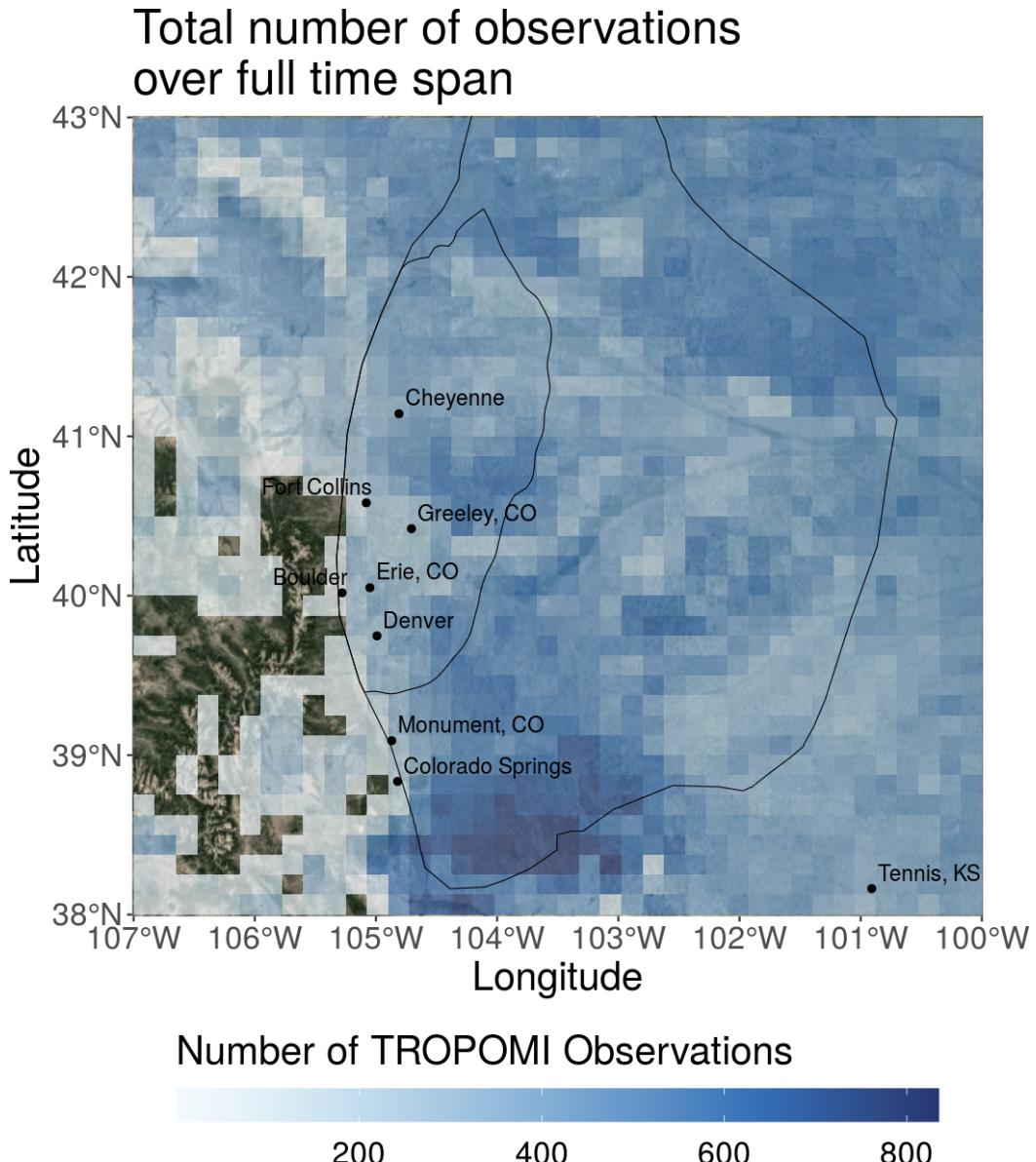
Figure 7: Number of data points available in each gridbox over the entire 22 month time span of the TROPOMI data. The outer boundary shows the DJ Basin, and the inner boundary shows the Niobara Play.

3.2 Smaller Time Scales

It is often informative to perform analysis on smaller time scales. Therefore, the TROPOMI data has been aggregated into the following temporal packages: gridded daily averages, gridded weekly averages, gridded monthly averages, and gridded seasonal averages. These products contain RES^2 gridbox averages for each time period. Note that the analysis in this section is performed using a 40×40 grid.

The sparsity of TROPOMI data varies across different time scales, as the instrument only passes over each geospatial point once or twice a day. Therefore, there are rarely more than 5 observations per gridbox when computing daily averages (using a 40×40 grid). A season average, on the other hand, will include data from dozens of TROPOMI overpasses. Figure 8 shows the expected number of observations per gridbox for each timescale. Specifically, it shows a histogram of counts per gridbox for each timescale, computed using a 40×40 grid. Note that gridboxes containing zero observations are excluded from these histograms.
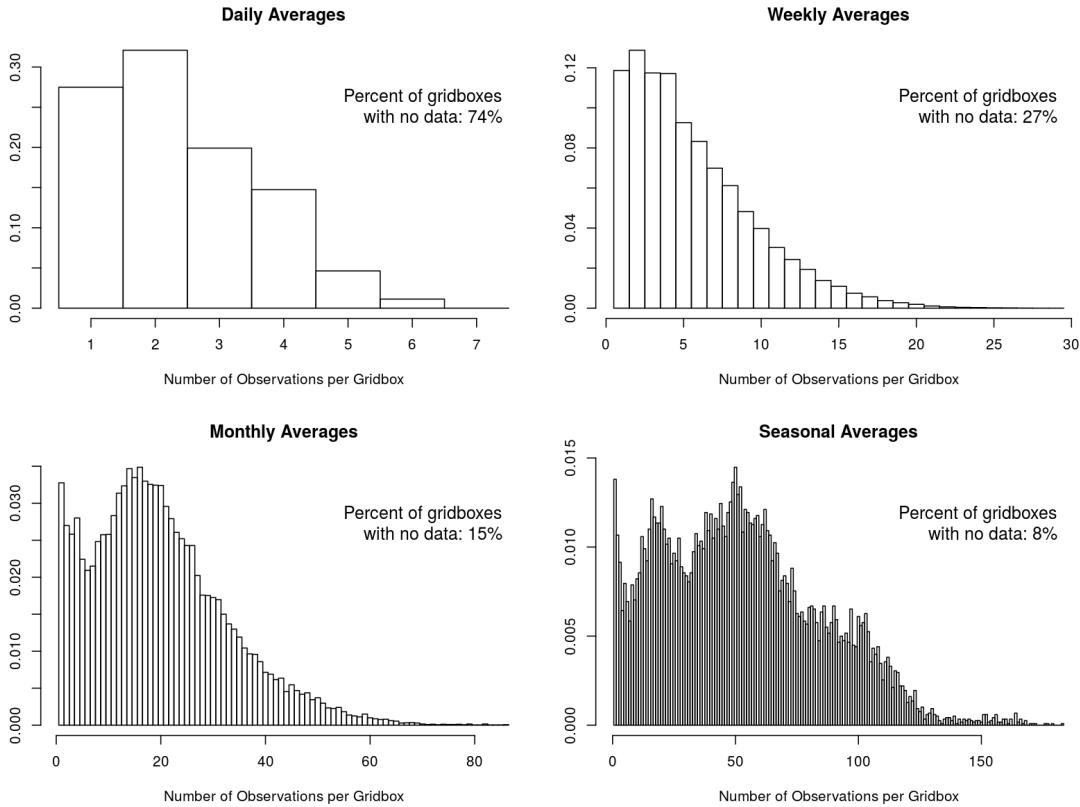
Figure 8: Histograms showing the relatively frequency of the number of observations per gridbox. These histograms were computed using a 40×40 grid size. Note that each subplot has a different scale on both the x-axis and y-axis. The percent of gridboxes containing no TROPOMI measurements is listed on each plot.

To get a sense of what the data looks like at these different timescales, sample plots have been included for each timescale. The grid size used in the following plots is 40×40. Plots in the left column use all available TROPOMI methane data. Plots in the right column use only the high quality data, after accounting for cloud cover, surface albedo, and other features that impair the retrieval algorithm. The dates included in each plot are listed as row labels.

The Payne Institute
for Public Policy
MINES

3.2.1 Daily Plots

On a daily timescale, the TROPOMI data is quite sparse. If a gridbox includes any data, there are most likely to be around 2 observations. Two examples are provided below. Figures 9 and 10 both plot one week's worth of daily averages. Figure 9 is an example of an especially sparse week, and Figure 10 is an example of an especially abundant week. Blank spots in the plot grid indicate that there was no data for that particular time period.
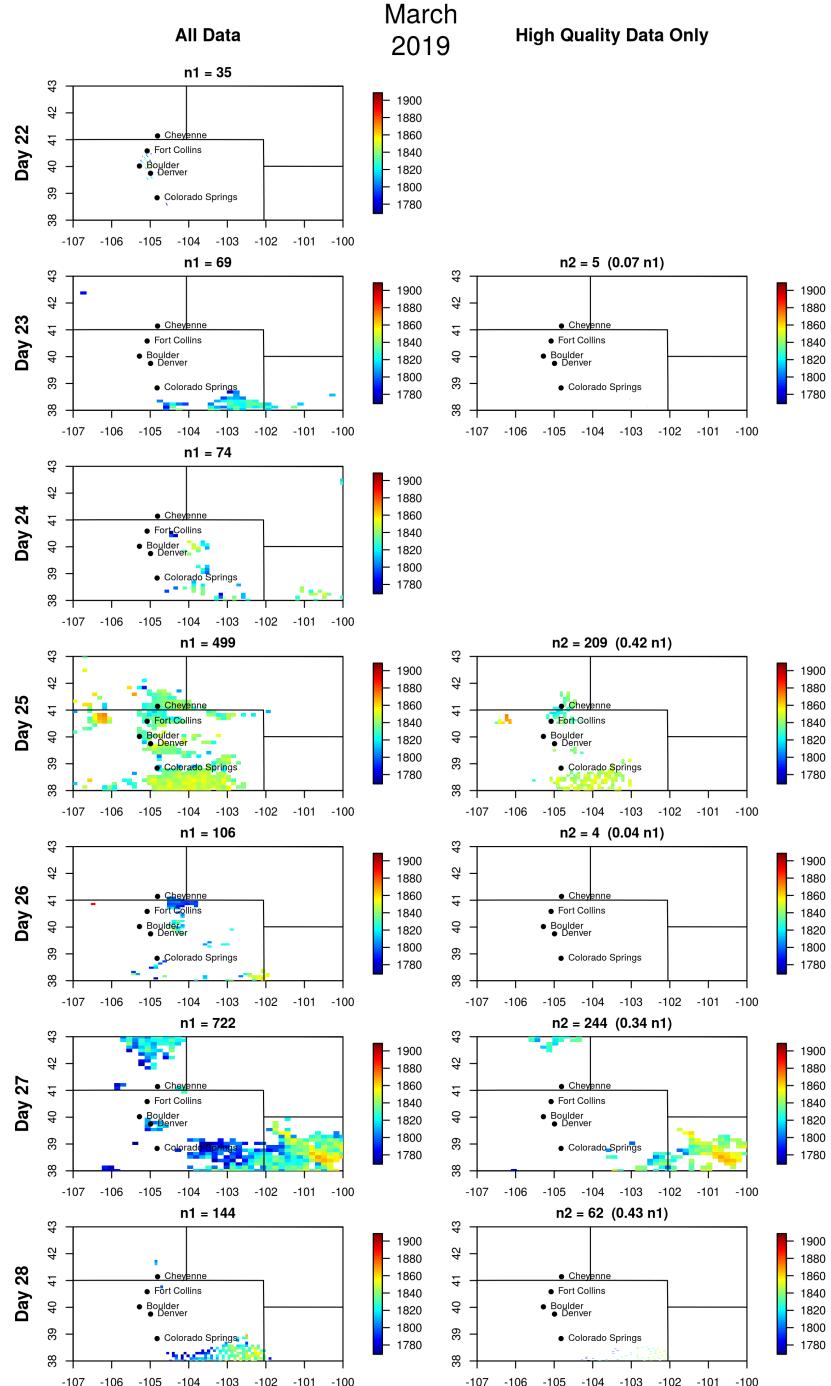

Figure 9: An example of a week with a relatively low amount of data. There are 1,649 total observations, 524 of which are high quality.
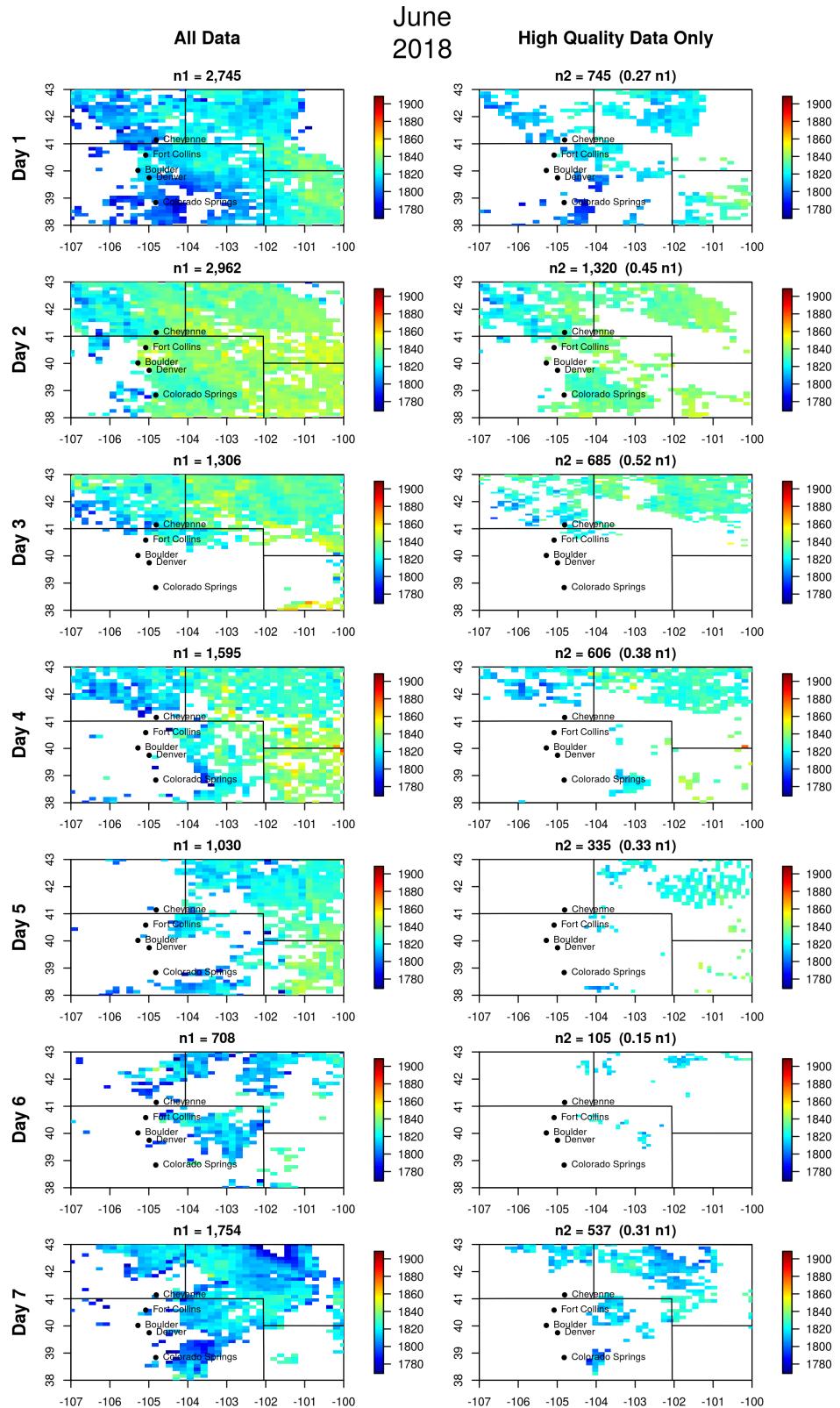
Figure 10: An example of a week with a relatively high amount of data. There are 12,100 total observations, 4,333 of which are high quality.

3.2.2 Weekly Plots

Week averaged data is still quite sparse. If a gridbox includes any data, there are most likely to be around 2 to 5 observations. However, it is possible to see gridboxes with up to 15 observations unlike the daily averages. Two examples are plotted below. Figure 11 shows a month of weekly averages with especially sparse conditions. Figure 12 shows a month of weekly averages with especially abundant conditions. Note that the last row of plots in Figures 11 and 12 contain more than 7 days of data, so that a full month could be represented.
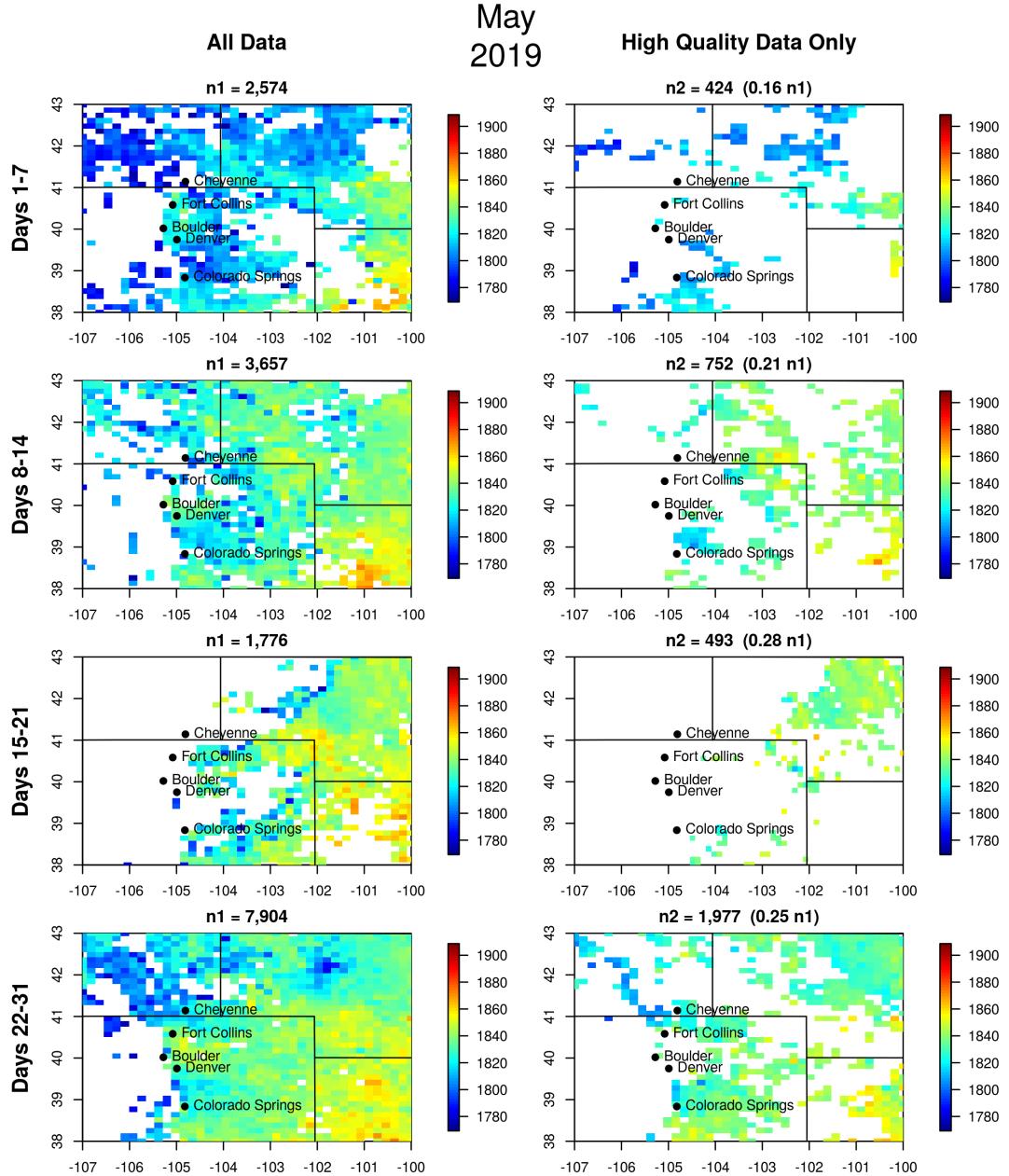
Figure 11: An example of a month with a relatively low amount of data. There are 15,911 total observations, 3,646 of which are high quality.
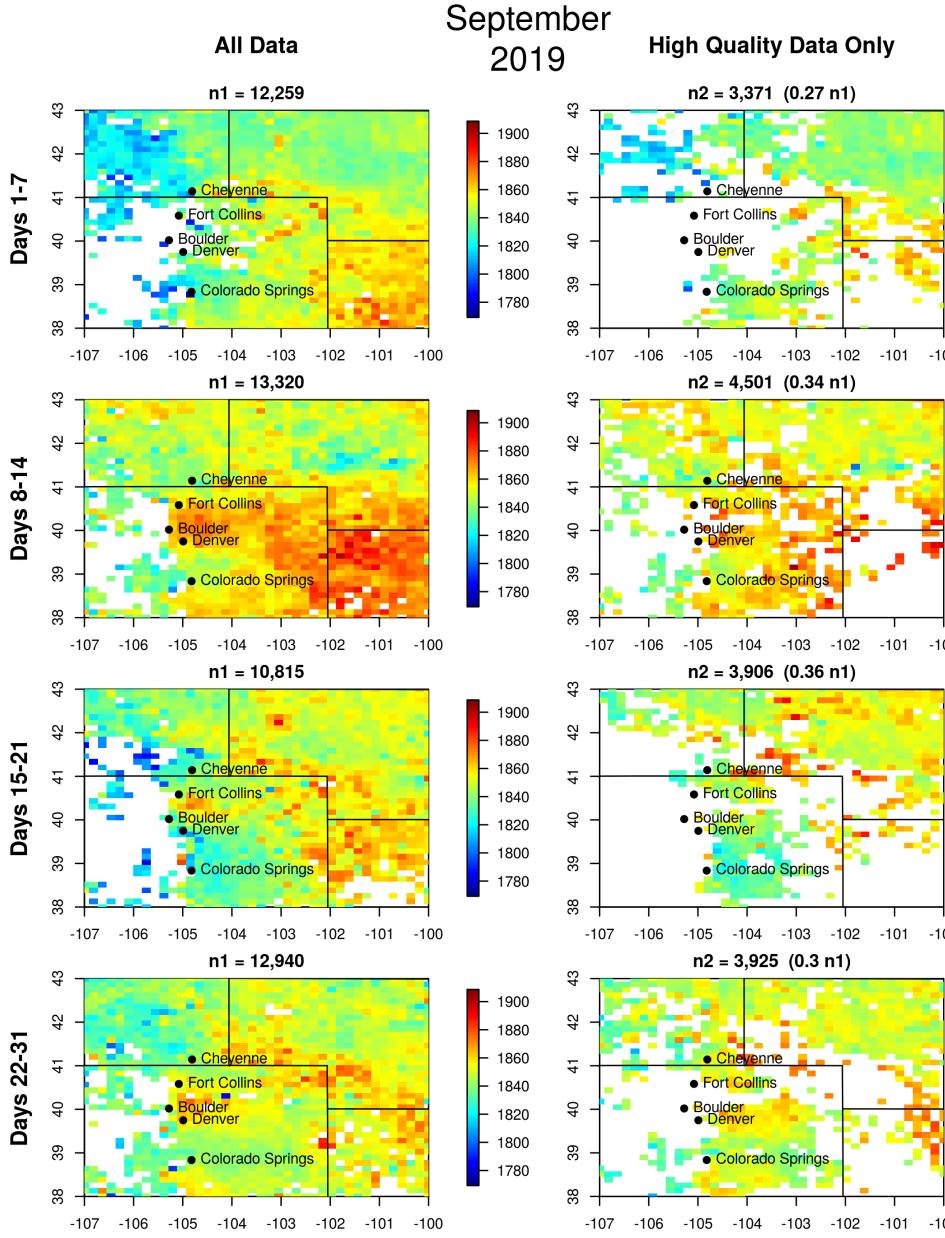
Figure 12: An example of a month with a relatively high amount of data. There are 49,334 total observations, 15,703 of which are high quality.

3.2.3 Monthly and Seasonal Plots

On the monthly and seasonal timescales, we begin to see less data scarcity. If a gridbox includes any data, there are most likely to be around 18 observations when using monthly averages and 50 observations when using seasonal averages. It is possible to see gridboxes with up to 60 observations when using monthly averages and up to 125 observations when using seasonal averages. Two examples are plotted below. Figure 13 shows three monthly averages with especially sparse conditions. This figure also includes an average over the entire season. Figure 14 shows three monthly averages with especially abundant conditions. This figure also includes an average over the entire season.

