Visualising daily PM10 pollution data in an open-cut mining valley of New South Wales, Australia - Part I: identification of spatial and temporal variation patterns

Ningbo Jiang1, Matthew L. Riley1, Merched Azzi1, Praveen Puppala1, Hiep Duc1, Giovanni Di Virgilio1

1 New South Wales Department of Climate Change, Energy, the Environment and Water, Sydney, Australia

# Supplementary Materials

The following supporting information can be downloaded at [**https://www.mdpi.com/article/**XXX](https://www.mdpi.com/article/XXX):

Figure S1 Box plots of daily PM10 data by station for data in 2012–2022 (including exceptional event days). The lower and upper boundaries of the box are respectively the 25th and 75th percentile; the horizontal line inside the box represents the median; asterisks represent extreme values, cases with values more than 3 box-lengths from the upper or lower edge of the box; dots denote outliers, cases with values between 1.5 and 3 box-lengths from the upper or lower edge of the box; horizontal lines connected to two ends of the box correspond to the largest or smallest observed values that are not outliers. Red dashed line shows the Australian national standard of 50 µg/m3 for daily PM10.

Figure S2 Identification of two air quality clusters/subregions in the Upper Hunter Valley based on Varimax rotated principal component analysis (RPCA) on daily PM10 data for 2012-2022 (including exceptional event days). Left panel: key of station number; middle panel: scatter plot of loadings for first two rotated principal components (PC1, PC2); right panel: map showing two air quality subregions with red balloons indicating station locations in the UHAQMN. Base map source: Google.

Figure S3 SE subregion temporal variability patterns. (a) The first principal component (PC1) scores used for the wavelet analysis, derived from RPCA of all-day dataset (exceptional event data included) where missing data were replaced with overall median for each station. (b) The local normalised wavelet power spectrum of (a) using the Morlet wavelet. The contour lines are at normalised variances of low to high values shown in dark (blue) to bright (light) colours. The thick black contour encloses regions of greater than 95% confidence for a red-noise process with lag-1 coefficient. Regions under the bowl-shape curve on either end indicates the “cone of influence”, where edge effects become important. (c) The scale-averaged wavelet power (variance) over the 1-365 days band for PC1 scores.

Figure S4 WNW subregion temporal variability patterns. (a) The second principal component (PC2) scores used for the wavelet analysis, derived from of all-day dataset (exceptional event data included) where missing data were replaced with overall median for each station. (b) The local normalised wavelet power spectrum of (a) using the Morlet wavelet. The contour lines are at normalised variances of low to high values shown in dark (blue) to bright (light) colours. The thick black contour encloses regions of greater than 95% confidence for a red-noise process with lag-1 coefficient. Regions under the bowl-shape curve on either end indicates the “cone of influence”, where edge effects become important. (c) The scale-averaged wavelet power (variance) over the 1-365 days band for PC2 scores.

Figure S5 Monthly mean PM10 levels and total number of poor air quality days (with PM10 levels > 50 µg/m3) for stations in the (a) WNW and (b) SE subregions. Rows are sorted by the “All months” column (multi-year station means or count of poor air quality days). Data: daily PM10 measurements in 2012–2022 (including exceptional events). Colour scale: green – relatively low value; yellow – near medium value; red – relatively high value.

Figure S6 Annual mean PM10 levels and total number of poor air quality days (with PM10 levels > 50 µg/m3) for stations in the (a) WNW and (b) SE subregions. Rows are sorted by the “All years” column (multi-year station means or total number of poor air quality days). Data: daily PM10 measurements in 2012–2022 (including exceptional event days). Colour scale: green - relatively low value; yellow – near medium value; red - relatively high value.