Note that the first three rows of plots in Figures 13 and 14 show the monthly averages, and the fourth row shows the seasonal average.
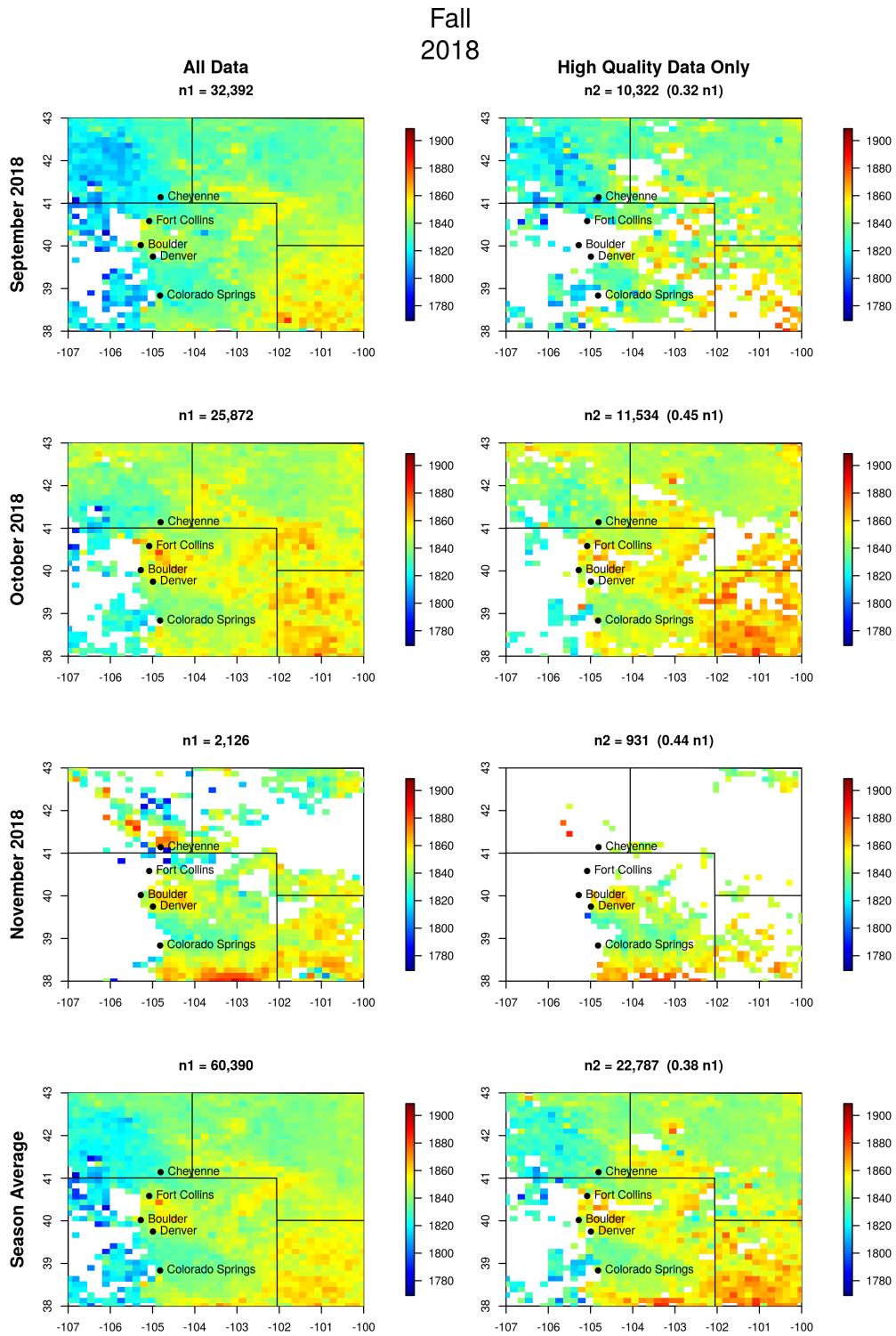
Figure 13: An example of a season with a relatively low amount of data. There are 60,390 total observations in the season, 22,787 of which are high quality.
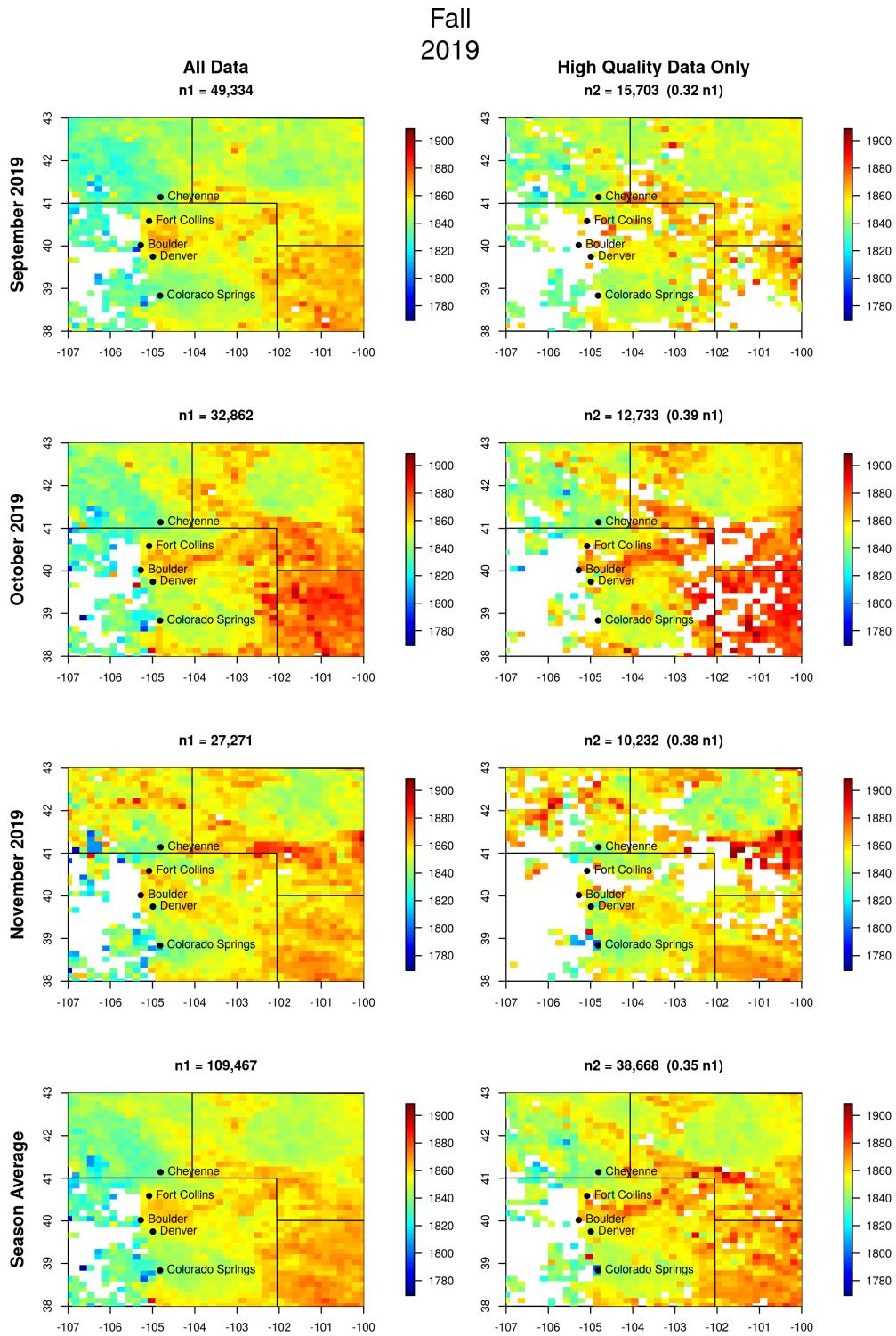
Figure 14: An example of a season with a relatively high amount of data. There are 109,467 total observations in the season, 38,668 of which are high quality.

3.2.4 Additional Spatial Plot

Additional spatial plots can be found in the appendix. For brevity, all of the TROPOMI data is not plotted at each timescale. The following plots are included for each averaging timescale. These plots are simply meant to give an idea of what the TROPOMI data looks like.

- Daily averages: January 1, 2019 - February 28, 2019.

- Weekly averages: March 1, 2019 - December 31, 2019.

- Monthly and seasonal averages: April 2018 - February 2020 (all TROPOMI data).

4 Time Series Plots

To get a better sense of the TROPOMI data over time, we have created a sample of time series plots. These plots show methane concentrations in five particular locations over the entire time span of the TROPOMI data, from May 2018 through February 2020. The five locations we have selected are:

1. Tennis, KS

2. Erie, CO

3. Boulder, CO

4. Greeley, CO

5. Monument, CO

These locations are highlighted in green in Figure 15. Tennis, KS is intended to show the high methane concentrations we observe in the southeast corner of our region over time. Monument, CO is intended to show a region with consistently low methane concentrations over time. The remaining three locations are intended to show methane concentrations in the Niobara Play over time.

To create these time series, we first determine which gridboxes contain the five locations listed above. We then plot the methane concentration in these gridboxes over time, creating separate figures for each different timescale. However, note that the daily timescale is not used for these time series plots. This is because any given gridbox has a high likelihood of containing no TROPOMI data when working with daily averages. Specifically, about 3 out of 4 gridboxes tend to contain no data on the daily timescale. Therefore, plotting daily methane concentrations results in many large gaps of 4 to 5 days with no data. Additionally, daily methane concentrations fluctuate greatly, which introduces a high level of noise into the daily time series plots. These two factors make the time series plot on a daily basis nearly uninterpretable.
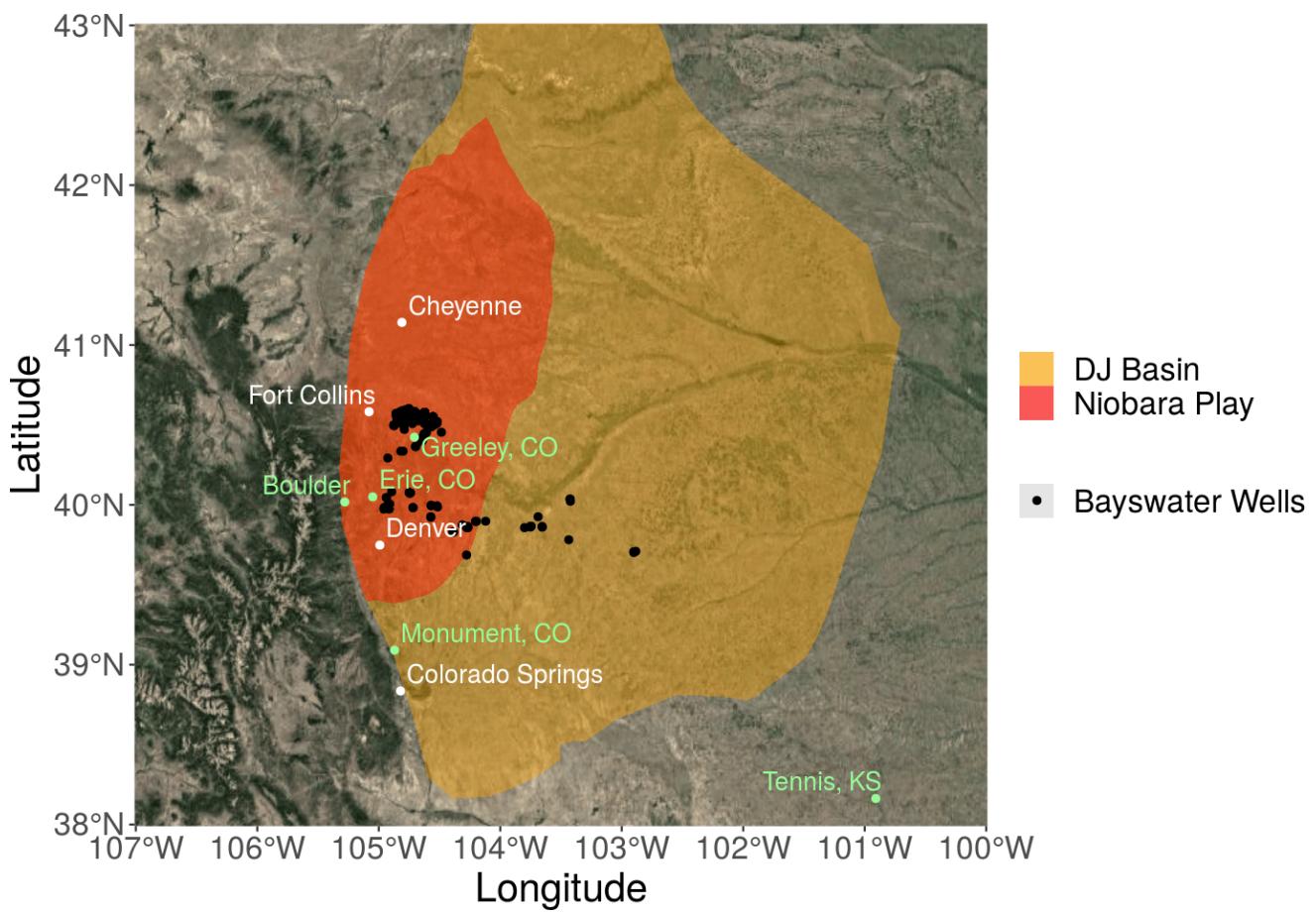
Figure 15: Our region of interest. Locations used in the time series analysis are highlighted in green.

4.1 Weekly Plots

Figures 16 and 17 show the time series using weekly averages. The weekly time series is quite noisy, as each weekly average contains a relatively small number of TROPOMI observations. Therefore, we have applied a Gaussian smoothing kernel to make the time series more interpretable, shown in Figure 17. This smoothing kernel averages each data point with the two values preceding it and the two values following it. Gaussian distributed weights are applied to the average so that the data point itself has the most influence on the average and values further from the data point have less influence. These weights make it so that the actual TROPOMI data is smoothed but not drastically altered by the surrounding data.

Note that there are significant gaps in the weekly time series, resulting from periods with no TROPOMI observations. However, we do begin to see a seasonal trend appear in the weekly time series. There appears to be higher methane concentrations in the fall and winter seasons, and lower methane concentrations in the spring and summer. This seasonal trend is more apparent in Figure 17.
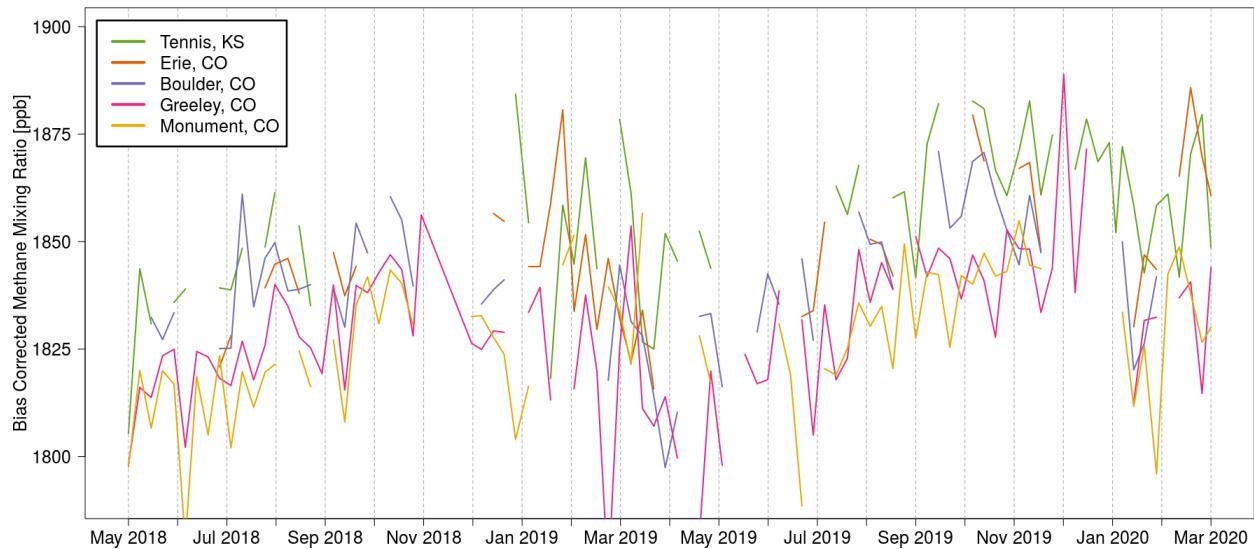
Figure 16: Time series using weekly averages.
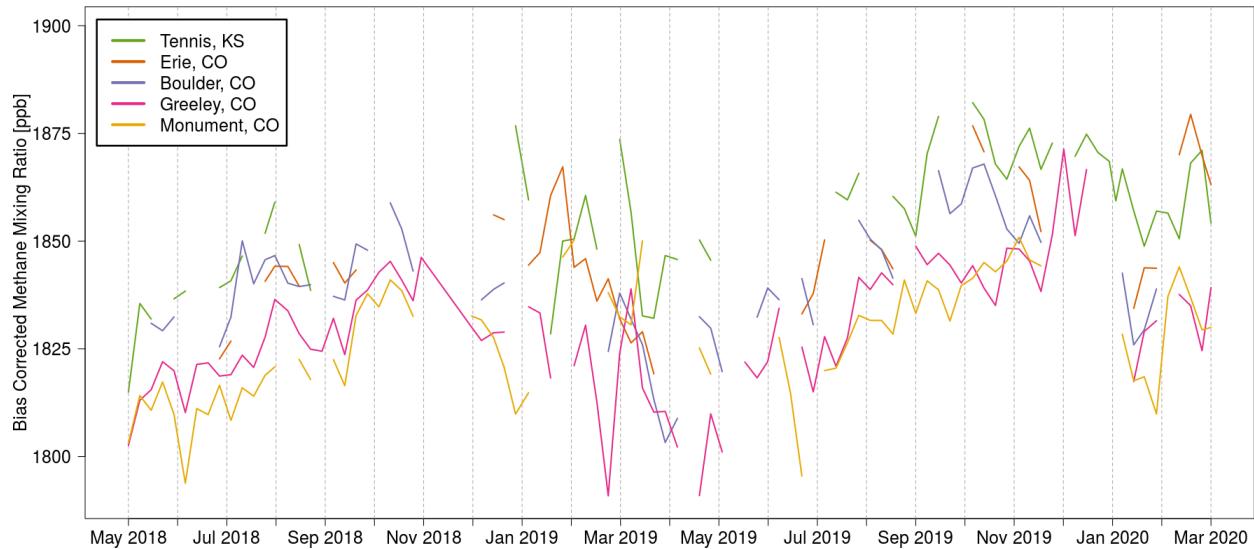
Figure 17: Time series using weekly averages with a Gaussian smoothing kernel applied.

19

4.2 Monthly Plots

Figures 18 through 20 show the time series using monthly averages. The data are less noisy than before, as each monthly average contains more TROPOMI observations. This makes it easier to see the seasonal trend discussed earlier. Like the weekly time series, however, we also apply a Gaussian smoothing kernel to the monthly time series. Figures 18 and 19 show the raw time series, and Figure 20 shows the smoothed time series.

Two types of plots are presented here: one with error bars and one with lines connecting the data points. The error bars show the uncertainty around the estimate of the mean, and the lines make it easier to see the seasonal trend. Note that the horizontal offsets in the error bar plots are for visual clarity and do not indicate different times within the month.

The error bars show two standard errors of the mean at each month-averaged data point. Standard error of the mean is defined as follows:

$$\mathrm{SE}(\bar{x}) = \frac{\sigma}{\sqrt{n}},$$

where σ is the population standard deviation and n is the sample size. Put in terms of this methane application, σ is the true deviation of atmospheric methane from its mean, and n is the number of recorded TROPOMI observations. Since we do not know the true deviation of atmospheric methane, we estimate it with the deviation of the methane in our sample of TROPOMI observations. This estimate, also referred to as $\mathrm{SE}(\bar{x})$, is defined as follows:

$$\mathrm{SE}(\bar{x}) = \frac{s}{\sqrt{n}},$$

where s is the sample standard deviation. Put in terms of this methane application, s is the deviation of TROPOMI methane observations from their mean. Note that in these plots we are interested in the monthly standard error in five different gridboxes. Therefore, s is really the deviation of TROPOMI methane observations falling within a given gridbox over the course of a given month. Similarly, n is the number of TROPOMI observations falling within the same gridbox during the same month. Therefore, a large error bar indicates that either: (1) there are not many observations falling within a given gridbox during a given month, or (2) the observations within that gridbox are highly variable over the course of the month. These error bars are not included on the weekly time series, as the weekly data is too noisy for the bars to be easily interpreted.

It is important to note that these error bars are likely an underestimate of the true standard error and can be considered a lower bound on the error estimate. This is because the standard error calculation relies on the assumption of independent observations. Our data violates this assumption, as there is obvious temporal correlation. Put simply, if a gridbox has a high methane concentration for a given month, it is more likely that the same gridbox will also have a high methane concentration for the following month. Violating the assumption of independence means that our error bars tend to underestimate the true standard error. However, they are still included to give a sense of relative error in our data.

Finally, the seasonal trend discussed earlier is more apparent in the monthly time series. We see that fall and winter months tend to have higher methane concentrations than spring and summer months. As before, this seasonal trend is more apparent in the smoothed time series shown in Figure 20.

*The* Payne Institute *for* Public Policy
MINES
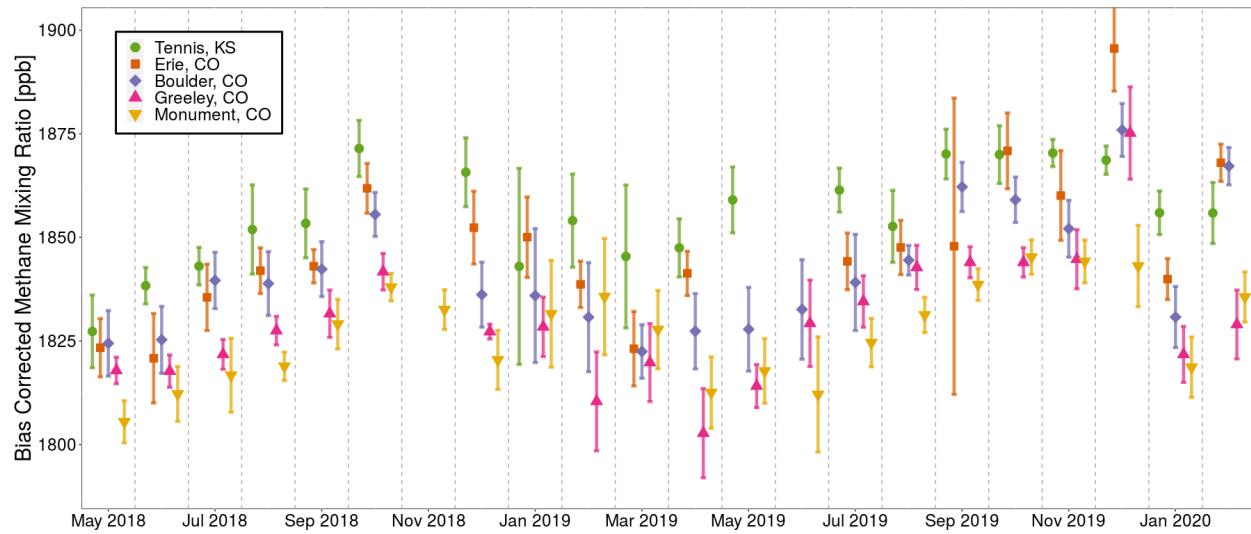
Figure 18: Time series using monthly averages. Error bars show two standard errors.
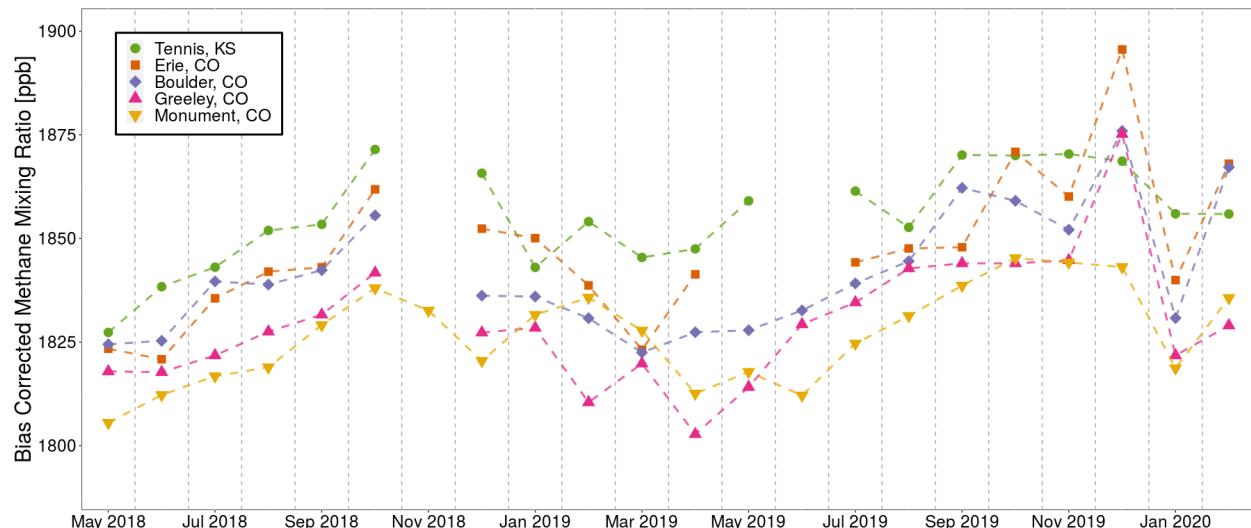
Figure 19: Time series using monthly averages. Dashed lines connect data points for visual clarity.
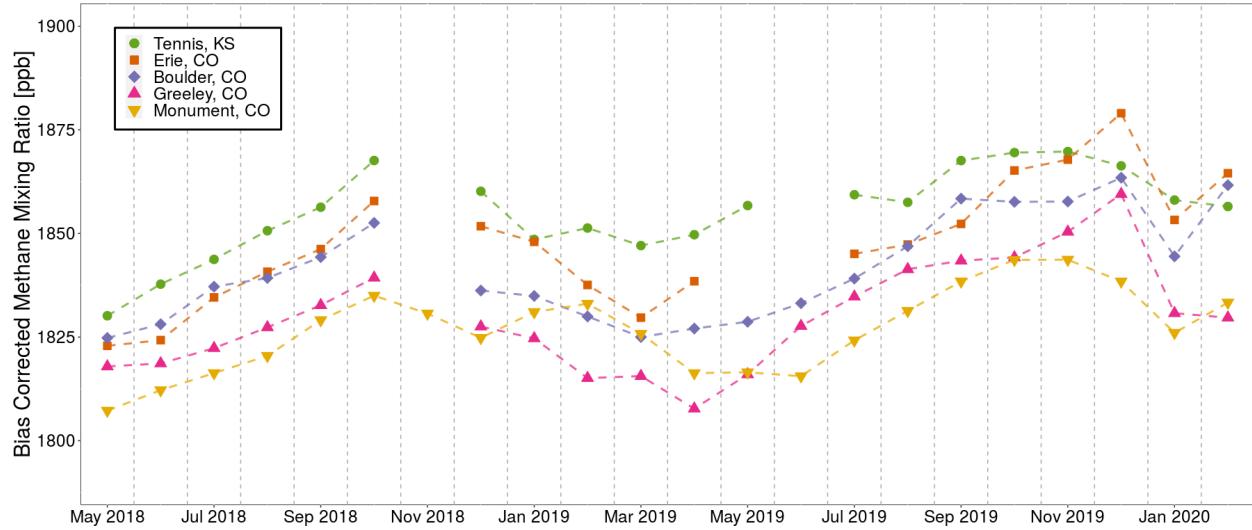
Figure 20: Time series using monthly averages with a Gaussian smoothing kernel applied. Dashed lines connect data points for visual clarity.