# Appendix Supplementary Materials

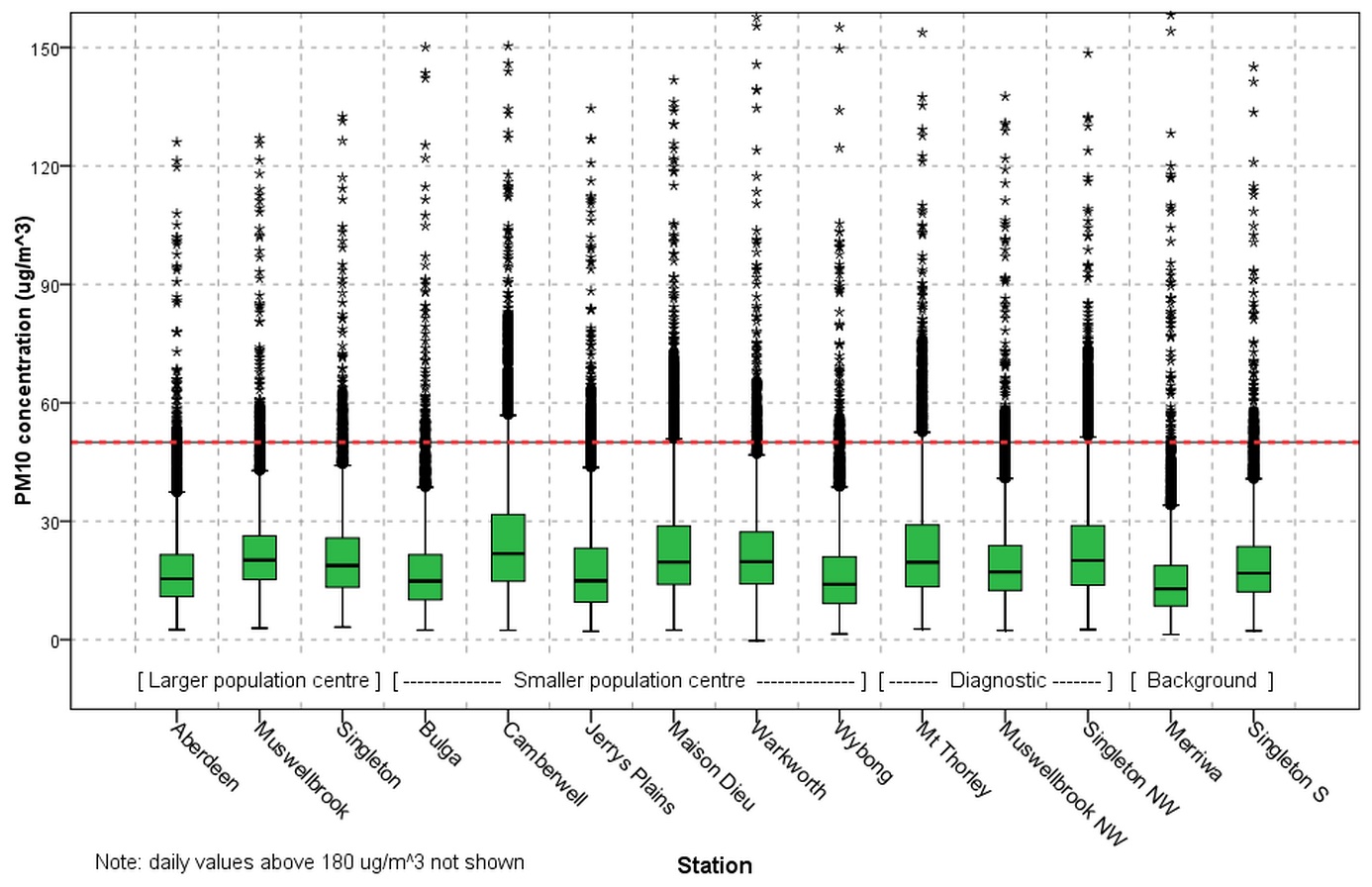


Figure S Box plots of daily PM10 data by station for data in 2012–2022 (including exceptional event days). The lower and upper boundaries of the box are respectively the 25th and 75th percentile; the horizontal line inside the box represents the median; asterisks represent extreme values, cases with values more than 3 box-lengths from the upper or lower edge of the box; dots denote outliers, cases with values between 1.5 and 3 box-lengths from the upper or lower edge of the box; horizontal lines connected to two ends of the box correspond to the largest or smallest observed values that are not outliers. Red dashed line shows the Australian national standard of 50 µg/m3 for daily PM10.

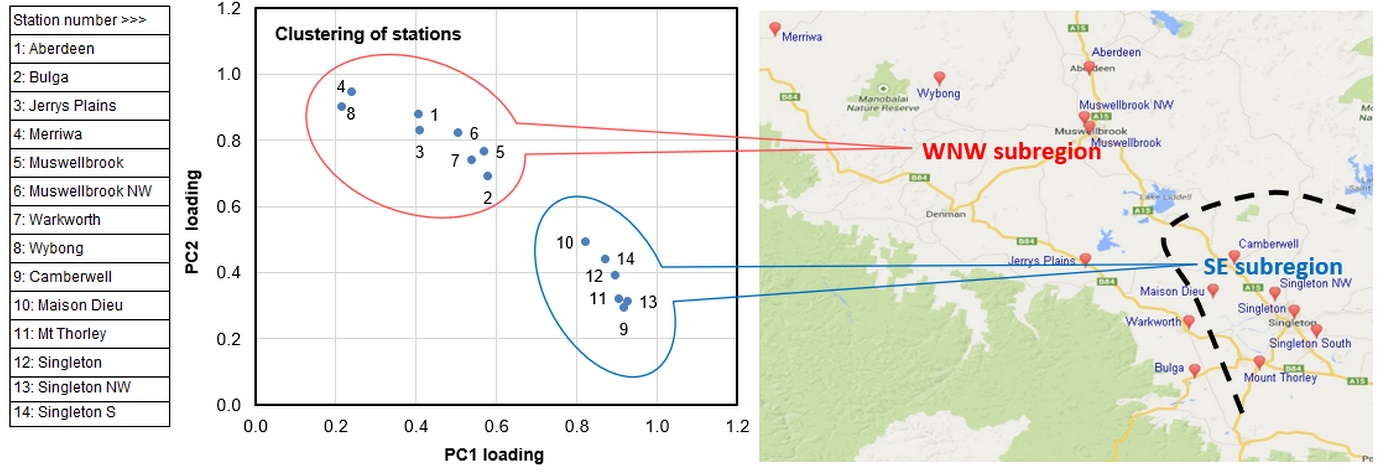


Figure S Identification of two air quality clusters/subregions in the Upper Hunter Valley based on Varimax rotated principal component analysis (RPCA) on daily PM10 data for 2012-2022 (including exceptional event days). Left panel: key of station number; middle panel: scatter plot of loadings for first two rotated principal components (PC1, PC2); right panel: map showing two air quality subregions with red balloons indicating station locations in the UHAQMN. Base map source: Google.