4.3 Seasonal Plots

Figures 21 and 22 show the time series using seasonal averages. Figure 21 shows the standard error bars discussed earlier, and Figure 22 includes connecting lines. No smoothing kernel is applied to the seasonal time series, as the seasonal averages are already quite smooth. Also note that the spring 2018 season only contains data from May 2018, as this is when TROPOMI data became publicly available.

The seasonal data are less noisy than the monthly and weekly data, as each seasonal average contains more TROPOMI observations. As a result, the seasonal trend discussed earlier is quite apparent here. The fall and winter seasons typically have larger methane concentrations than the spring and summer seasons. There also appears to be an overall upward linear trend in the methane time series. However, the series is too short to quantify this trend reliably.

Finally, it is much easier to differentiate between the five locations when using the seasonal average. Tennis has uniformly higher methane concentrations than the other locations, which supports the spatial plots shown earlier. These plots show consistently higher concentrations of methane in the southeast corner of our region of interest. Monument typically has lower methane concentrations than the other locations. The three locations within the Niobara Play largely fall between Tennis and Monument with the exception of Greeley, which occasionally falls below Monument.
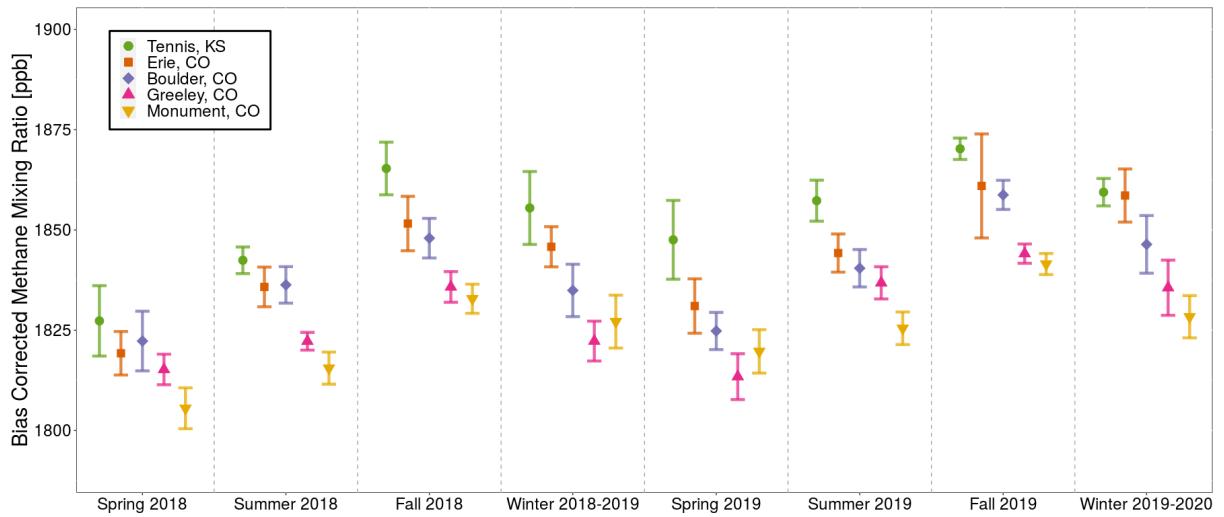
Figure 21: Time series using seasonal averages. Error bars show two standard errors.
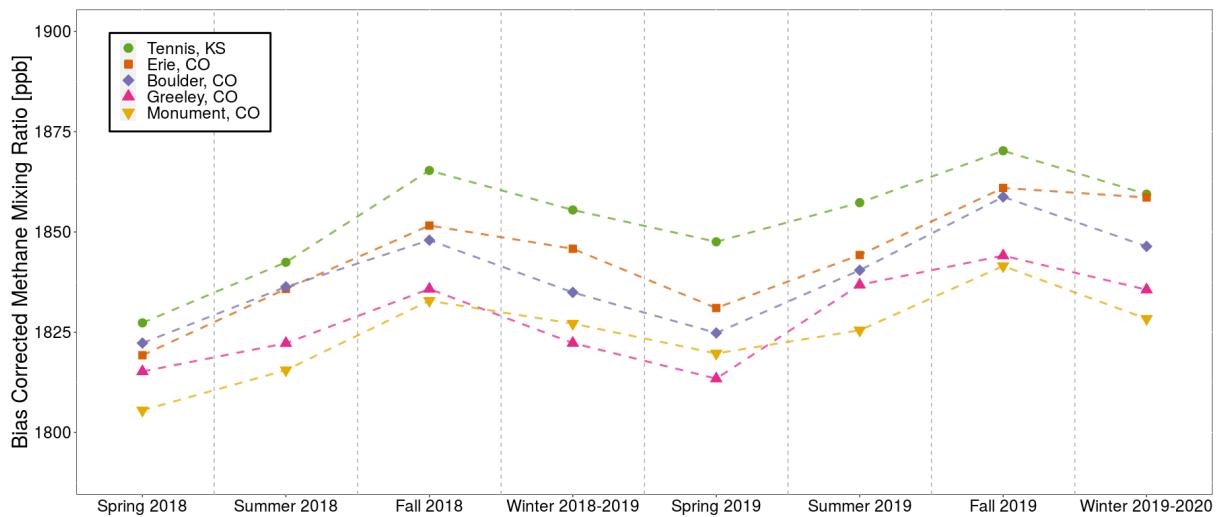
Figure 22: Time series using seasonal averages. Dashed lines connect data points for visual clarity.

The Payne Institute
for Public Policy

5 Data Analysis

5.1 Locations Consistently above the Regional Average

One question we might ask is: how often is each gridbox above the regional average? Looking for gridboxes that are consistently higher than average might help spot recurring areas of high methane emissions. The following plots attempt to answer this question.

We have again broken this analysis up into four different temporal scales: daily, weekly, monthly, and seasonal. On each time scale, we count the number of times a given gridbox average is above the regional average and then convert this number to a proportion. Take, for instance, the daily timescale. For a given day, we compute the following quantities: (1) the average methane value over our entire region of interest, and (2) the average methane value for each gridbox. We then check which gridbox averages are above the regional average on that given day. We do this for each day and count the total number of times each gridbox falls above the regional average. This number is then converted to a proportion and plotted.

Figures 23 through 26 show these proportions for the four different timescales.
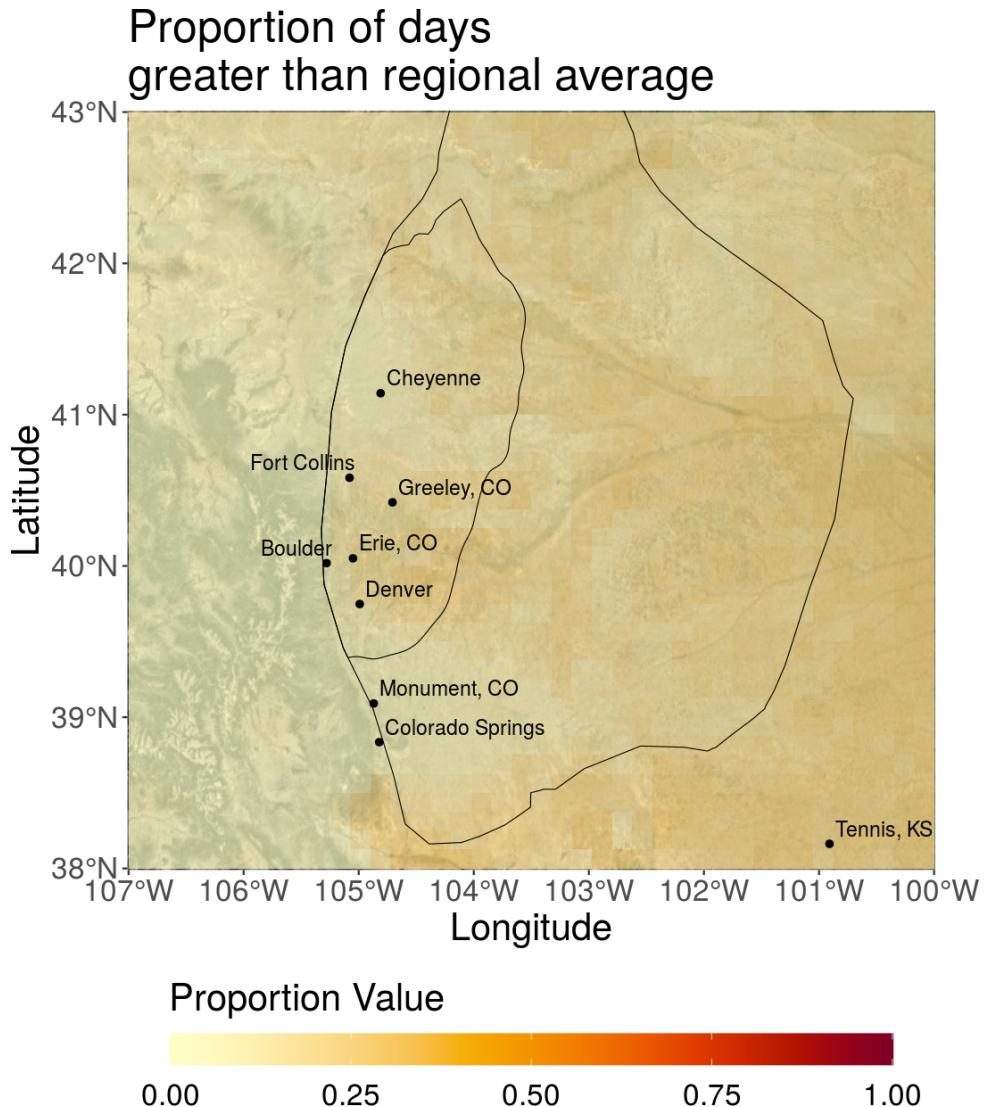
Figure 23: Proportion of days greater than the daily regional average, broken up by gridbox. The outer boundary shows the DJ Basin, and the inner boundary shows the Niobara Play.
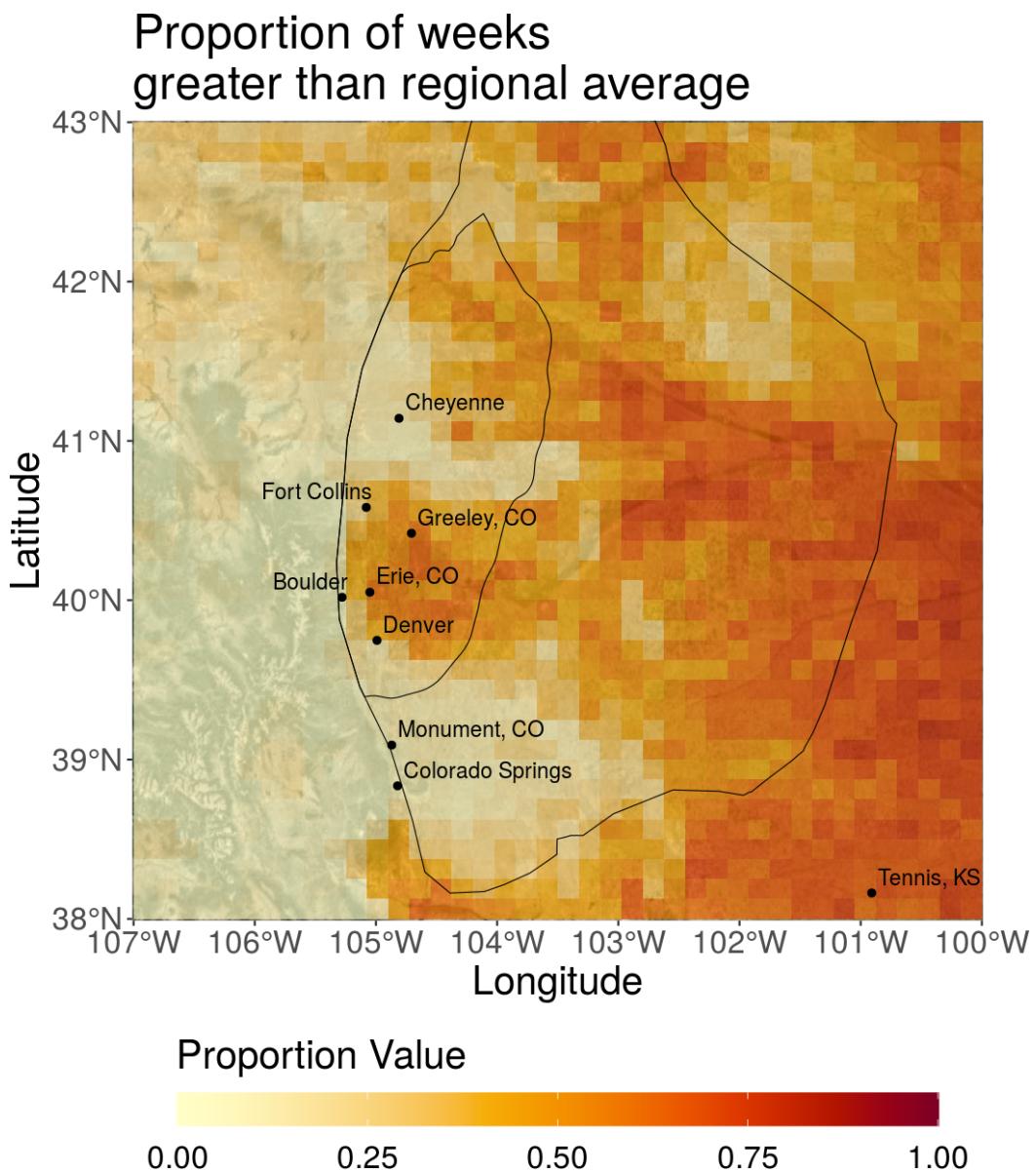
Figure 24: Proportion of weeks greater than the weekly regional average, broken up by gridbox. The outer boundary shows the DJ Basin, and the inner boundary shows the Niobara Play.
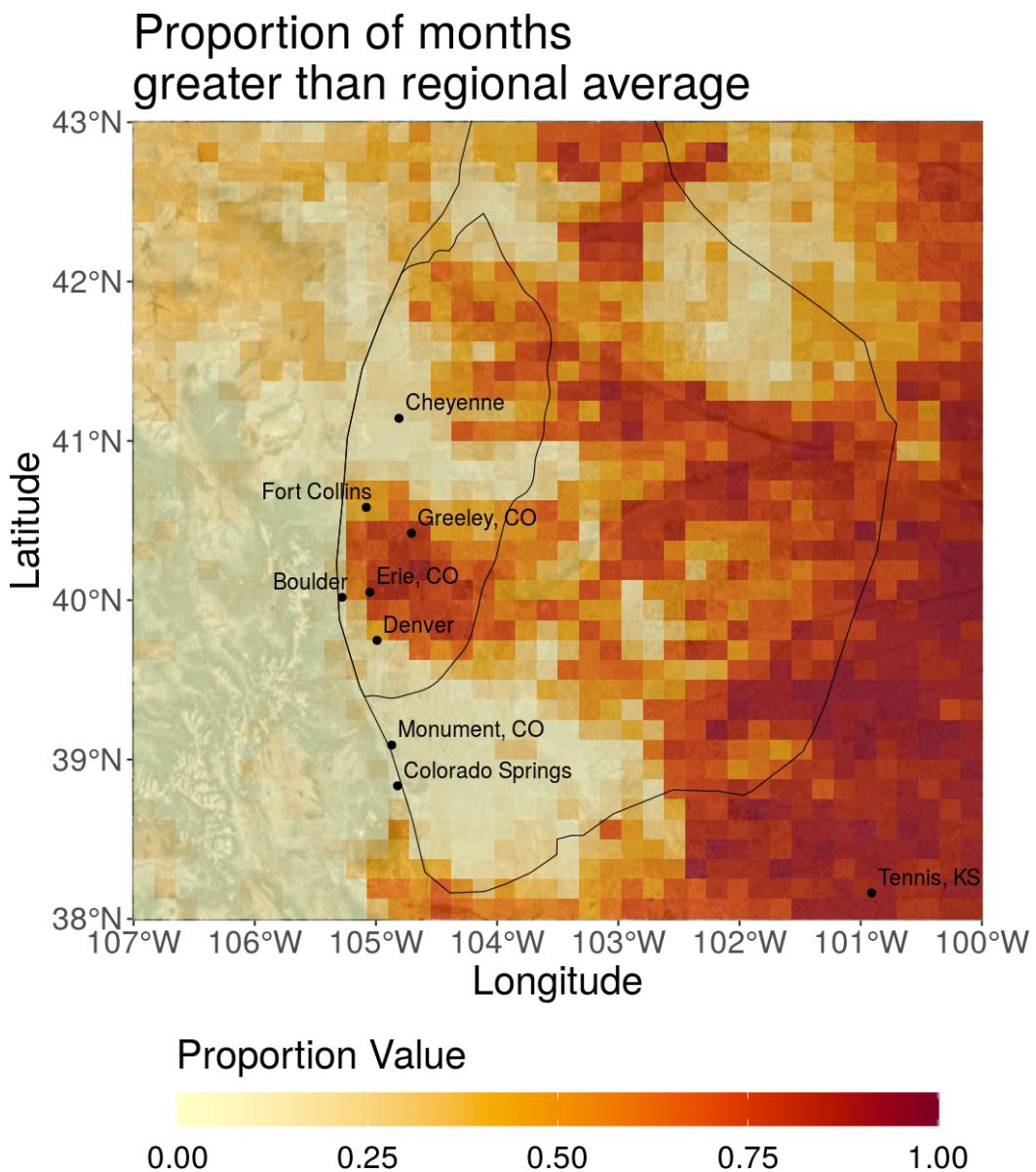
Figure 25: Proportion of months greater than the monthly regional average, broken up by gridbox. The outer boundary shows the DJ Basin, and the inner boundary shows the Niobara Play.
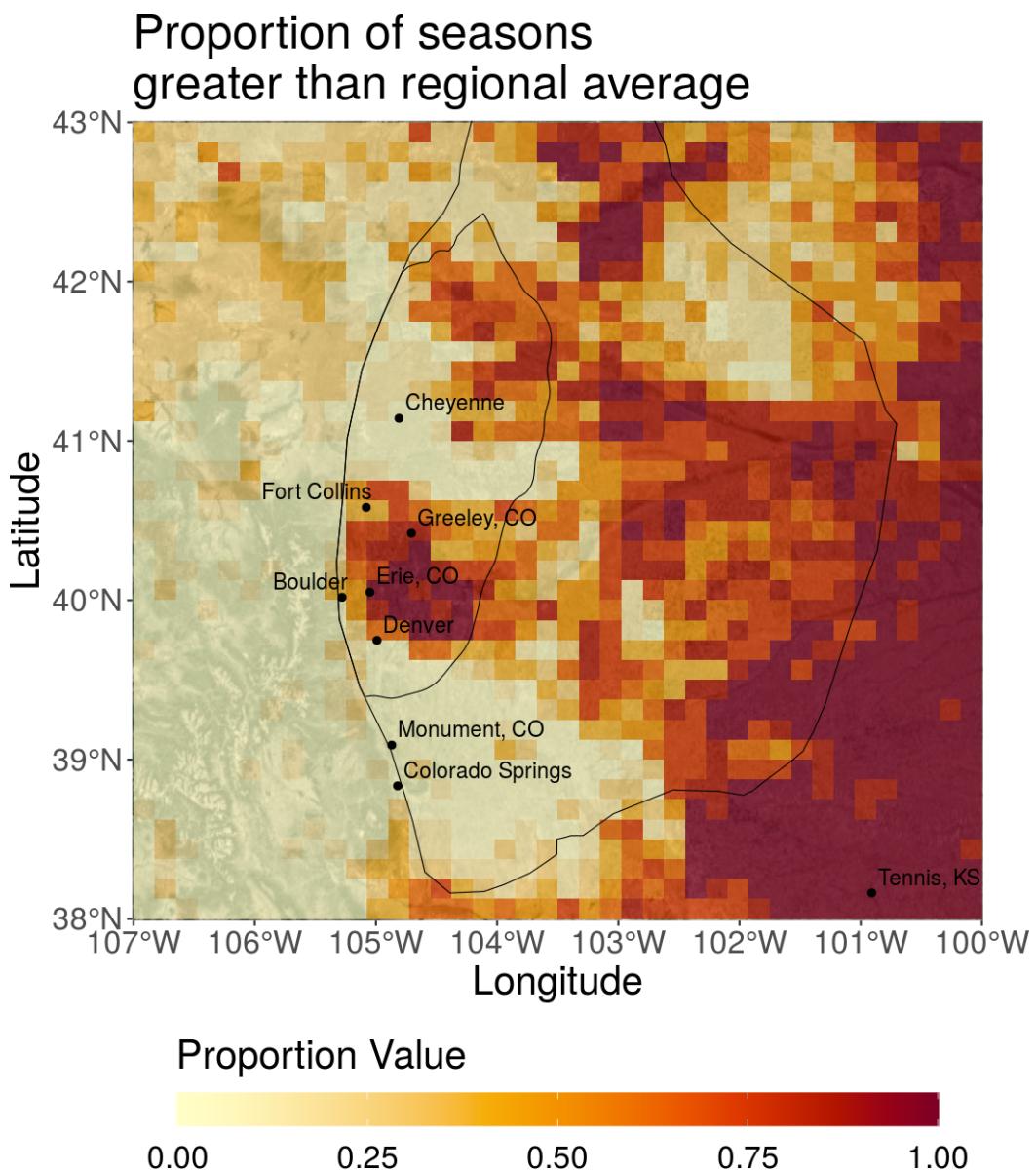
Figure 26: Proportion of seasons greater than the seasonal regional average, broken up by gridbox. The outer boundary shows the DJ Basin, and the inner boundary shows the Niobara Play.

First, it is important to note that the proportion values over the mountains in the southwest corner of the region are unreliable. This is because there are rarely any reliable data collected in this area, as seen in the earlier spatial plots.

These proportions tell us that the gridbox averages on the eastern side of the region are often higher in methane than the regional average. This could potentially be due to higher background methane concentrations in Oklahoma, Kansas, and Nebraska. The gridboxes north of Denver, east of Boulder, and south of Fort Collins are also frequently above the regional average. This could be due to agriculture or oil and gas developments in this area.

Moving from daily to seasonal timescales increases the proportion above average for many gridboxes. This makes sense because the larger time scales have much more data to average. Take, for instance, the proportion above average on the daily timescale. Methane concentrations fluctuate day-to-day much more than they do month-to-month or season-to-season. Therefore, it is unlikely that a given gridbox will be above the regional average every single day, regardless of whether it is typically a large methane emitter. However, we have much more data to average on the seasonal timescale. Therefore, it is fairly likely that a given gridbox will be above the seasonal average every single season if it is typically a large methane emitter. This is because small day-to-day fluctuations get averaged out when considering an entire season of data.