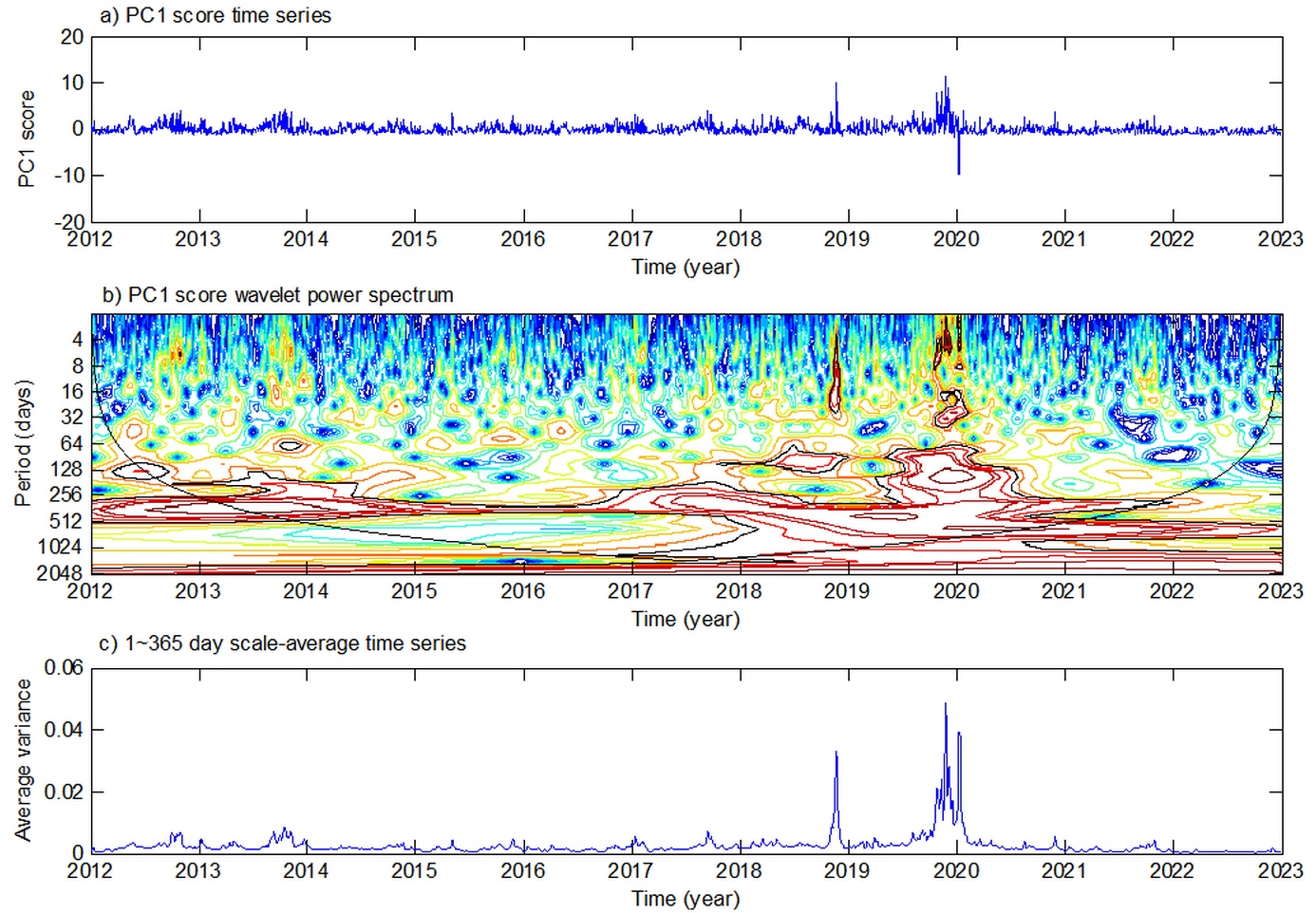


Figure S SE subregion temporal variability patterns. (a) The first principal component (PC1) scores used for the wavelet analysis, derived from RPCA of all-day dataset (exceptional event data included) where missing data were replaced with overall median for each station. (b) The local normalised wavelet power spectrum of (a) using the Morlet wavelet. The contour lines are at normalised variances of low to high values shown in dark (blue) to bright (light) colours. The thick black contour encloses regions of greater than 95% confidence for a red-noise process with lag-1 coefficient. Regions under the bowl-shape curve on either end indicates the “cone of influence”, where edge effects become important. (c) The scale-averaged wavelet power (variance) over the 1-365 days band for PC1 scores.

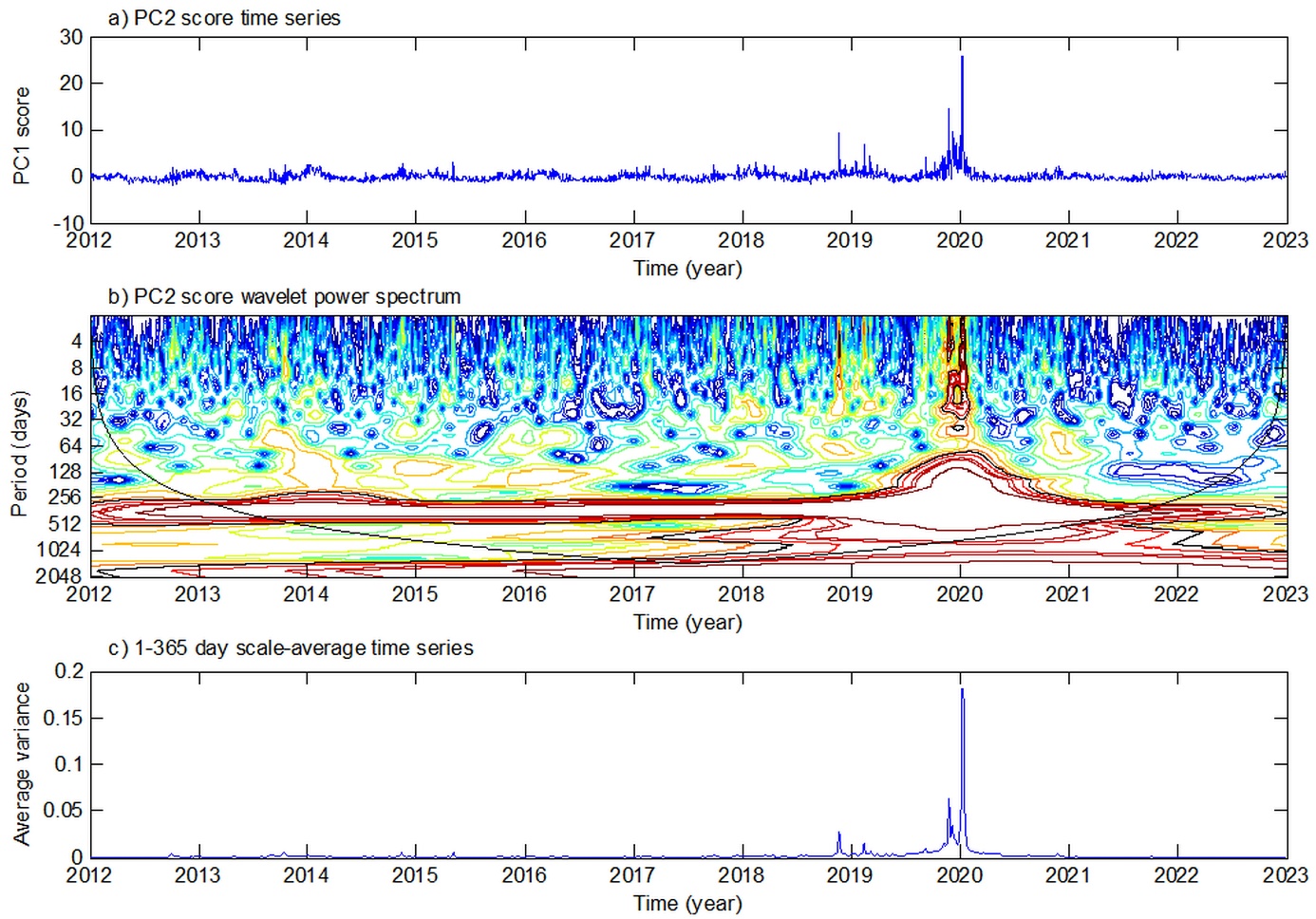


Figure S WNW subregion temporal variability patterns. (a) The second principal component (PC2) scores used for the wavelet analysis, derived from of all-day dataset (exceptional event data included) where missing data were replaced with overall median for each station. (b) The local normalised wavelet power spectrum of (a) using the Morlet wavelet. The contour lines are at normalised variances of low to high values shown in dark (blue) to bright (light) colours. The thick black contour encloses regions of greater than 95% confidence for a red-noise process with lag-1 coefficient. Regions under the bowl-shape curve on either end indicates the “cone of influence”, where edge effects become important. (c) The scale-averaged wavelet power (variance) over the 1-365 days band for PC2 scores.