5.2 Locations Containing Consistently High Methane Concentrations

Similarly, we might be interested in asking: how often is the methane concentration in each gridbox among the top 25% in the region? This analysis could also help spot recurring areas of high methane emissions. Particularly, it could reveal gridboxes with the most intense emissions.

We have again broken this analysis up into four different temporal scales: daily, weekly, monthly, and seasonal. On each timescale, we count the number of times the average methane in a given gridbox is within the top 25% in the region. We then convert this number to a proportion. Take, for instance, the monthly timescale. For a given month, we compute the average methane value in each gridbox. We then check which gridboxes are among the top 25% based on their average methane value for that given month. We then do this for each month and count the total number of times each gridbox is among the top 25%. This number is then converted to a proportion and plotted.

Figures 27 through 30 show these proportions for the four different timescales.
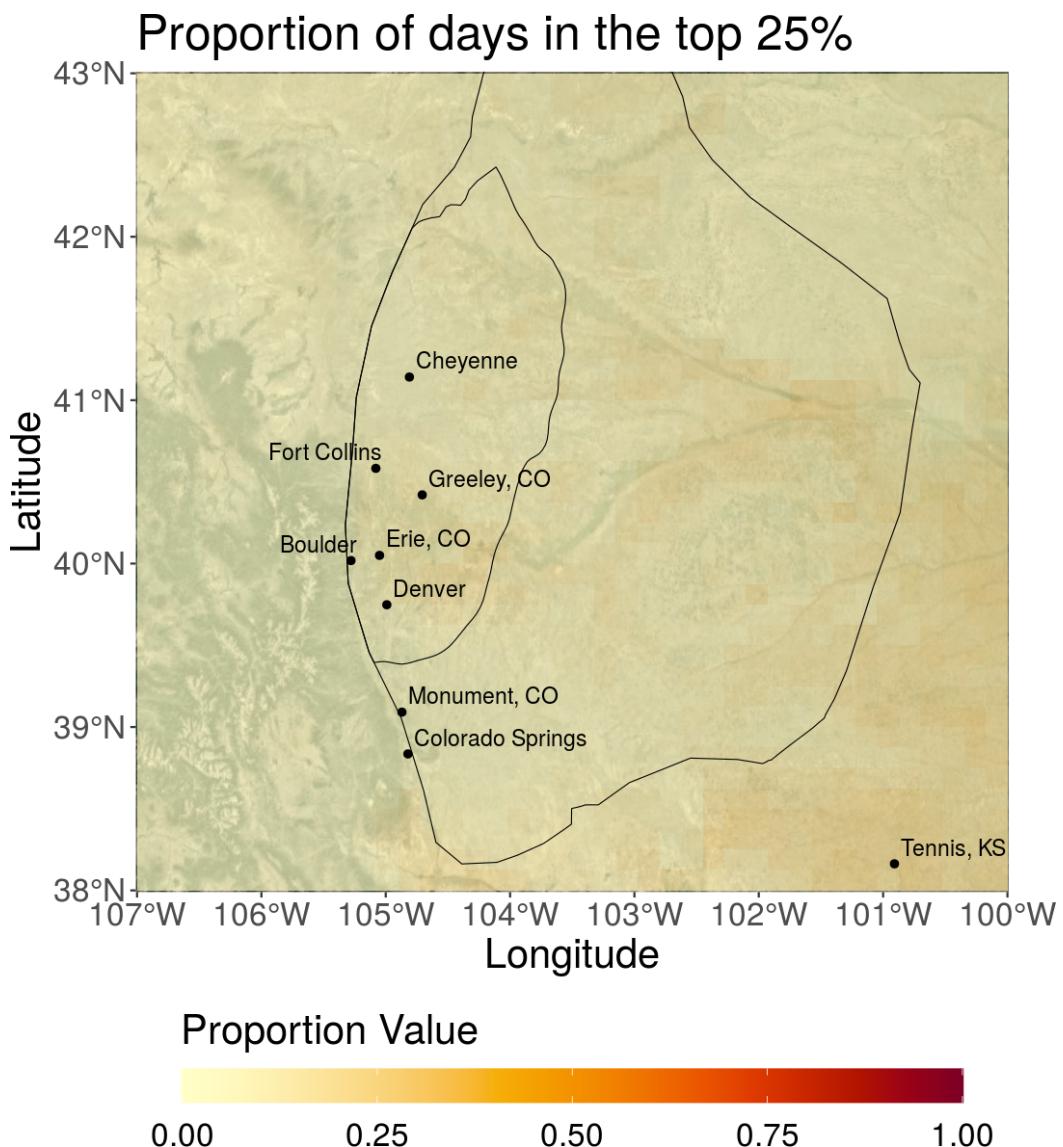
Figure 27: Proportion of days in the top 25%, broken up by gridbox. The outer boundary shows the DJ Basin, and the inner boundary shows the Niobara Play.
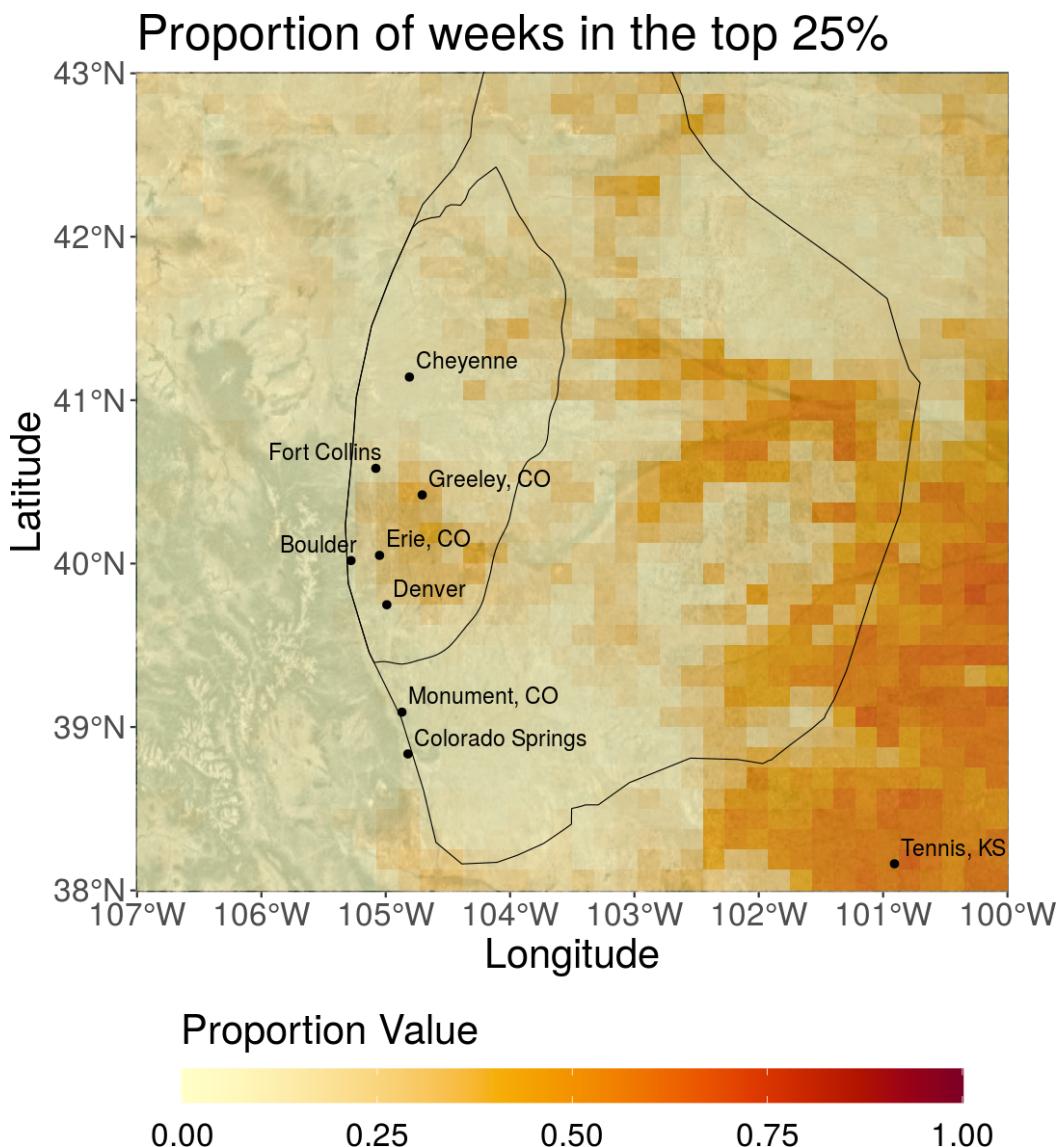
Figure 28: Proportion of weeks in the top 25%, broken up by gridbox. The outer boundary shows the DJ Basin, and the inner boundary shows the Niobara Play.
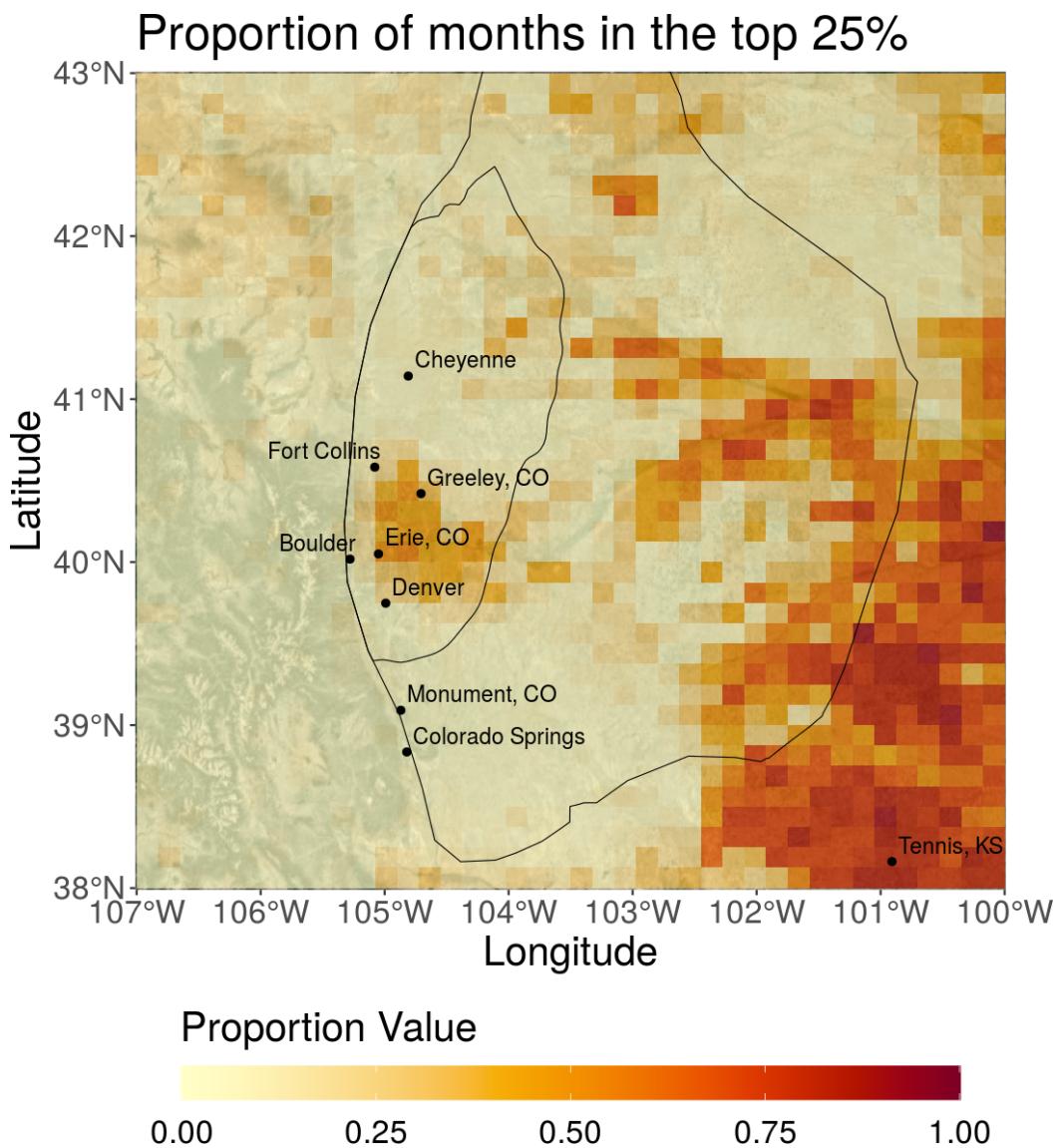
Figure 29: Proportion of months in the top 25%, broken up by gridbox. The outer boundary shows the DJ Basin, and the inner boundary shows the Niobara Play.
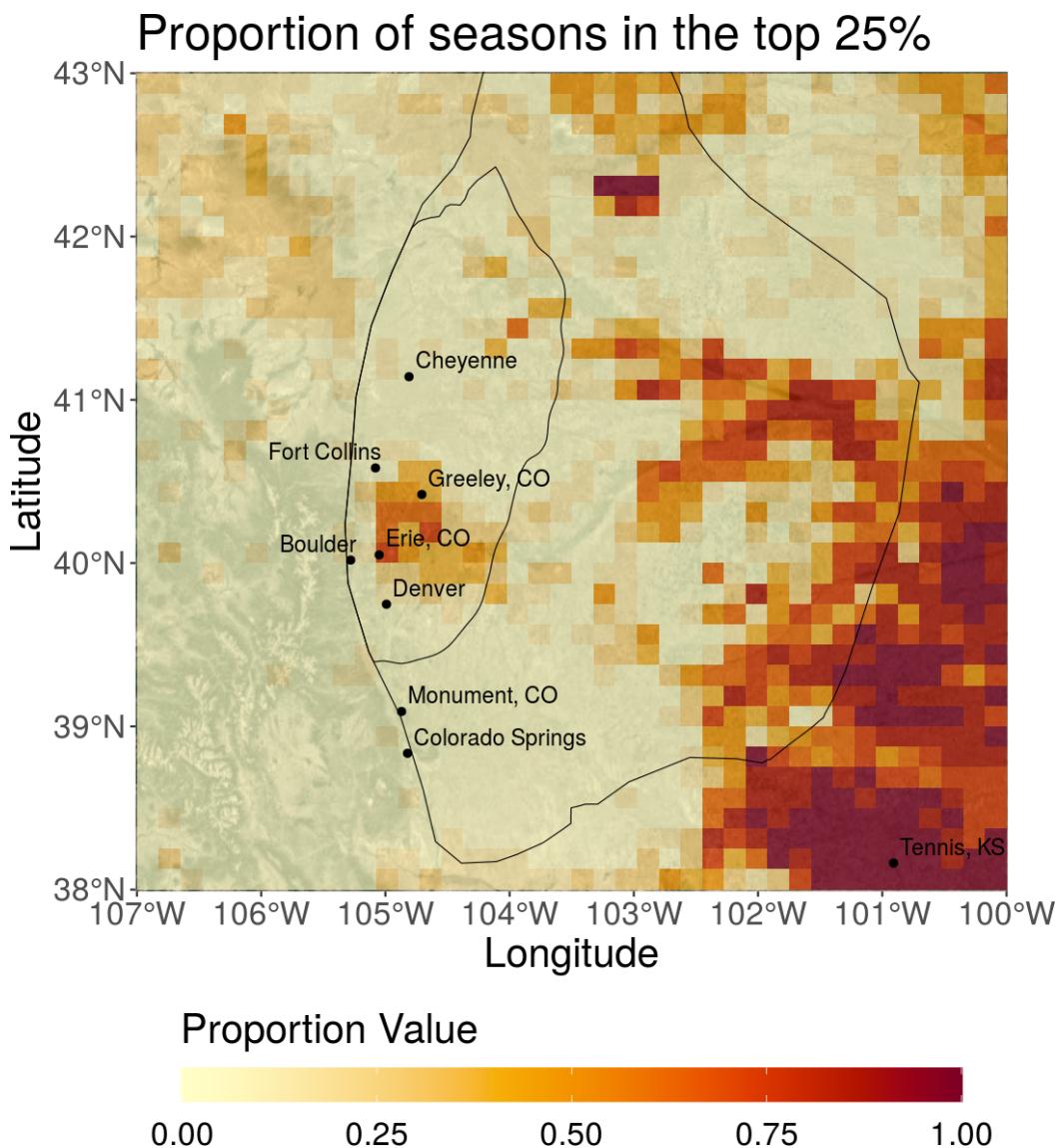
Figure 30: Proportion of seasons in the top 25%, broken up by gridbox. The outer boundary shows the DJ Basin, and the inner boundary shows the Niobara Play.

Here we see that the southeast corner of our region is most likely to contain an average methane value in the top 25%. Again, this could be from larger background methane concentrations in Oklahoma and Kansas. We also see a larger probability of high methane concentrations in the area northeast of Denver, east of Boulder, and southeast of Fort Collins. This could be due to agriculture or oil and gas developments in this area.

As in the previous section, we see the proportions increase as we move from daily to seasonal timescales. Again, this could be because much more data are averaged on the seasonal timescale than the daily timescale. The daily averages with fewer observations are more likely to fluctuate over time. Therefore, it is not very likely that a given gridbox will consistently be in the top 25% day after day, regardless of whether it is typically a large methane emitter. The seasonal averages with more observations are less likely to fluctuate over time, as the outlying methane values get rolled into the averages. Therefore it is more likely that a given gridbox will consistently be in the top 25% season to season if it is typically a large methane emitter.

Note that the analysis in this section could be improved by either: (1) examining a smaller subset of our region, or (2) by removing the background methane concentrations that we see in the southeast and northeast corners of our region.

6 Conclusions

We are interested in studying emissions data in the DJ Basin. So far, we have collected methane data from the TROPOMI instrument, performed an initial exploratory analysis, and created daily, weekly, monthly, and seasonal data products.

Our initial analysis has revealed areas of consistently high methane concentrations. The most obvious of these areas is the southeast corner of our region. High methane concentrations in this region could be due to larger background methane concentrations in Oklahoma and Kansas. The area north of Denver, east of Bolder, and southeast of Fort Collins also shows consistently high concentrations of methane. This could be due to agriculture or oil and gas developments in this area. Finally, there is a "right caret" pattern of consistently high methane concentrations that follows the North and South Platte Rivers. High methane concentrations in this area could be due to wetlands associated with the Platte River or because of the agricultural presence in this area.

We have also performed an initial time series analysis. Viewing methane concentrations over time in five different locations has revealed two potential trends in the data. The first is a seasonal pattern, with higher methane concentrations in the fall and winter and lower concentrations in the spring and summer. The second is an upwards linear trend, indicating that methane concentrations might be generally increasing over time. However, it is important to note that only two years of methane data is available from the TROPOMI instrument. This makes quantifying either of these trends unreliable.

Going forward, our analysis could be improved by removing the high methane background we see in the southeast corner of the region. This could be achieved by simply focusing on a smaller subregion that does not contain this background or by subtracting off the methane concentrations attributed to surrounding states. However, more analysis into Oklahoma and Kansas is needed before any data can be removed or altered.

Finally, we hope to integrate this analysis with another source of data. One option is VIIRS DNB temporal profiles, which contain flaring data. With this data, we could potentially correlate methane concentrations to flaring. Another option is a different methane data source, such as onsite LDAR cameras or drone overflights. Comparing TROPOMI data to these more localized data sources could serve as validation of the various instruments.

The **Payne Institute** *for* **Public Policy**

References

[1] C. D. Elvidge, K. Baugh, M. Zhizhin, F. C. Hsu, and T. Ghosh, "VIIRS night-time lights," *International Journal of Remote Sensing*, vol. 38, no. 21, pp. 5860–5879, 2017. [Online]. Available: https://doi.org/10.1080/01431161.2017.1342050

[2] "Quantifying methane emissions from the largest oil-producing basin in the United States from space," *Science Advances*, vol. 6, no. 17, p. eaaz5120, apr 2020. [Online]. Available: https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.aaz5120

[3] J. P. Veefkind, I. Aben, K. McMullan, H. Förster, J. de Vries, G. Otter, J. Claas, H. J. Eskes, J. F. de Haan, Q. Kleipool, M. van Weele, O. Hasekamp, R. Hoogeveen, J. Landgraf, R. Snel, P. Tol, P. Ingmann, R. Voors, B. Kruizinga, R. Vink, H. Visser, and P. F. Levelt, "TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications," *Remote Sensing of Environment*, vol. 120, pp. 70–83, may 2012.

[4] "Colorado oil and gas conservation comission," https://cogcc.state.co.us/data2.html#/downloads, accessed: 2020-04-19.

[5] "A glimpse of colorado's agriculture," https://www.colorado.gov/pacific/sites/default/files/Glimpse%20of%20CO%20Agriculture%20Map%202.pdf, accessed: 2020-04-19.

[6] T. R. Christensen, A. Ekberg, L. Ström, M. Mastepanov, N. Panikov, M. Öquist, B. H. Svensson, H. Nykänen, P. J. Martikainen, and H. Oskarsson, "Factors controlling large scale variations in methane emissions from wetlands," *Geophysical Research Letters*, vol. 30, no. 7, apr 2003. [Online]. Available: http://doi.wiley.com/10.1029/2002GL016848

[7] D. I. Stern and R. K. Kaufmann, "Estimates of global anthropogenic methane emissions 1860-1993," *Chemosphere*, vol. 33, no. 1, pp. 159–176, jul 1996.

The Payne Institute
for Public Policy

Appendix A   Data and Code Availability

The TROPOMI data and `R` code used in this analysis can be found at:

> `https://github.com/wsdaniels/DJemissions.git`

This repository contains the following:

- *methane_data.* This directory contains TROPOMI methane data. The data is stored in `.RData` files.

- *scripts.* This directory contains the R scripts used in the analysis found in this report.