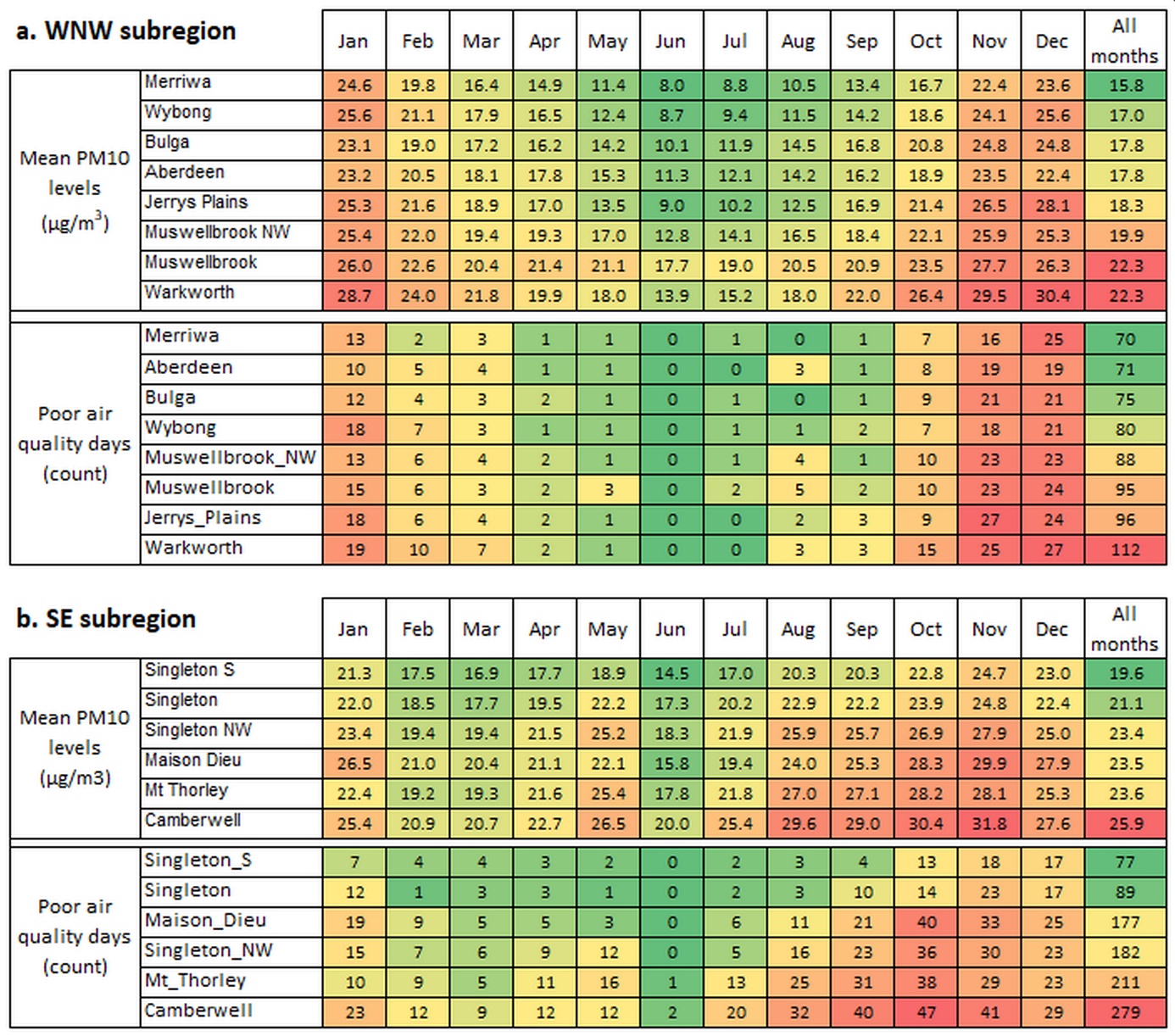


Figure S Monthly mean PM10 levels and total number of poor air quality days (with PM10 levels > 50 µg/m3) for stations in the (a) WNW and (b) SE subregions. Rows are sorted by the “All months” column (multi-year station means or count of poor air quality days). Data: daily PM10 measurements in 2012–2022 (including exceptional events). Colour scale: green – relatively low value; yellow – near medium value; red – relatively high value.

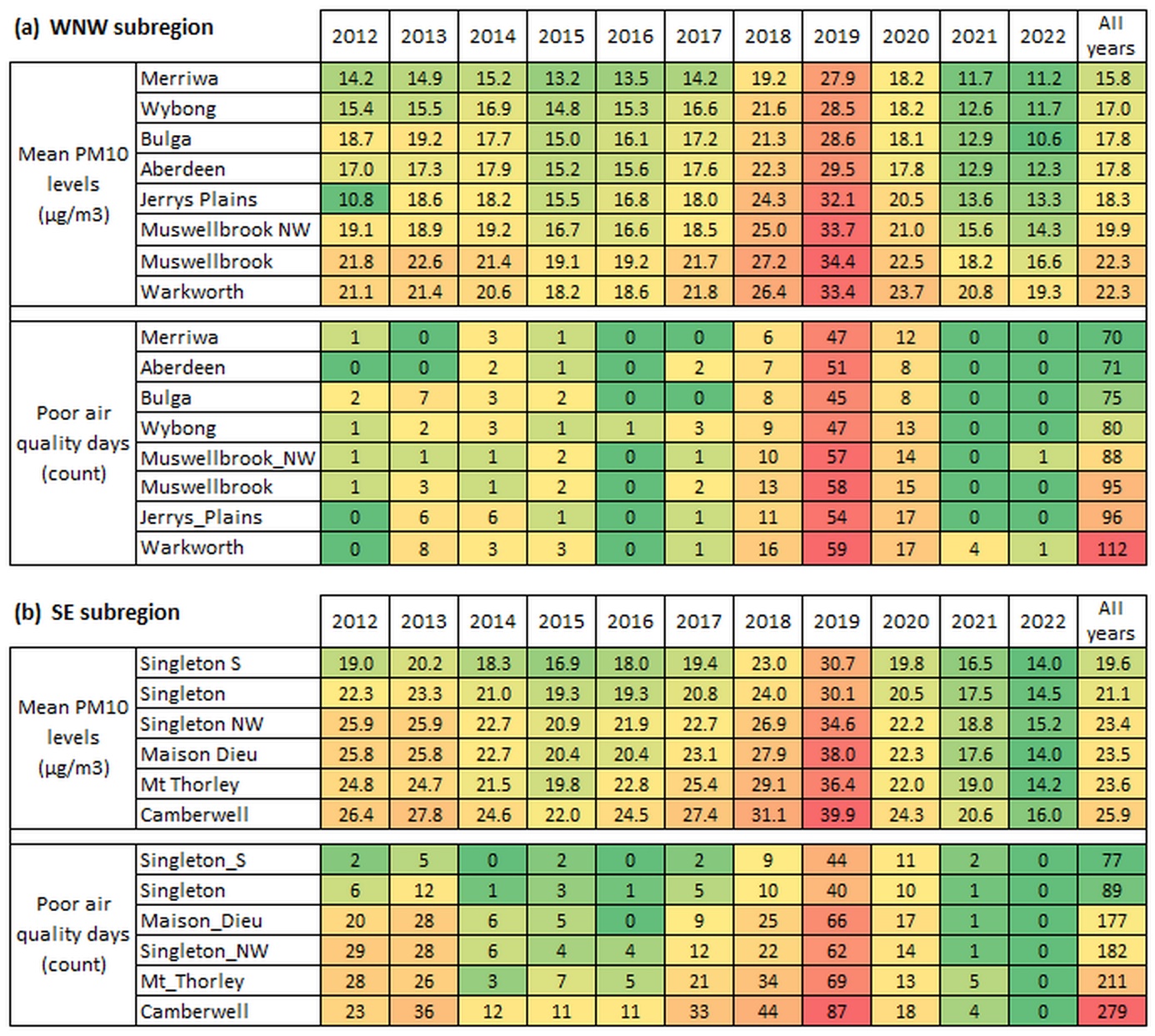


Figure S Annual mean PM10 levels and total number of poor air quality days (with PM10 levels > 50 µg/m3) for stations in the (a) WNW and (b) SE subregions. Rows are sorted by the “All years” column (multi-year station means or total number of poor air quality days). Data: daily PM10 measurements in 2012–2022 (including exceptional event days). Colour scale: green - relatively low value; yellow – near medium value; red - relatively high value.