# Appendix B   Daily Plots
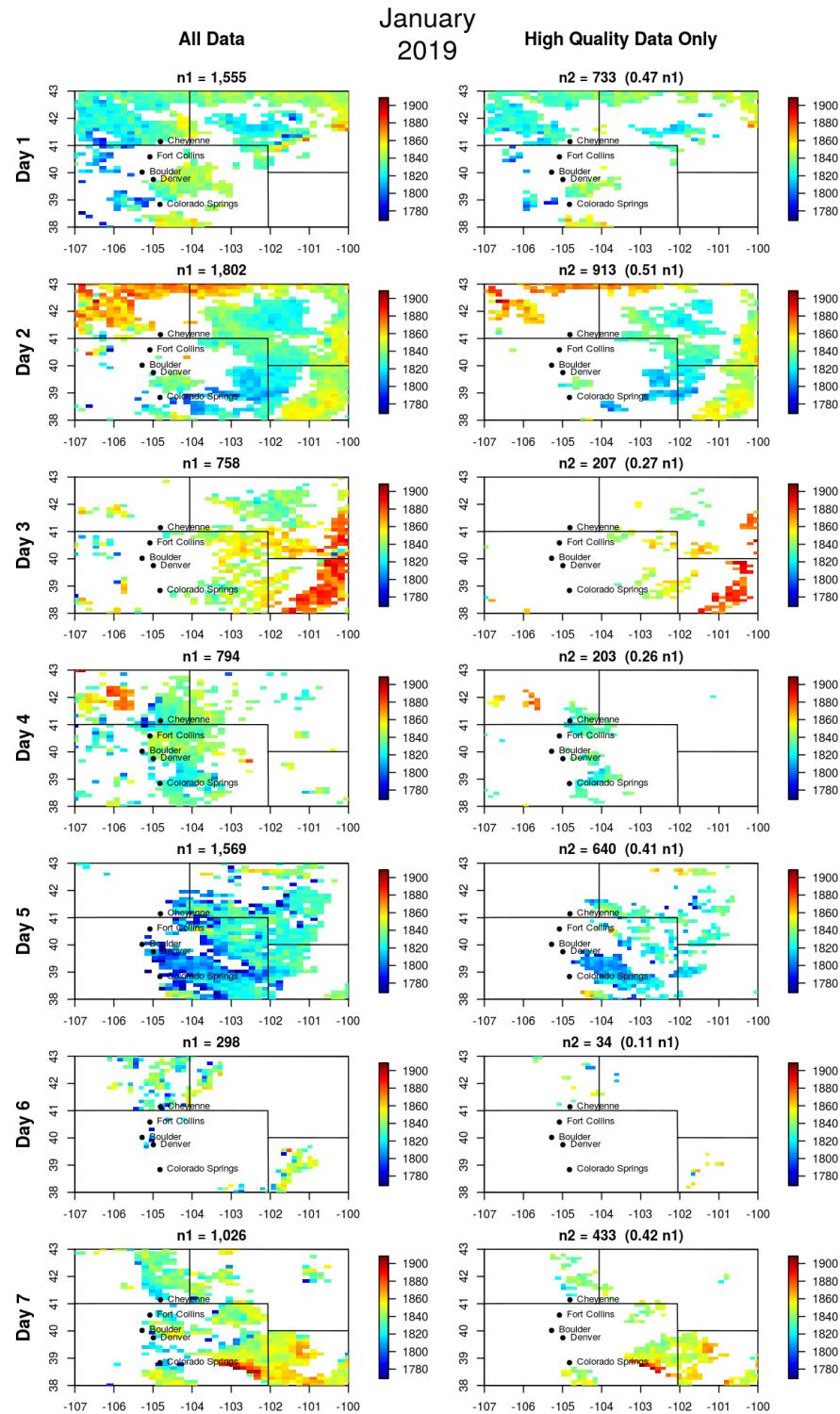
Figure 31
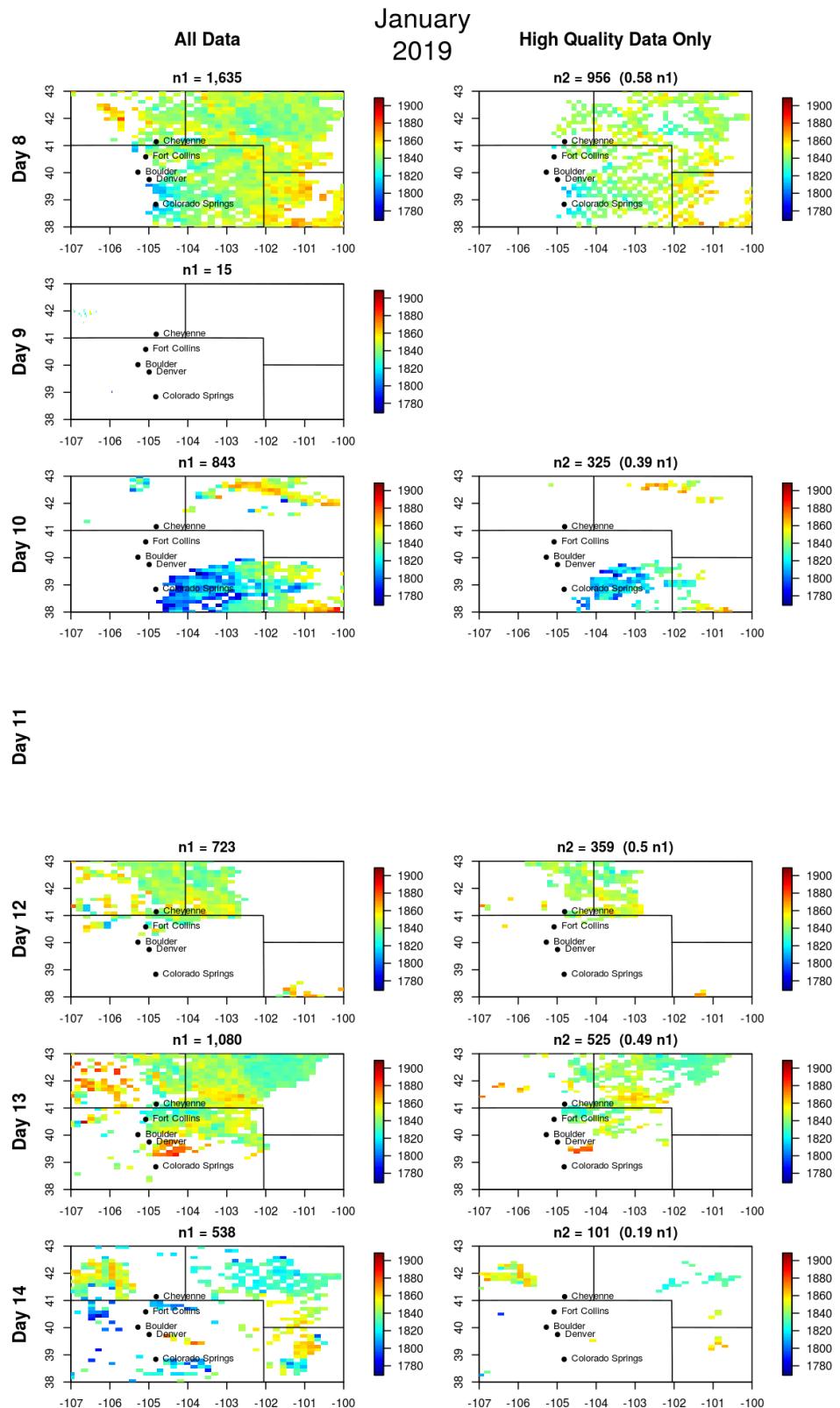
Figure 32
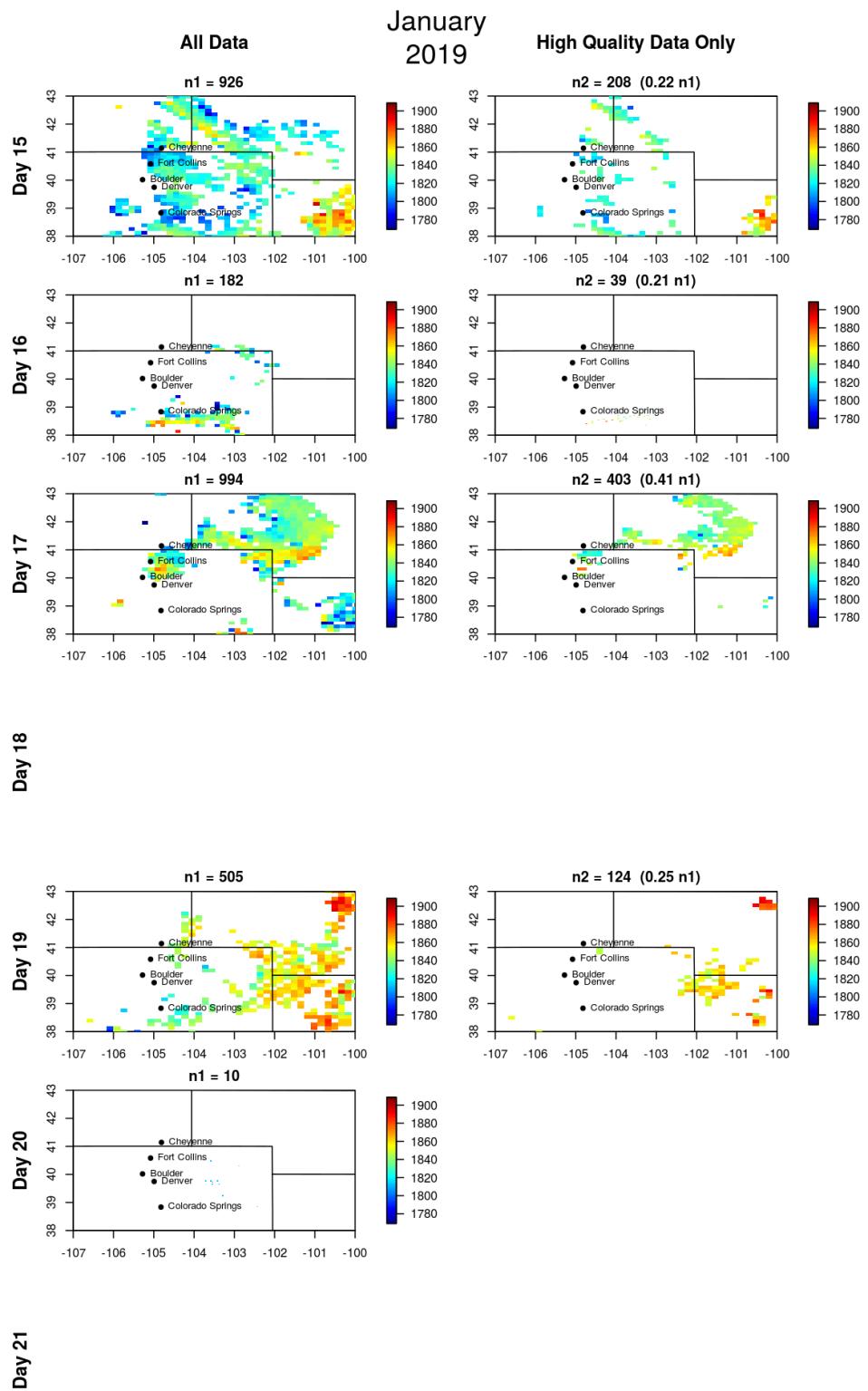
Figure 33
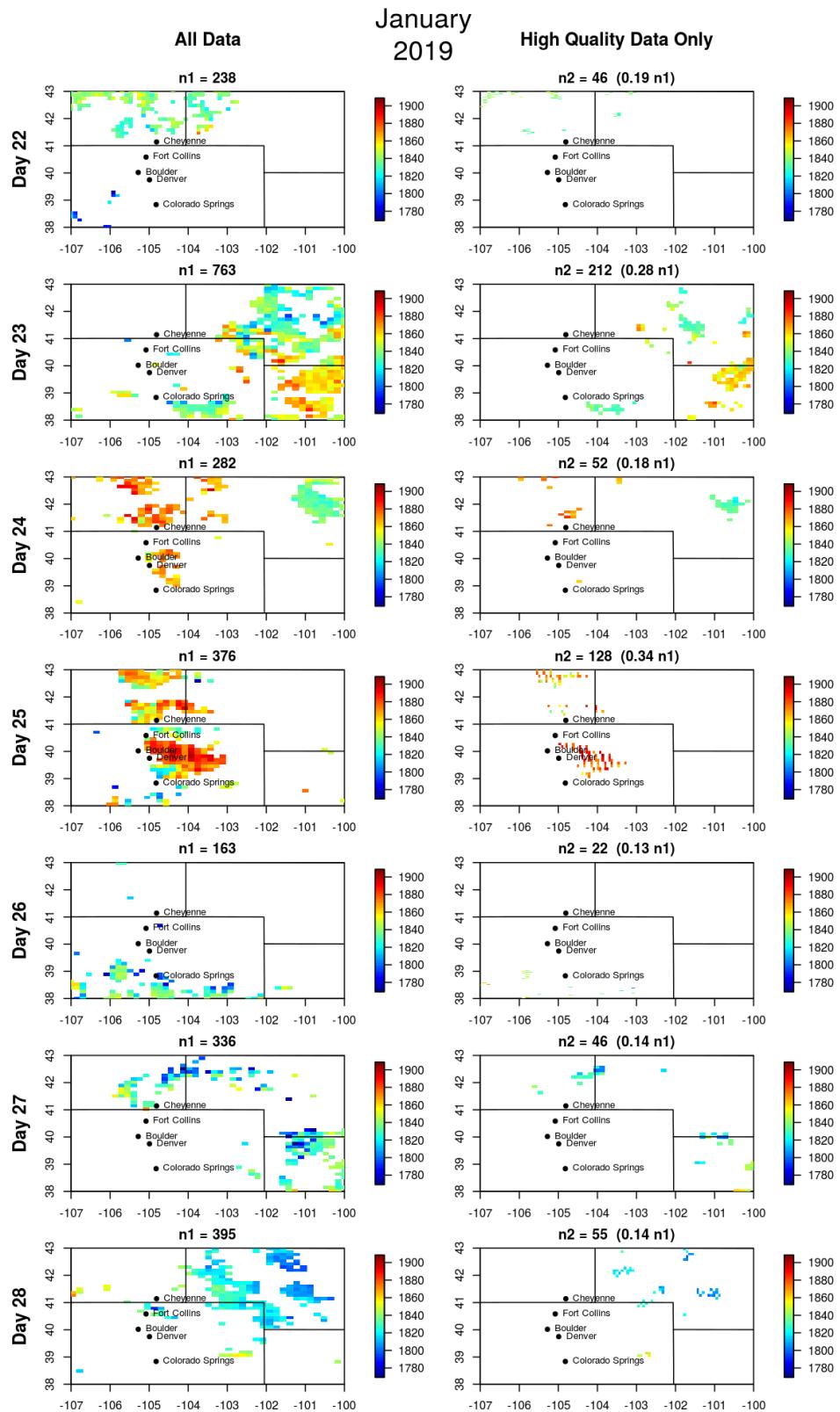
Figure 34
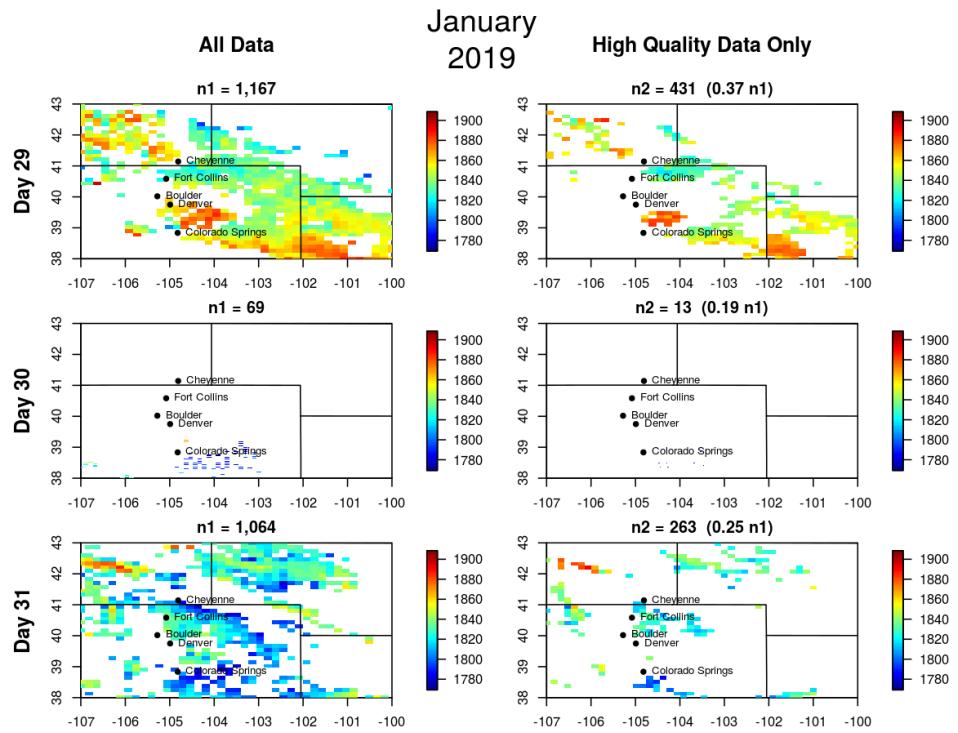
Figure 35
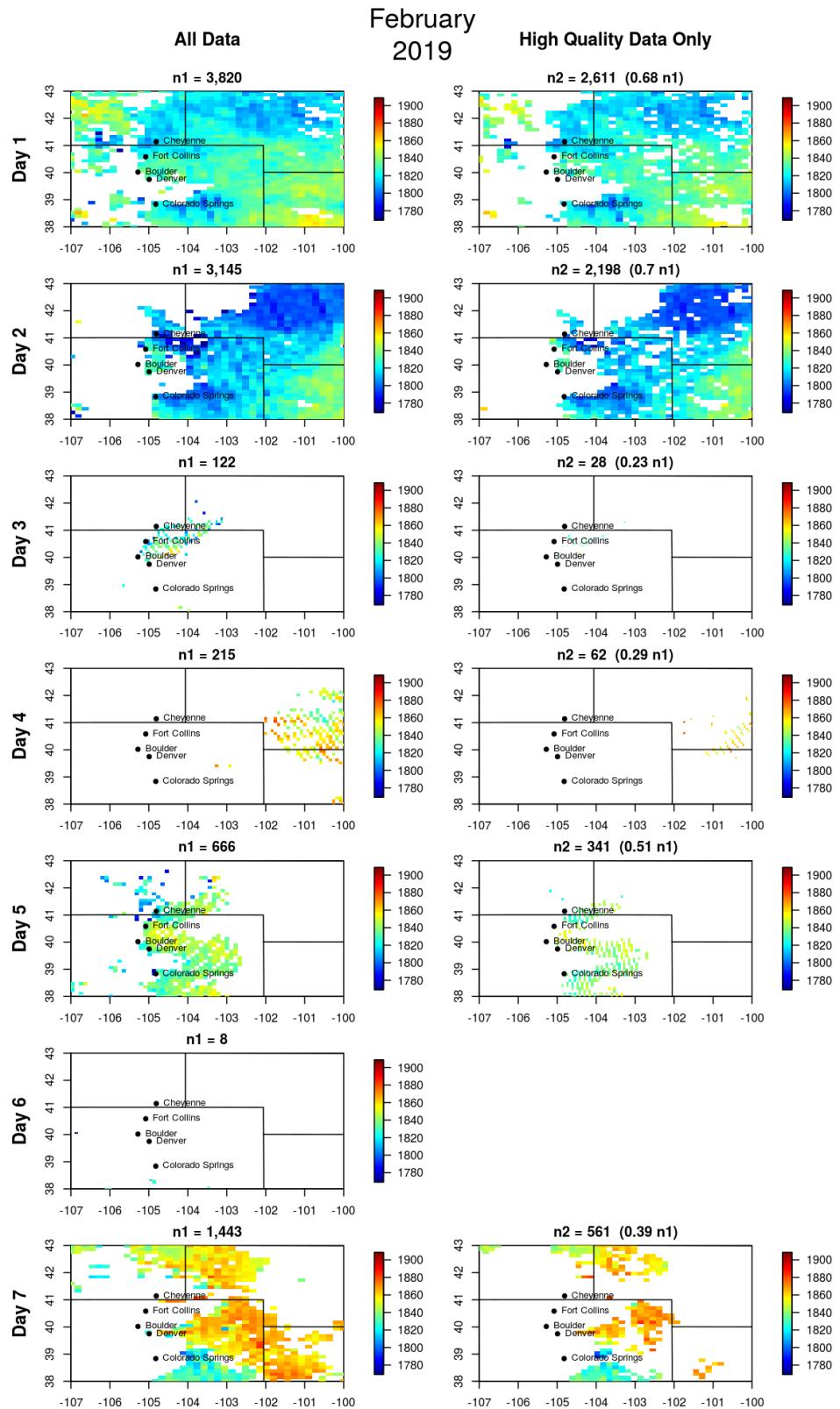
Figure 36

Figure 37

42
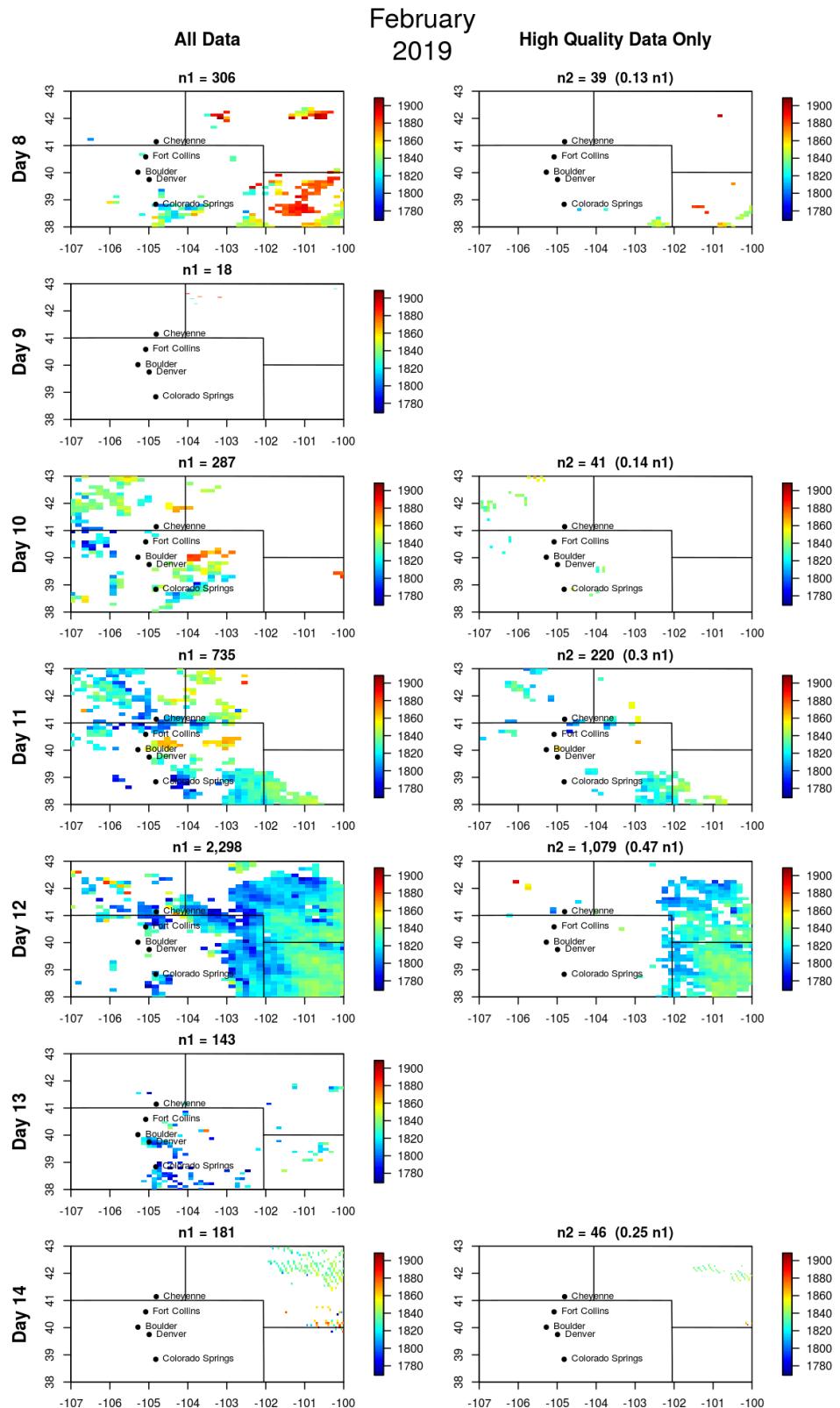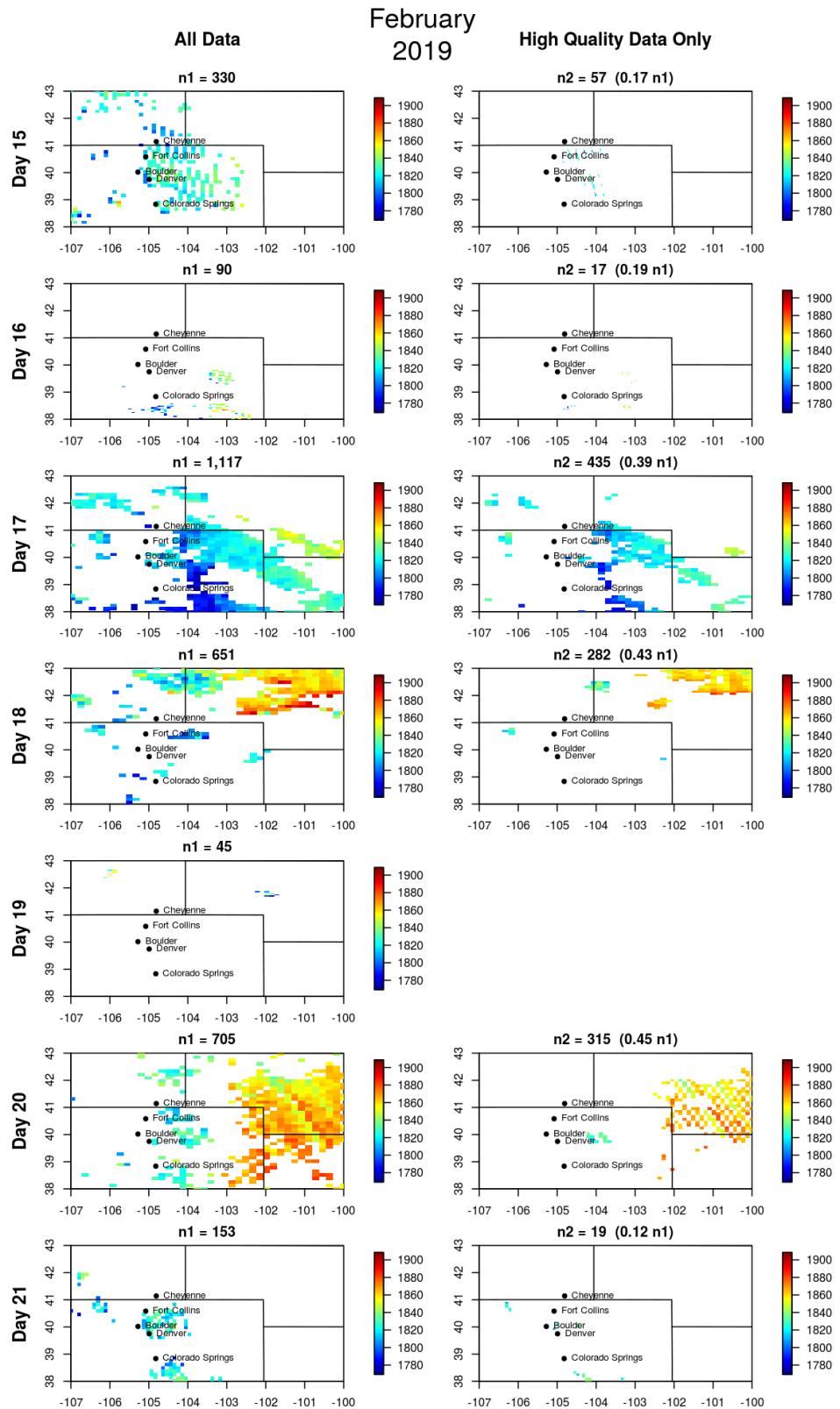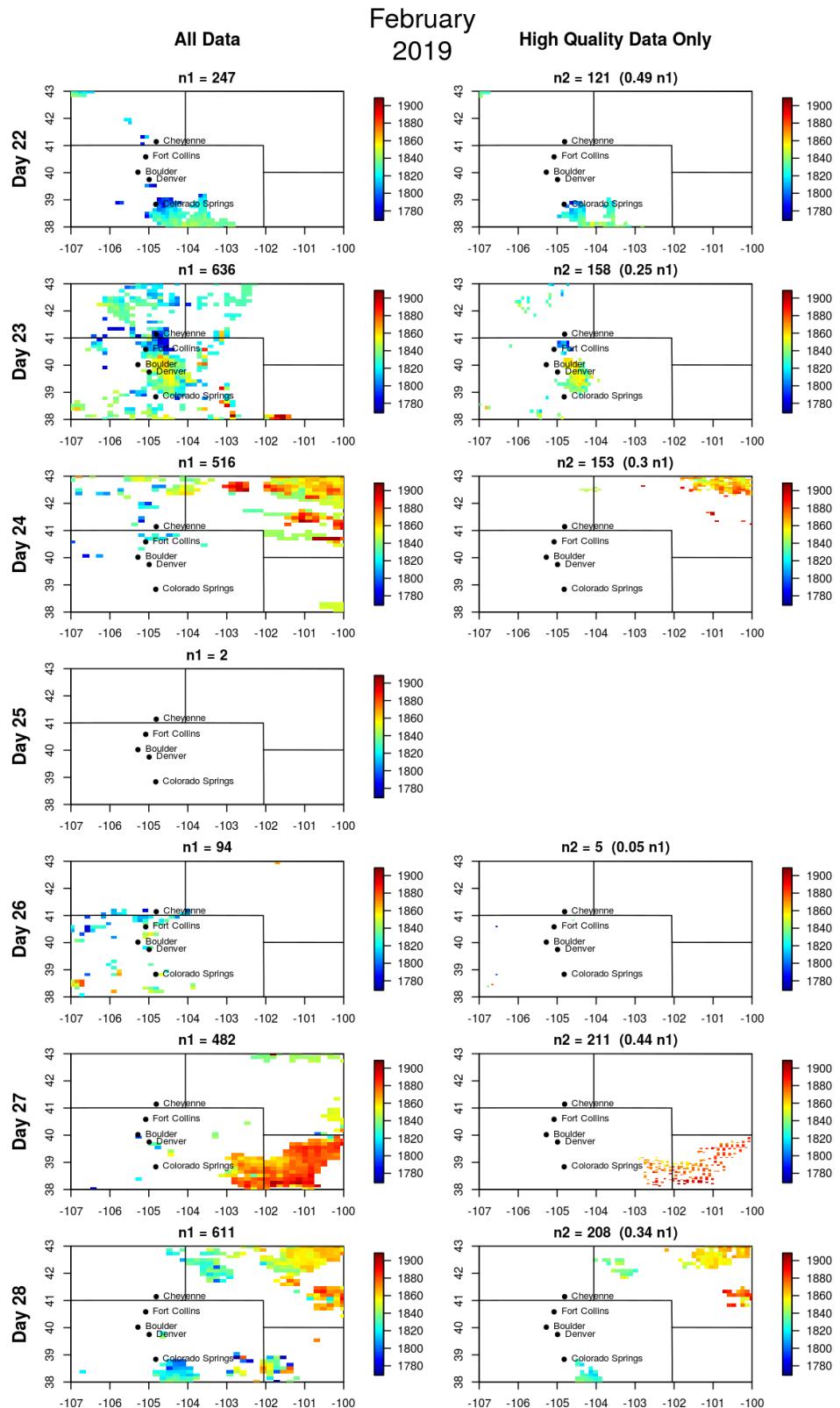
Figure 38

Figure 39

Appendix C   Weekly Plots
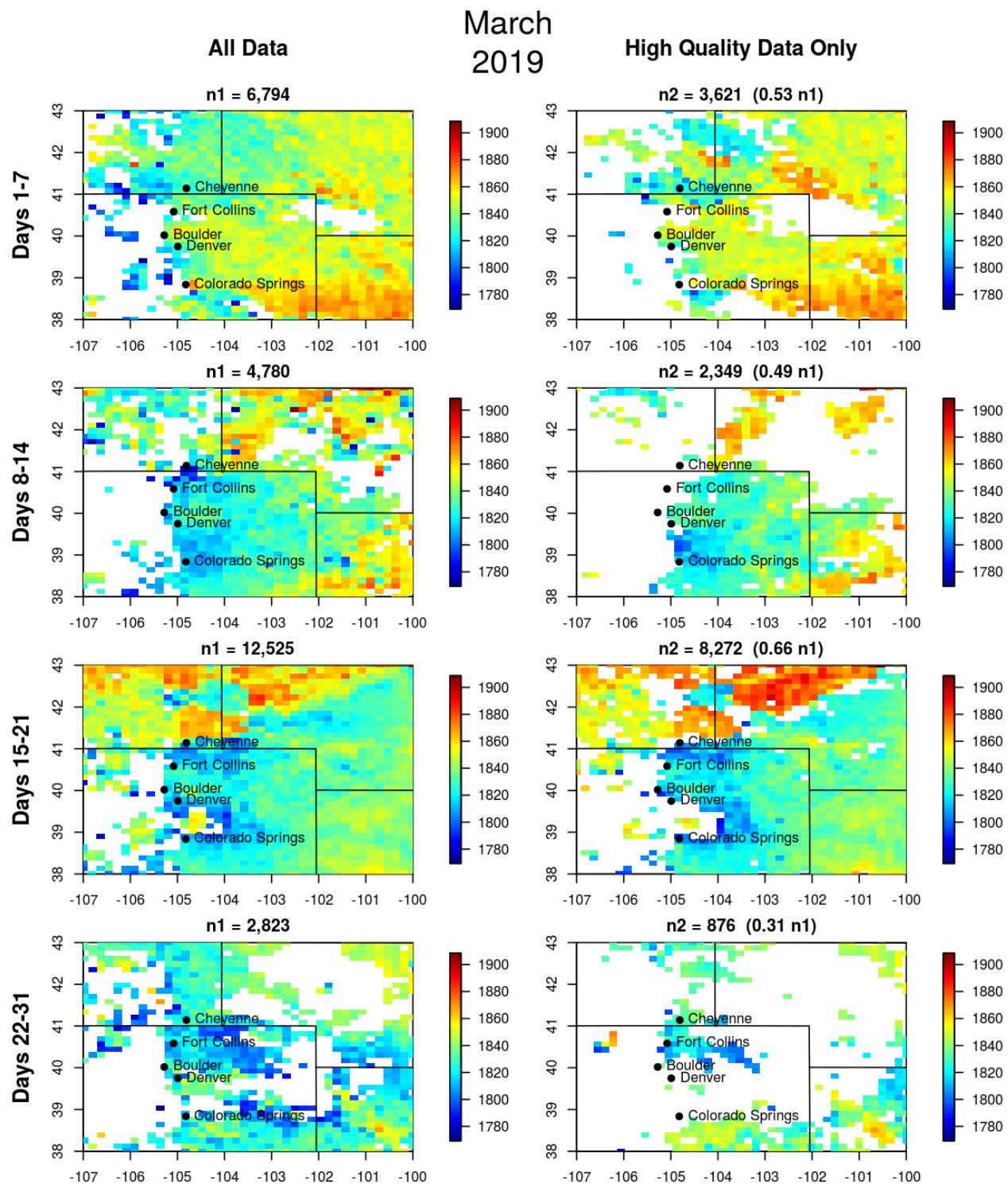
Figure 40
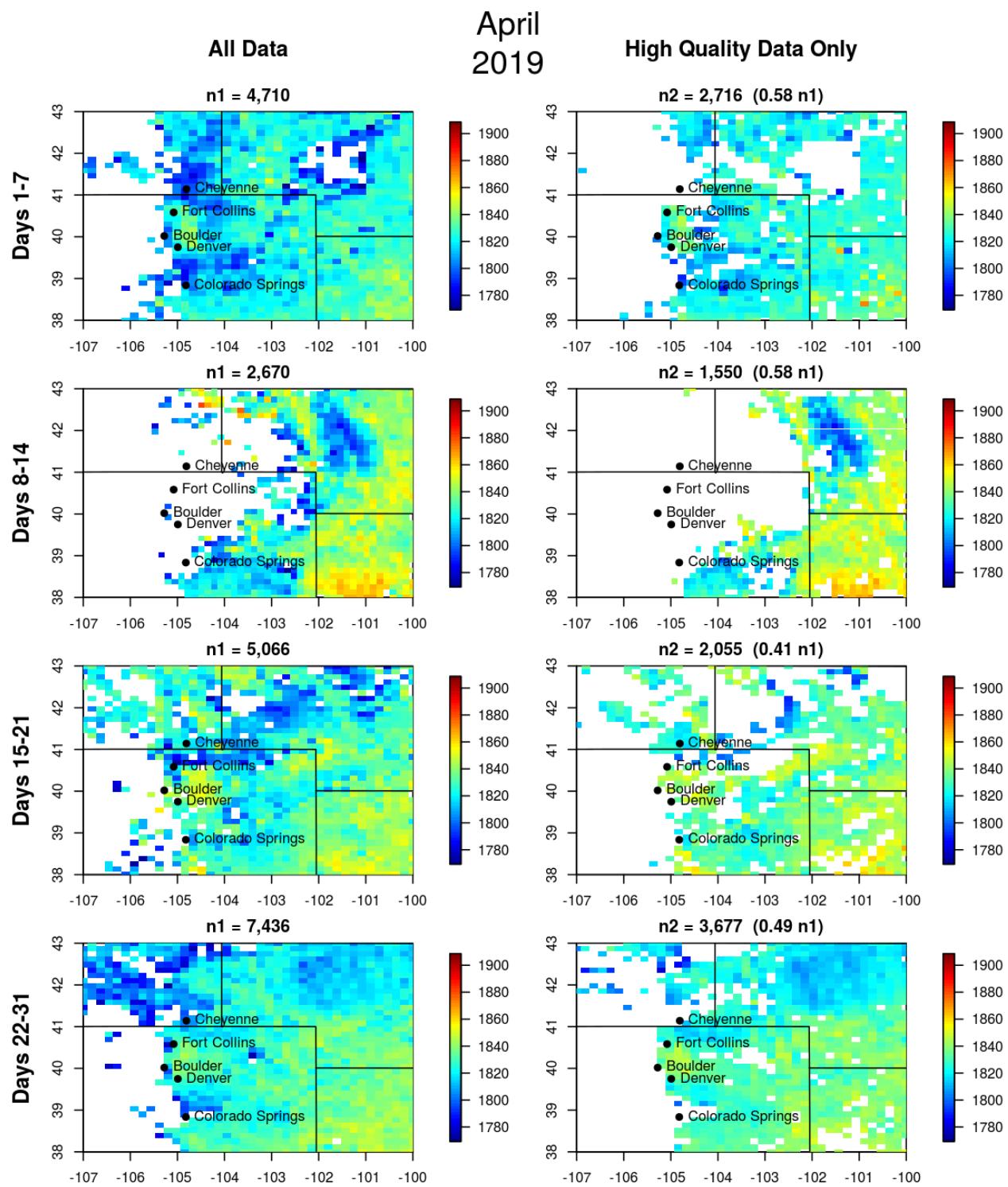
Figure 41
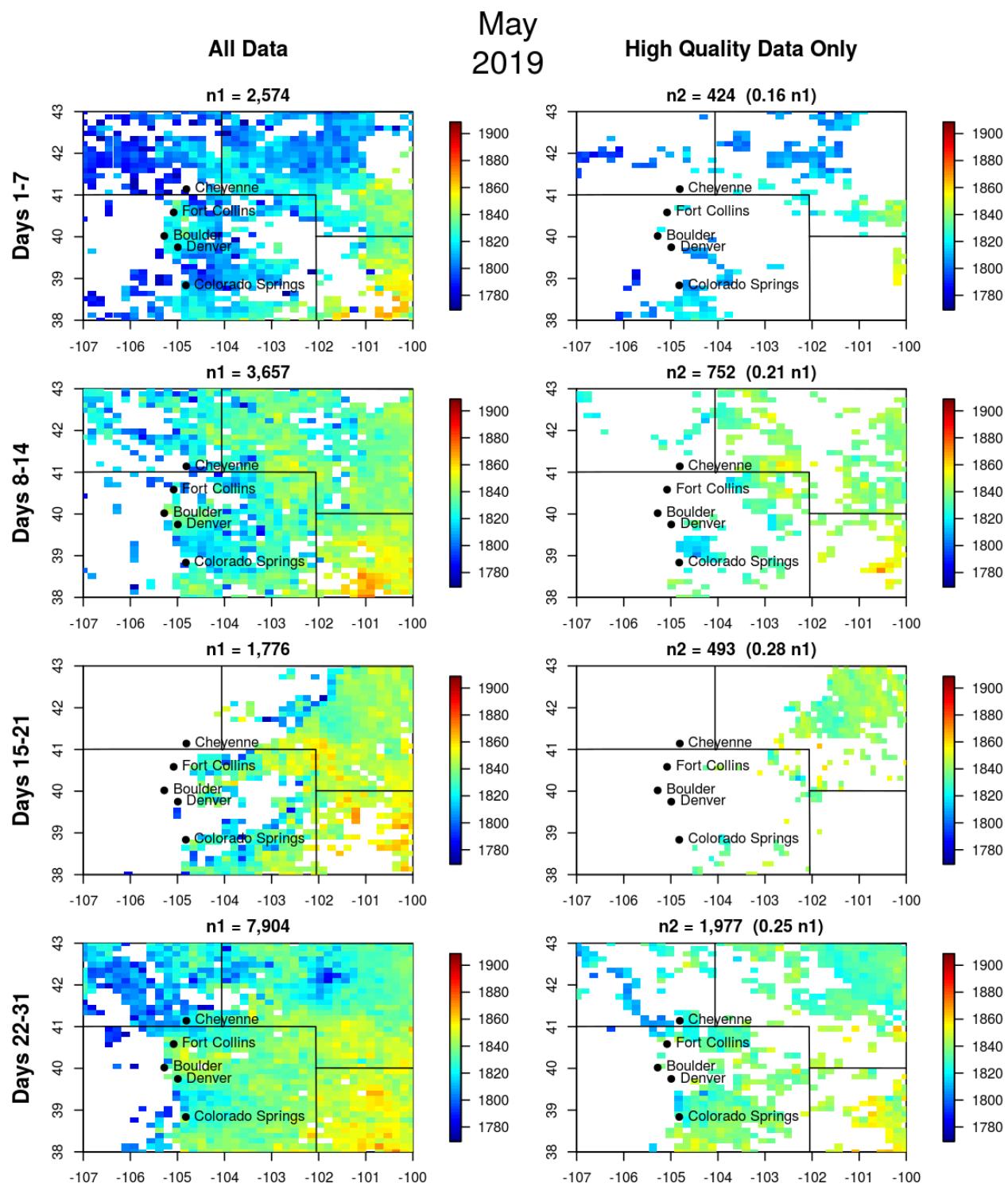
Figure 42

Figure 43

Figure 44

49

August
2019

Figure 45

The Payne Institute for Public Policy

September 2019

**All Data**

High Quality Data Only

Figure 46

October 2019

Figure 47

November 2019

Figure 48

December 2019

Figure 49

54
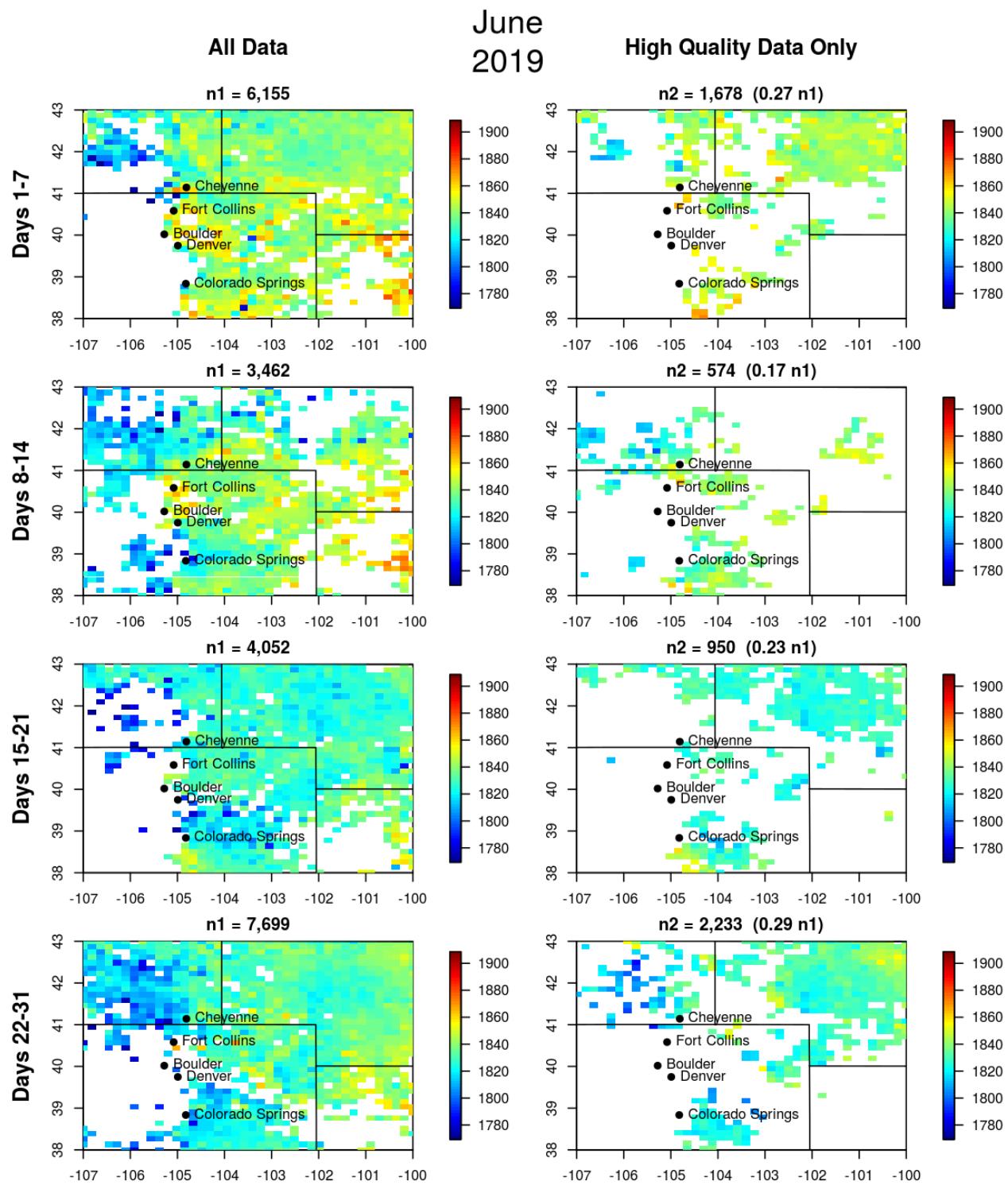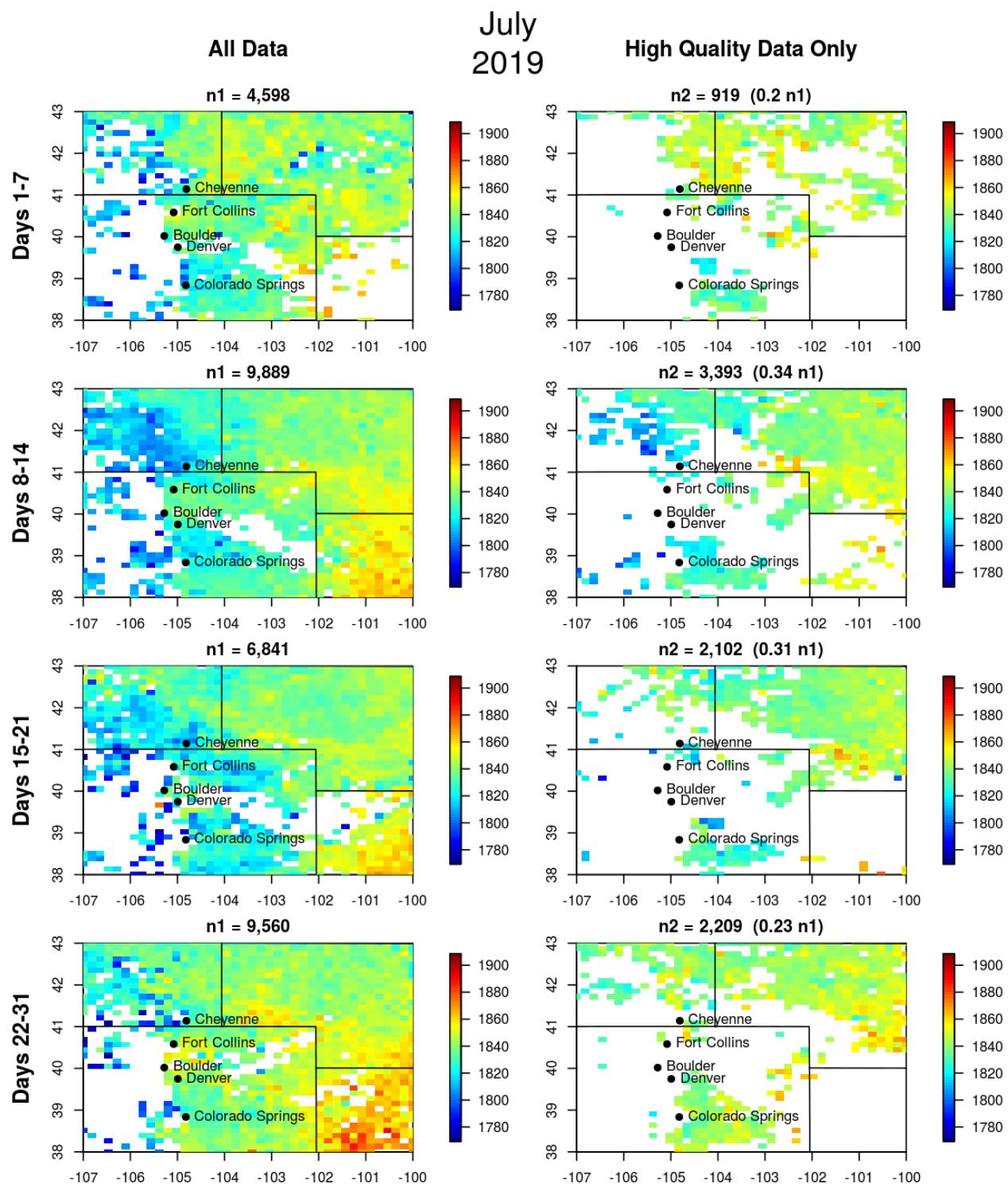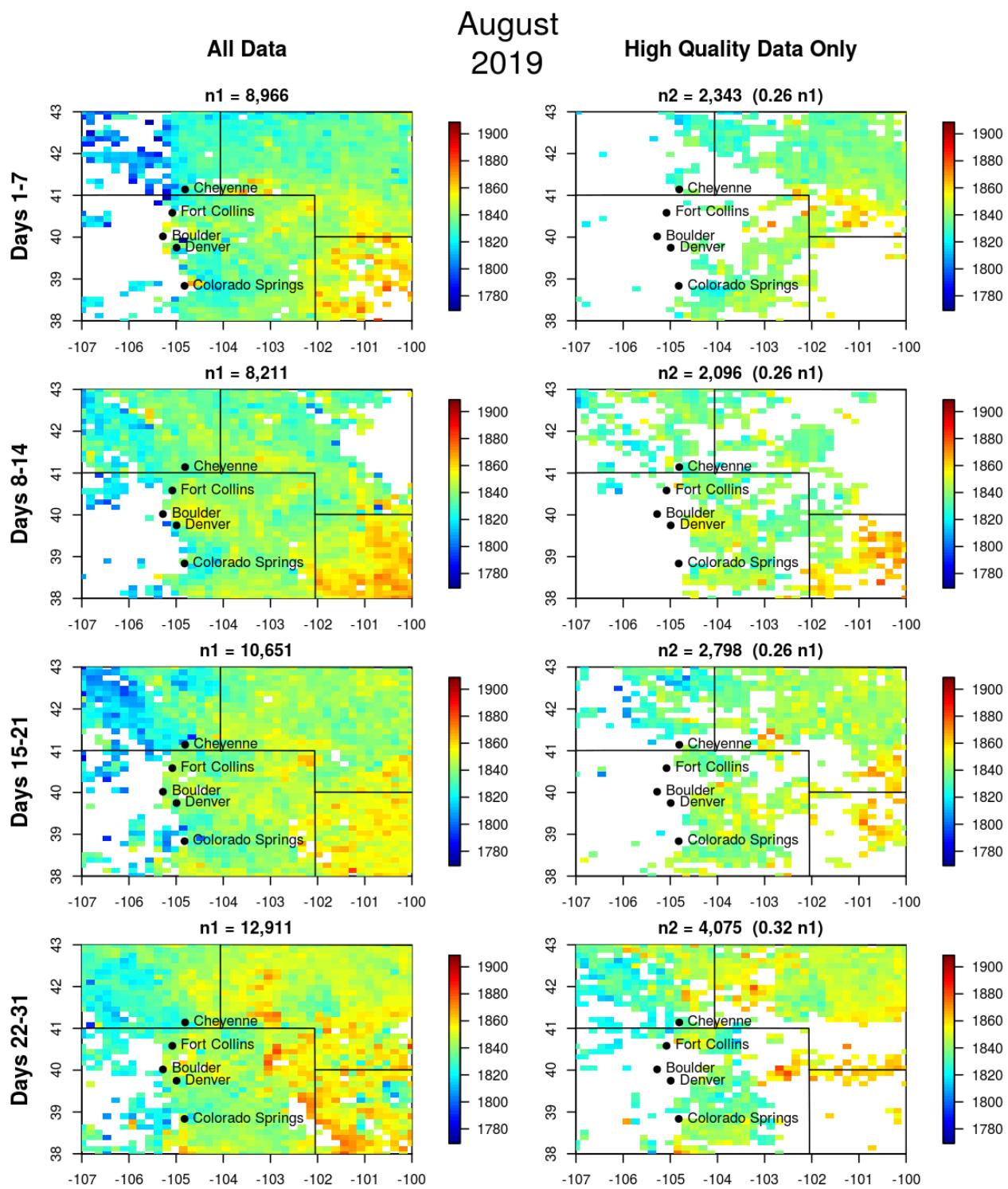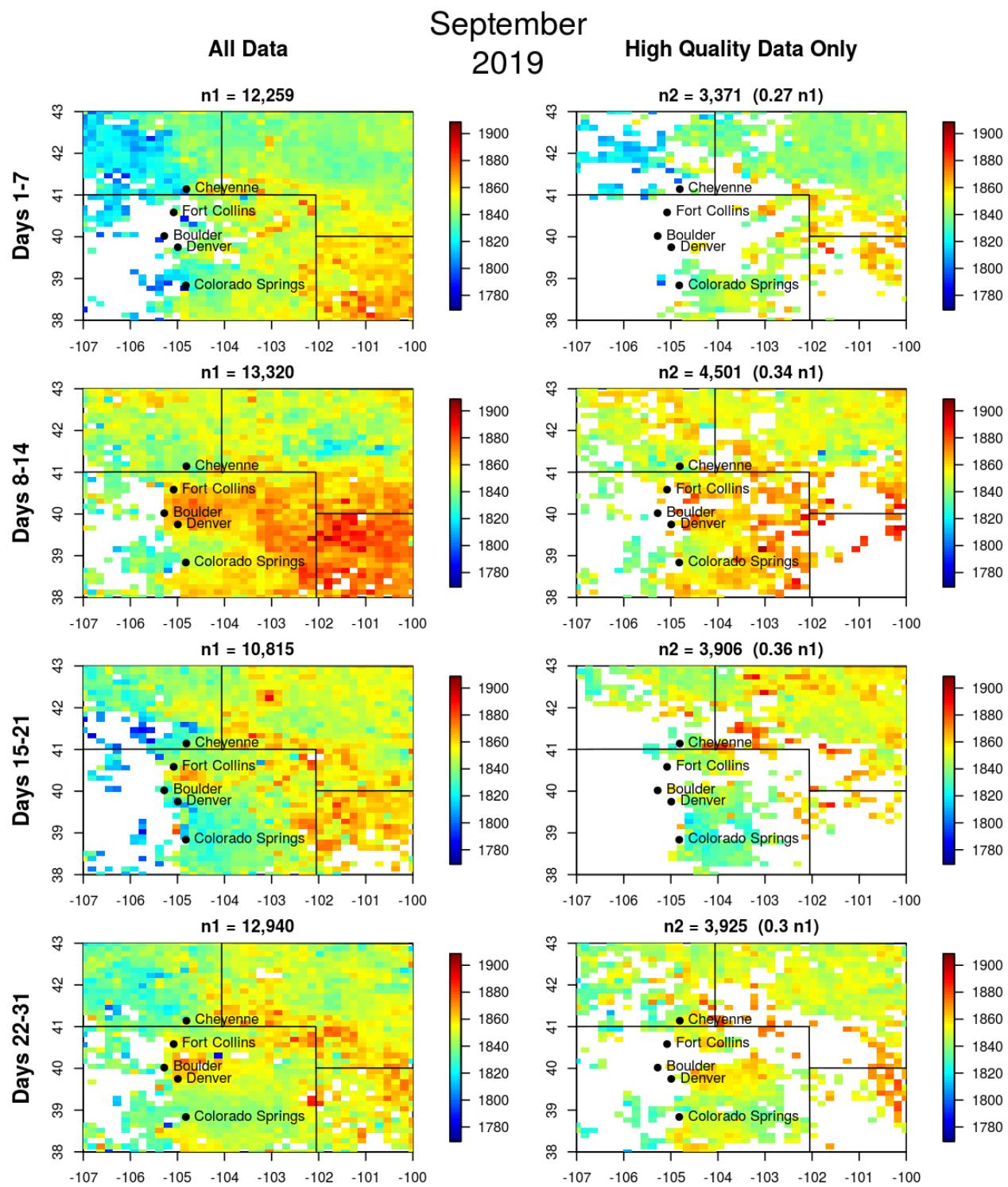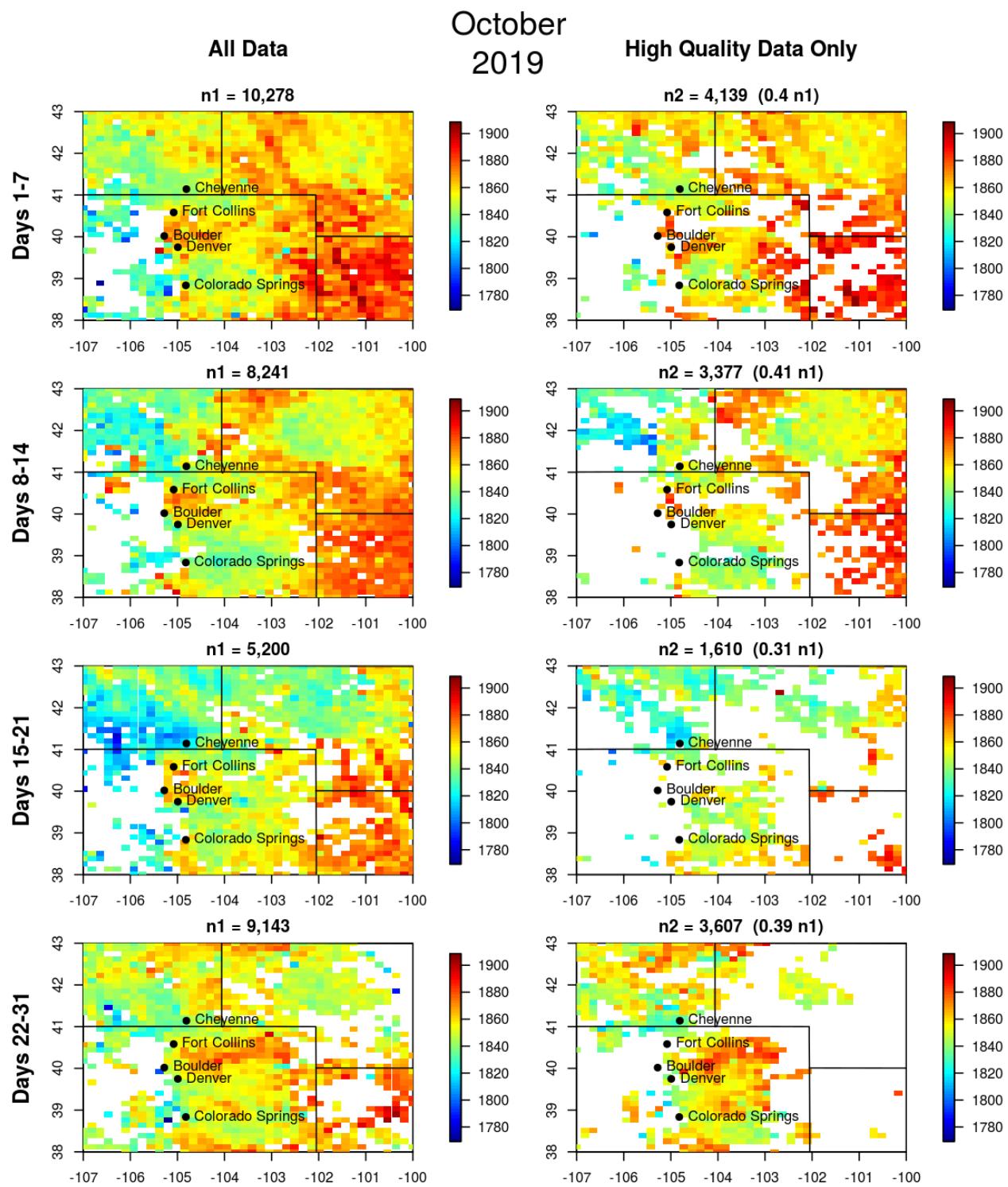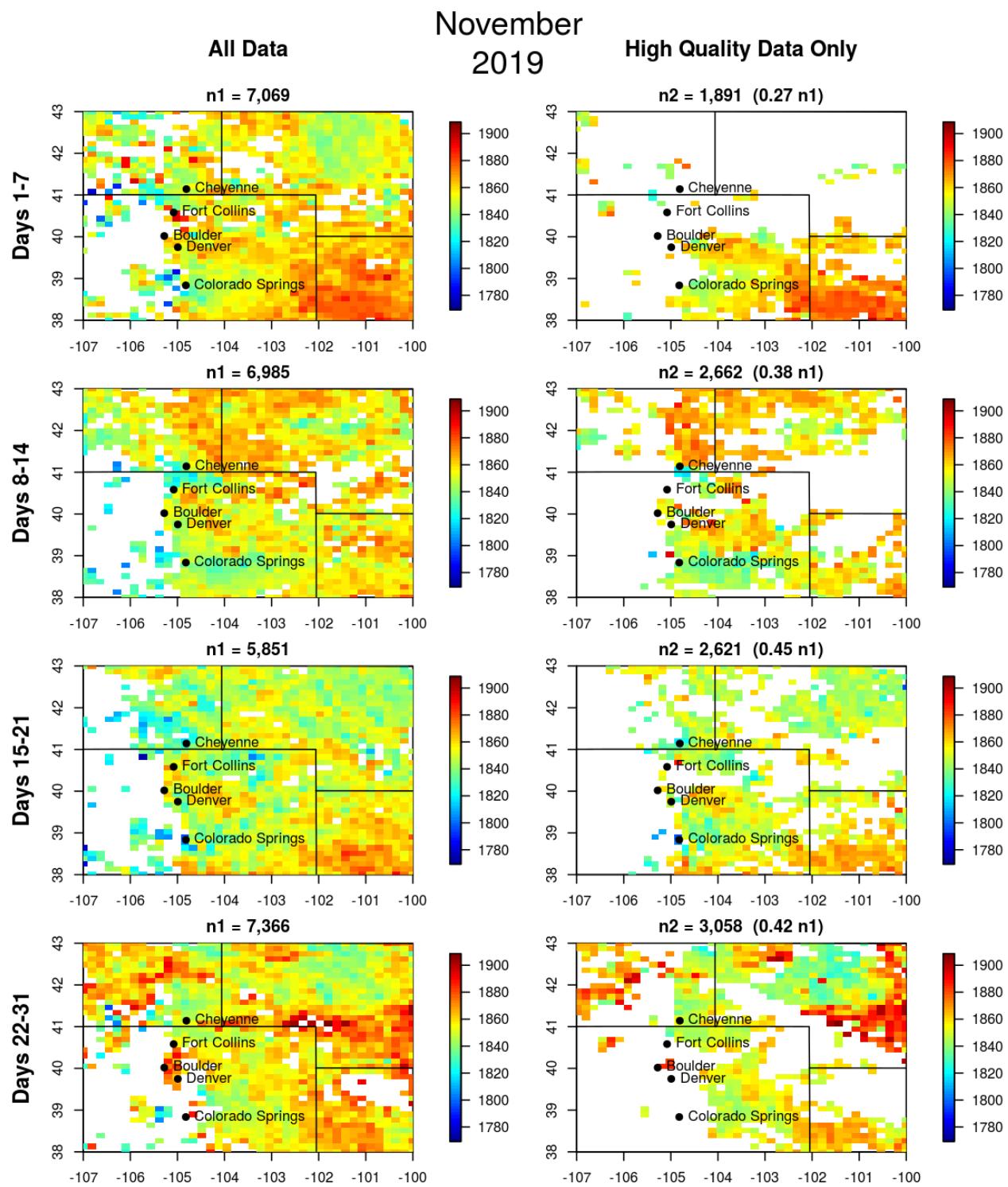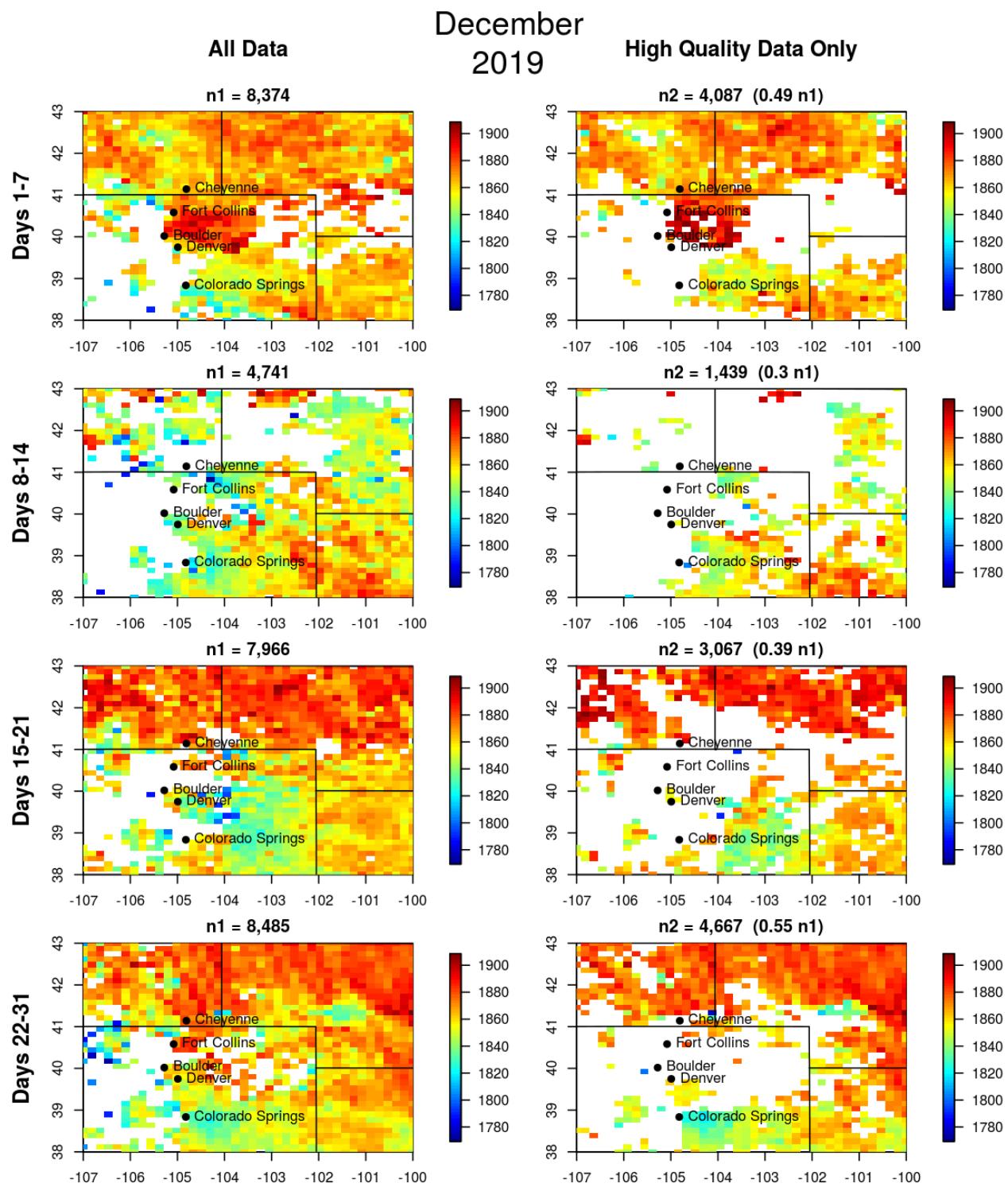
Appendix D   Monthly & Seasonal Plots

Spring 2018

All Data   High Quality Data Only
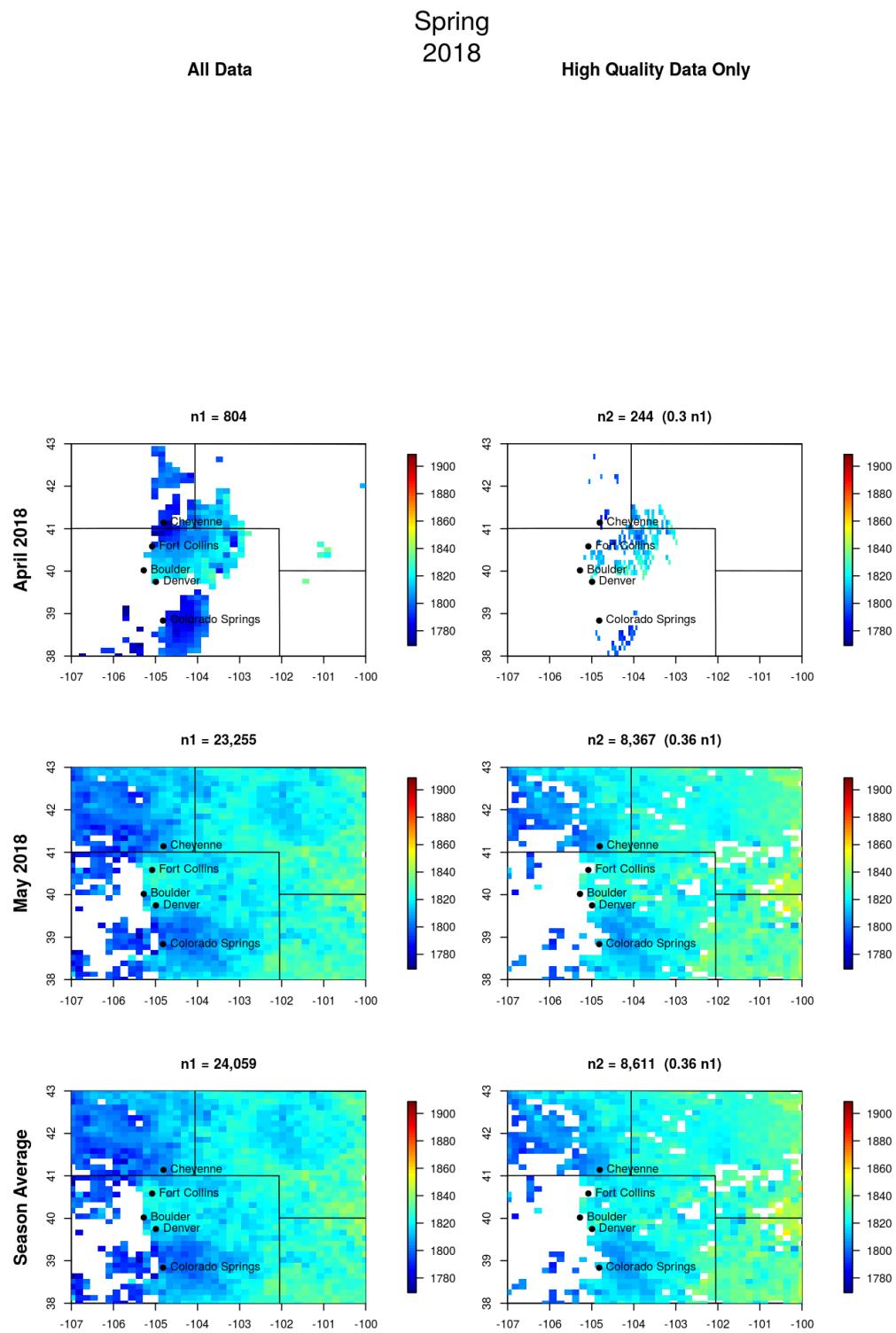
Figure 50: Spring 2018 monthly methane averages. Left column includes all data. Right column includes only high quality data. Bottom row is seasonal average.
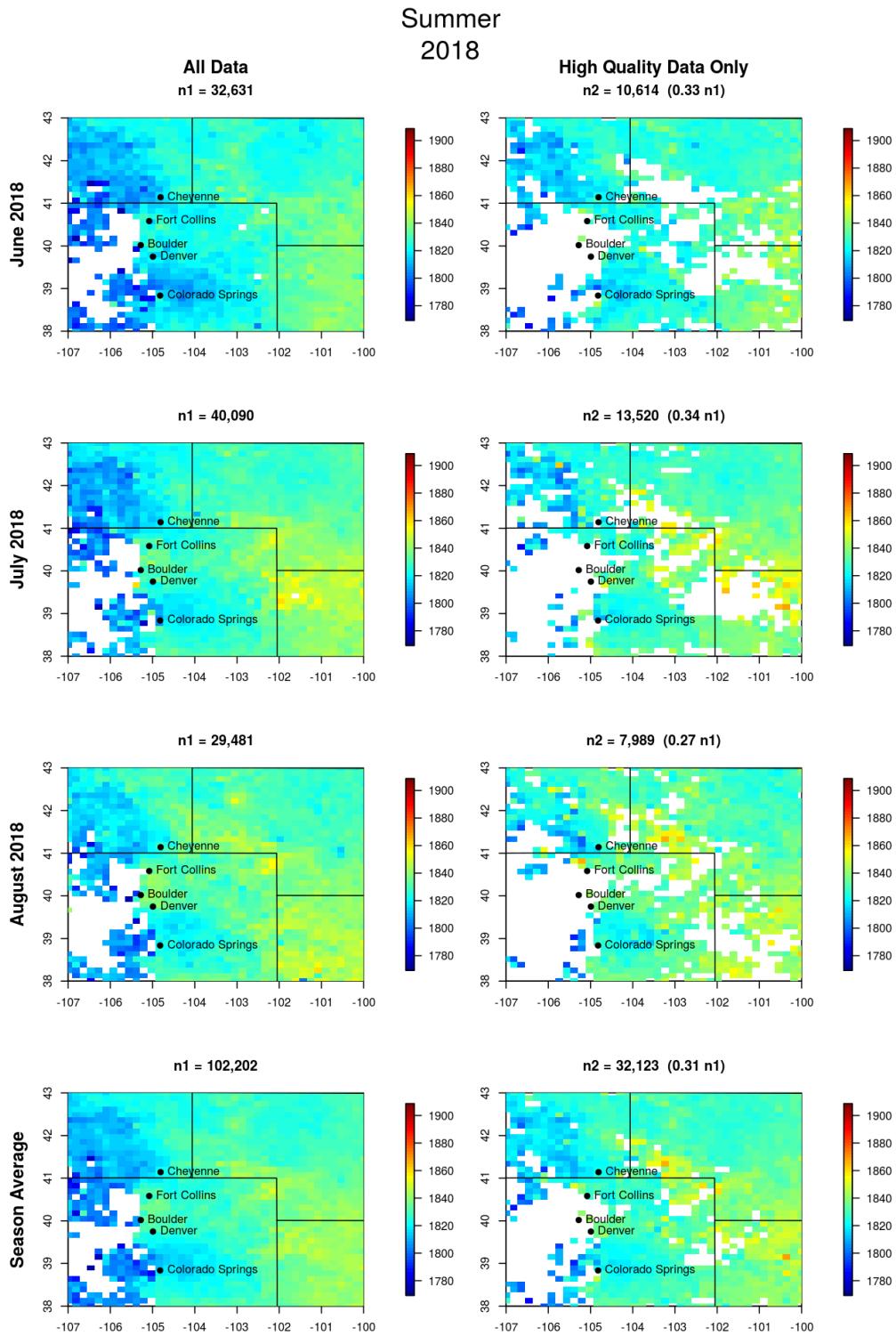
Figure 51: Summer 2018 monthly methane averages. Left column includes all data. Right column includes only high quality data. Bottom row is seasonal average.
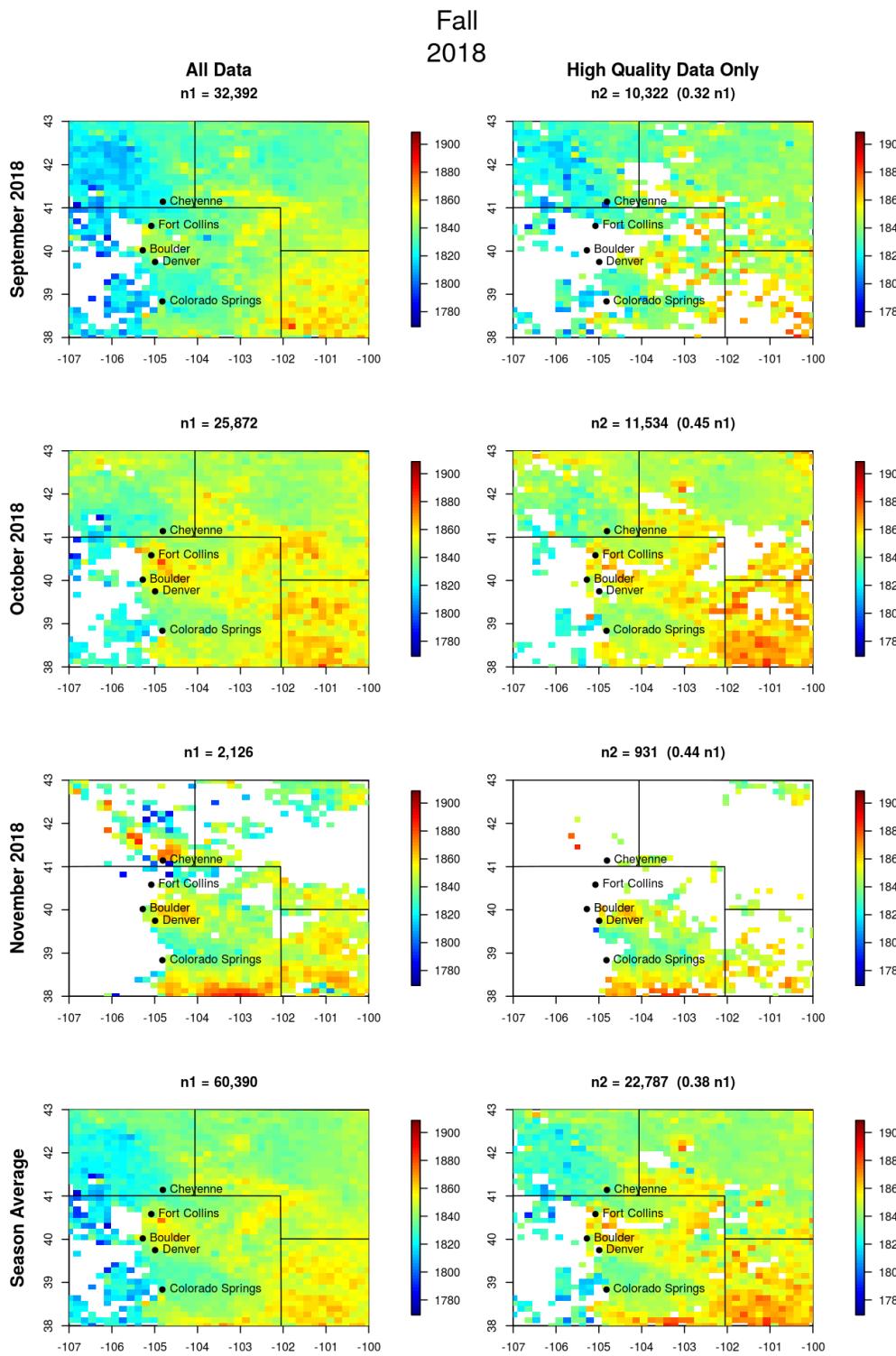
Figure 52: Fall 2018 monthly methane averages. Left column includes all data. Right column includes only high quality data. Bottom row is seasonal average.
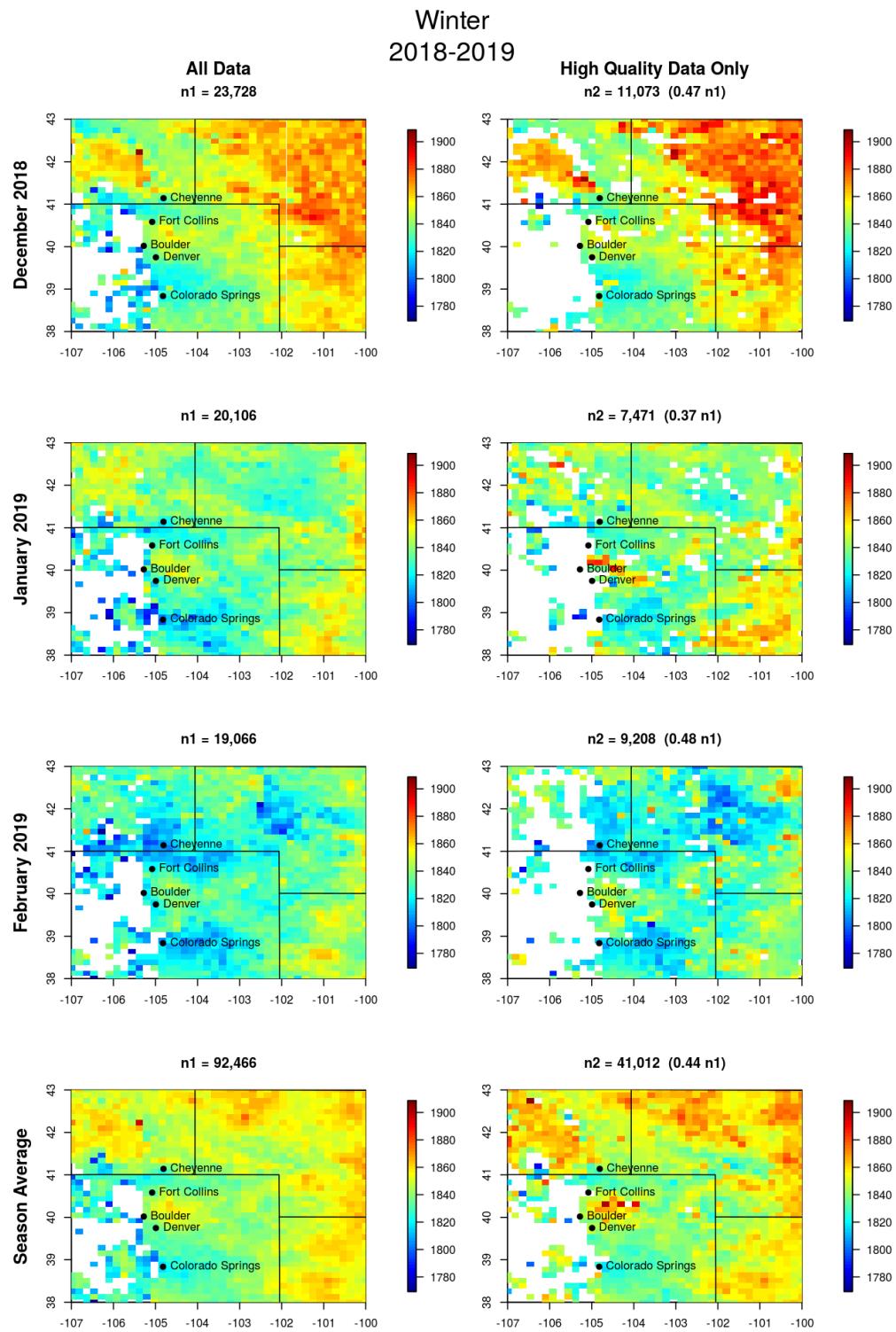
Figure 53: Winter 2018-2019 monthly methane averages. Left column includes all data. Right column includes only high quality data. Bottom row is seasonal average.
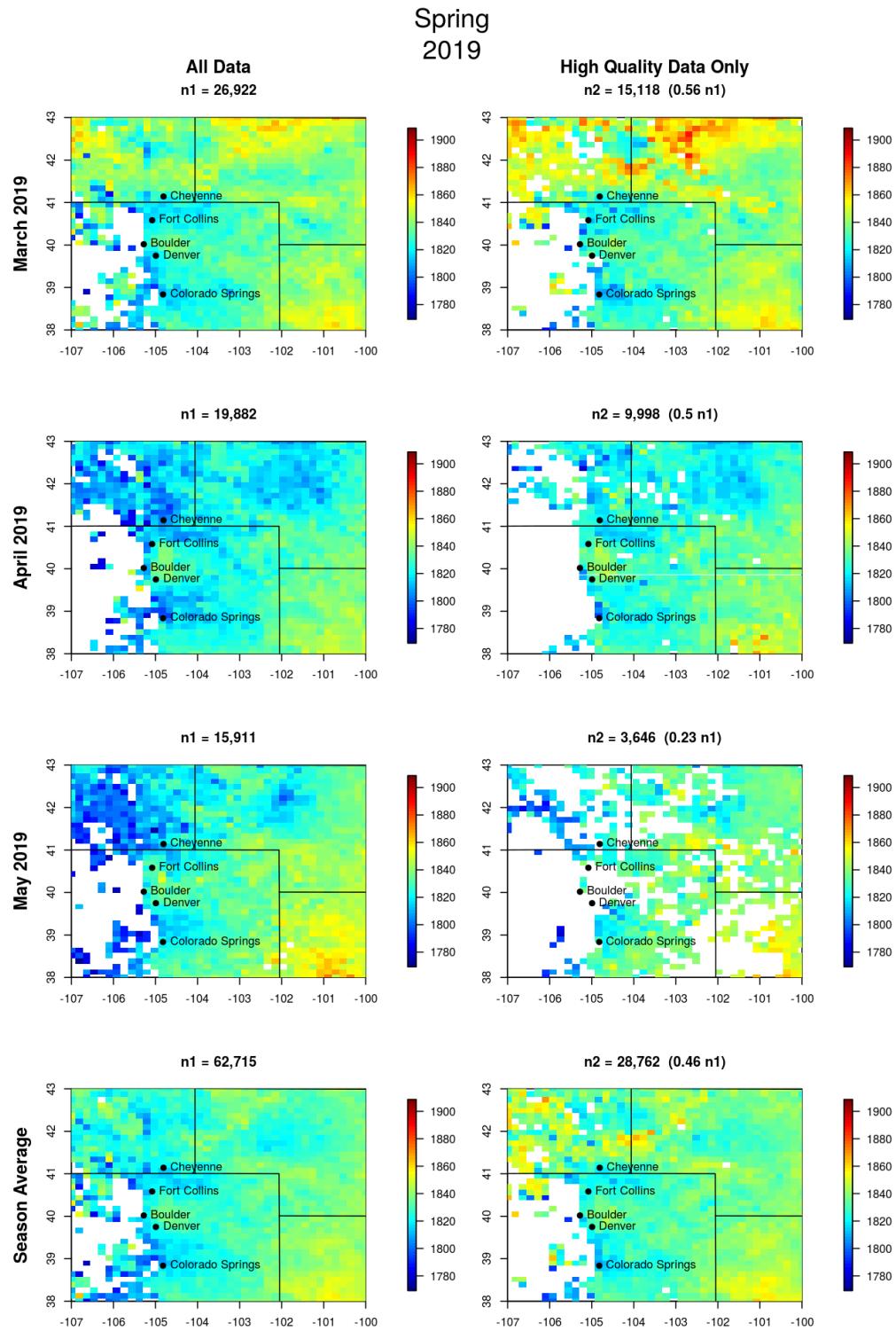
Figure 54: Spring 2019 monthly methane averages. Left column includes all data. Right column includes only high quality data. Bottom row is seasonal average.
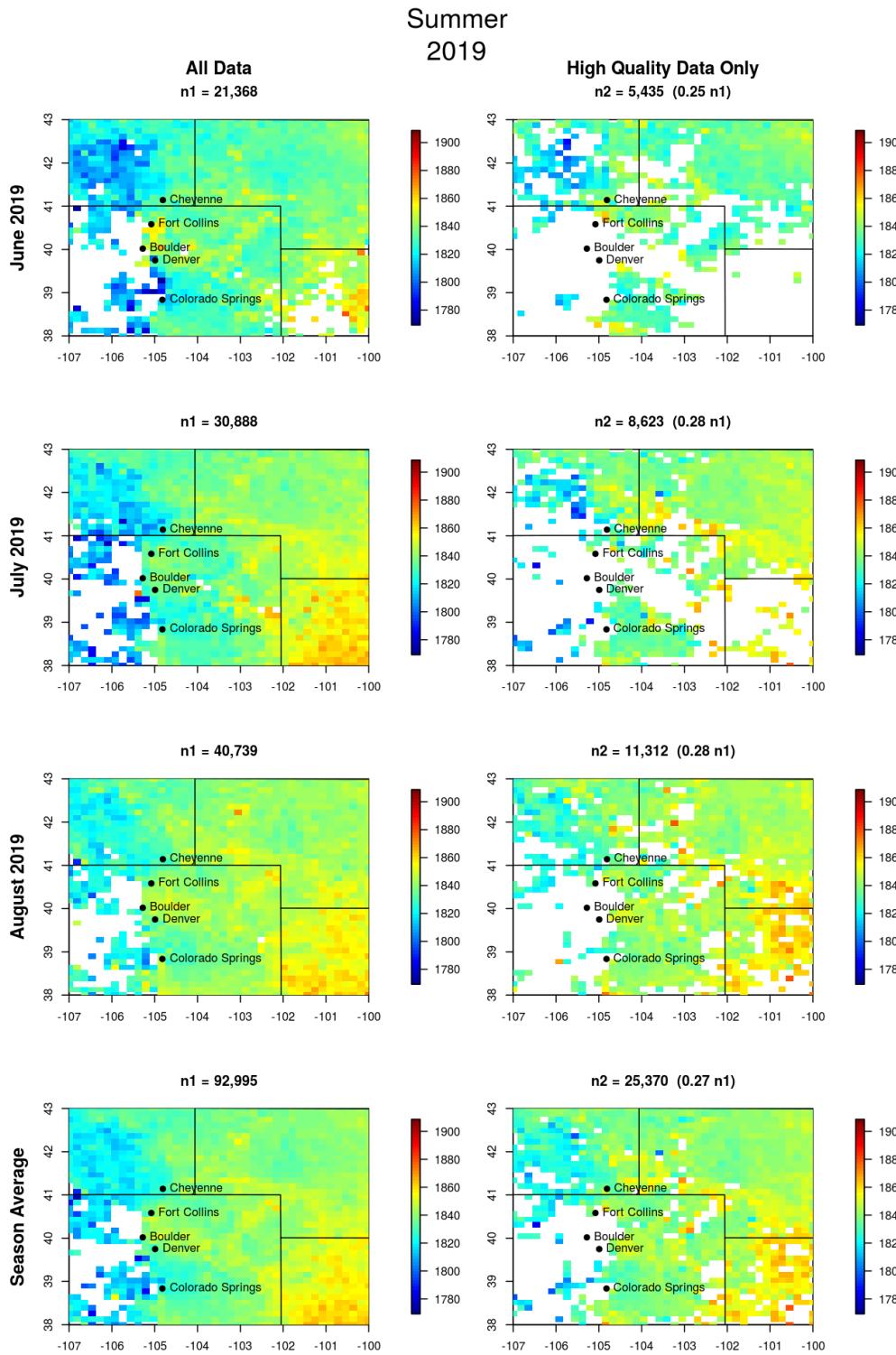
Figure 55: Summer 2019 monthly methane averages. Left column includes all data. Right column includes only high quality data. Bottom row is seasonal average.
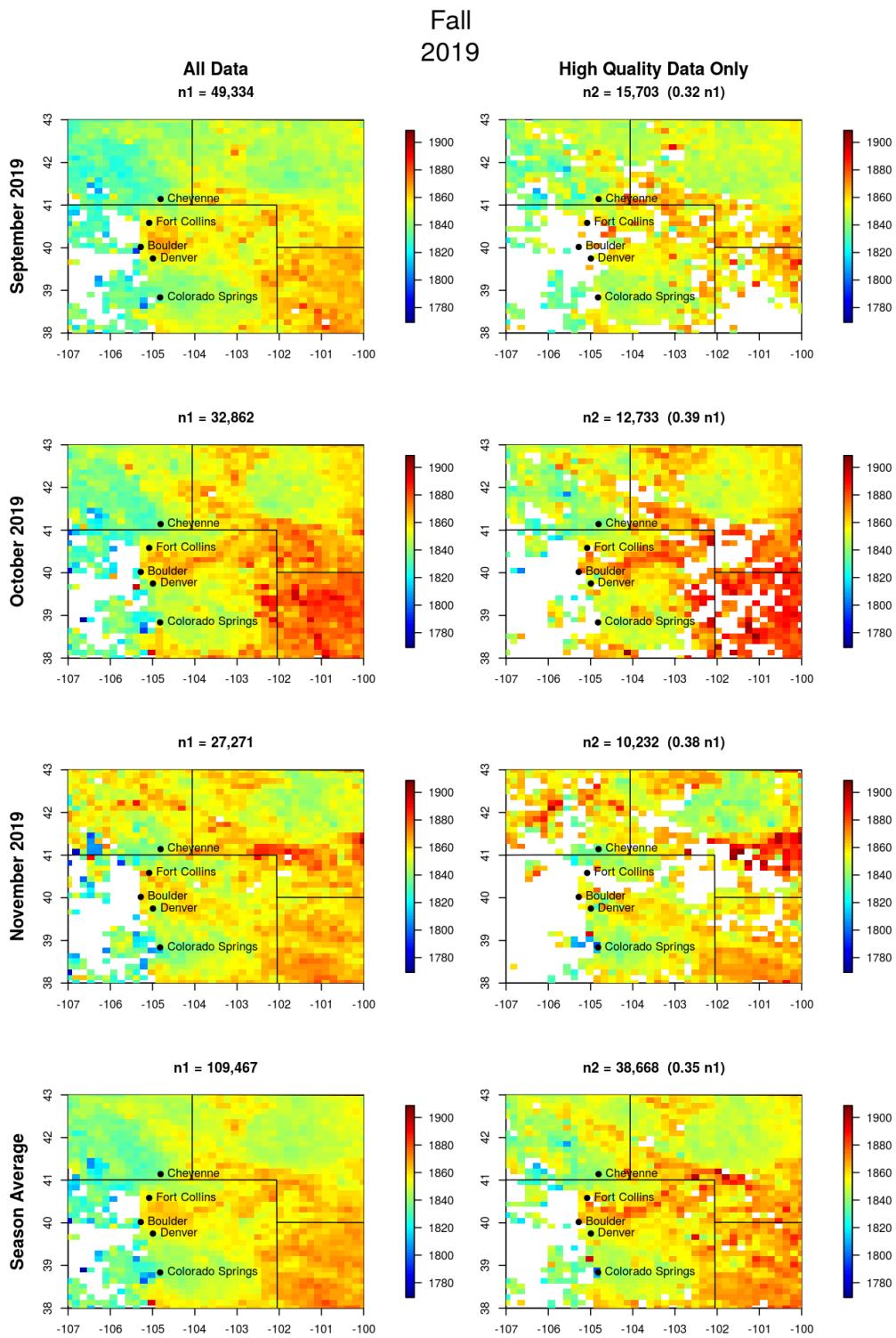
Figure 56: Fall 2019 monthly methane averages. Left column includes all data. Right column includes only high quality data. Bottom row is seasonal average.
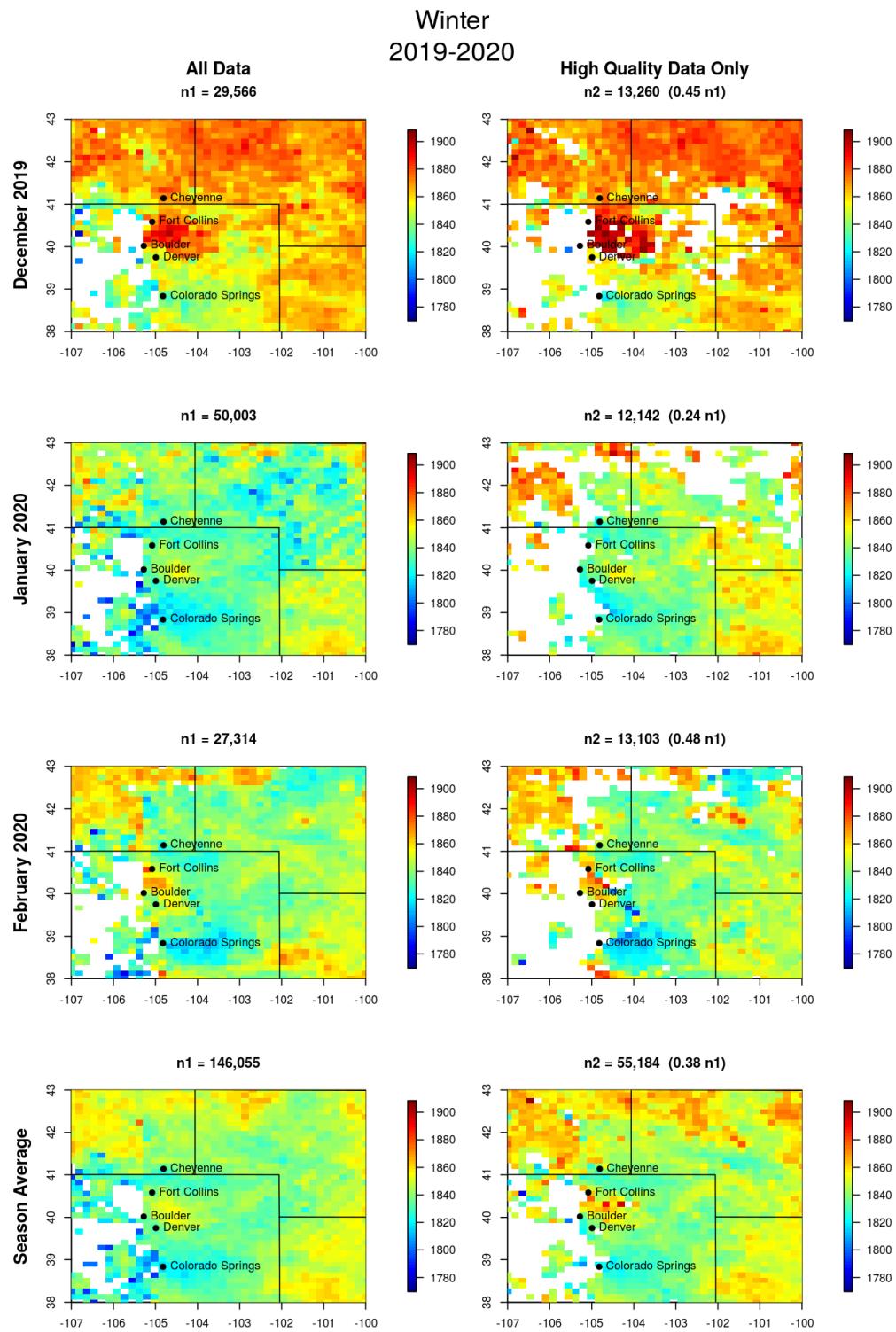
Figure 57: Winter 2019-2020 monthly methane averages. Left column includes all data. Right column includes only high quality data. Bottom row is seasonal average.

The Payne Institute for Public Policy

COLORADOSCHOOLOF**MINES**
EARTH ● ENERGY ● ENVIRONMENT

ABOUT THE AUTHORS

Dorit Hammerling
Associate Professor, Applied Mathematics and Statistics

   After 8 years working in the cement industry on process and quality control, Prof. Hammerling obtained a M.A. and PhD (2012) from the University of Michigan in Statistics and Engineering developing statistical methods for large satellite data. This was followed by a post-doctoral fellowship at the Statistical Applied Mathematical Sciences Institute in the program for Statistical Inference for massive data. Prof. Hammerling subsequently joined the National Center for Atmospheric Research, where she led the statistics group within the Institute for Mathematics Applied to the Geosciences and worked in the Machine Learning division before becoming an Associate Professor in the Department of Applied Mathematics and Statistics at the Colorado School of Mines in January 2019. Prof. Hammerling received the Early Investigator Award from the American Statistical Association, Section on Statistics and the Environment, in 2018.

William Daniels
Colorado School of Mines MS in Statistics

William Daniels is currently pursuing a MS in Statistics at the Colorado School of Mines. He recently obtained a BS in Engineering Physics from Mines, graduating with the Physics Faculty Distinguished Graduate Award in 2019. His work in the Physics Department involved modeling and simulation of certain transient luminous events in the upper atmosphere. His current work in the Applied Mathematics and Statistics Department involves statistical inference problems related to large satellite data.

Morgan D. Bazilian
Director, Payne Institute for Public Policy and Professor of Public

   Morgan Bazilian is the Director of the Payne Institute and a Professor of public policy at the Colorado School of Mines. Previously, he was lead energy specialist at the World Bank. He has over two decades of experience in the energy sector and is regarded as a leading expert in international affairs, policy and investment. He is a Member of the Council on Foreign Relations.

The Payne Institute *for* Public Policy

COLORADO SCHOOL OF MINES
EARTH ● ENERGY ● ENVIRONMENT

ABOUT THE PAYNE INSTITUTE

The mission of the Payne Institute at Colorado School of Mines is to provide world-class scientific insights, helping to inform and shape public policy on earth resources, energy, and environment. The Institute was established with an endowment from Jim and Arlene Payne, and seeks to link the strong scientific and engineering research and expertise at Mines with issues related to public policy and national security.

The Payne Institute Commentary Series offers independent insights and research on a wide range of topics related to energy, natural resources, and environmental policy. The series accommodates three categories namely: Viewpoints, Essays, and Working Papers.

For more information about the Payne Institute please visit:
https://payneinstitute.mines.edu/

or follow us on Twitter or LinkedIn: