atmosphere

Supplementary Material for Modelling Cyclists' Multi-Exposure to Air and Noise Pollution with Low-Cost Sensors—The Case of Paris

Jérémy Gelb and Philippe Apparicio *

Institut National de la Recherche Scientifique, Centre Urbanisation Culture Société, Montréal, QC H2X 1E3, Canada ; jeremy.gelb@ucs.inrs.ca
* Correspondence: philippe.apparicio@ucs.inrs.ca

1. Model Priors

All of the selected priors are "weakly informative". This is justified by the lack of information about the expected values for the studied parameters in Paris and by the fairly large amount of data, so strong priors are not required.

For all the linear coefficients predicting NO_2, we have selected a normal prior with a mean of 0 and a standard deviation of 10 (normal(0,10)). For noise, the same coefficients have a normal prior also centered on 0 with a standard deviation of 2 (normal(0,2)). These priors are conservative because they are centered on 0 and are used to help the NUT (No-U-Turn) sampler to converge without sampling meaningless values. Indeed, it is very unlikely that one of these parameters has an impact greater than 30 µg/m3 or 6 dB(A).

For the random effects, we followed the recommendation of Gelman [1] and used proper priors with a half-Cauchy distribution. The scale parameters for both day and participants groups for NO_2 are 20 and the scale parameters for noise exposure are 2. This reflects prior knowledge that NO_2 might be more variable across days and sensors than noise.

In the R package named *brms* (Bayesian Regression Models using 'Stan'), spline variability is modeled as a random effect due to the proximity between them. Therefore, the same types of priors are used (half-Cauchy). For NO_2, the scale parameters selected are 20 (distance from road), 20 (Time) and 35 (X and Y coordinates); for noise, they are 8, 8 and 16 respectively. This reflects the prior knowledge that for both noise and NO_2, the space parameter might have more variations than time of the day or distance from roads and that NO_2 concentrations have a higher variance than noise levels.

Finally, for the MA (Moving Average) terms, we specified a prior as normal (0.6,0.2) with a lower restriction of -1 and an upper restriction of 1 reflecting the prior knowledge that both data series are strongly autocorrelated.

2. Calculation of Urban Morphological Indexes

Slope (%)

Data source: the elevation is reported by the Garmin watch every second
Calculation: (Altitude at the end of the segment – Altitude at the beginning of the segment) / Segment length * 100

Low speed area (0-1)

Data source: https://opendata.paris.fr/explore/dataset/zones-30
Calculation: geographical intersection with the segment

Industrial activity land use density (%)

Data source: http://data.iau-idf.fr/datasets/75ab6d1bb58043729a8b05172937cfb0_40
Calculation: Part of a 50m buffer around the segment that is covered by industrial activities

Vegetation density (%)

Data source: https://opendata.apur.org/datasets/hauteur-vegetation-2015
Calculation: Part of a 50m buffer around the segment that is covered by tree canopy

Sky view factor index (%)

Data source: https://opendata.paris.fr/explore/dataset/volumesbatisparis/information/
Calculation: SAGA GIS version 2.3.2
Interpretation: part of the sky visible when the obstruction caused by buildings is considered. Proxy for the canyon shape of a street.

Fetch Index (%)

Usually defined as: the distance the wind can travel across a specific environment
Wind data source: https://www.infoclimat.fr
Building data source : https://opendata.paris.fr/explore/dataset/volumesbatisparis/information/
Calculation: the index is calculated as described by Figure S1:

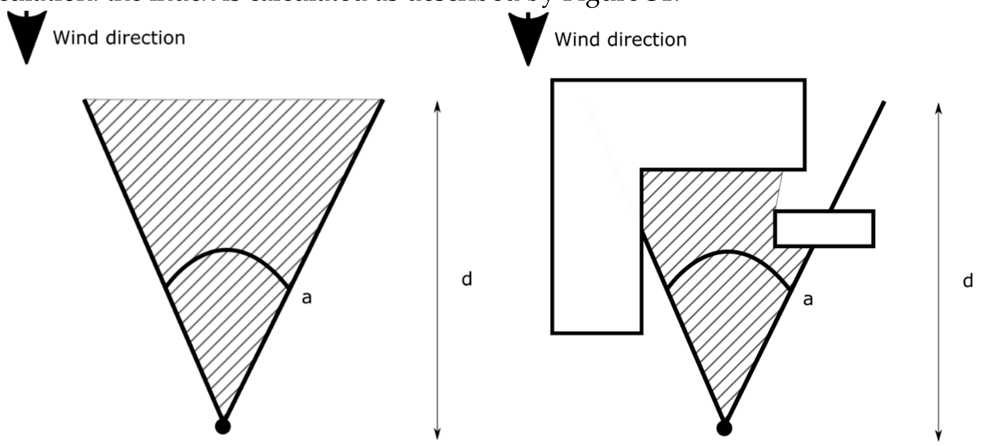


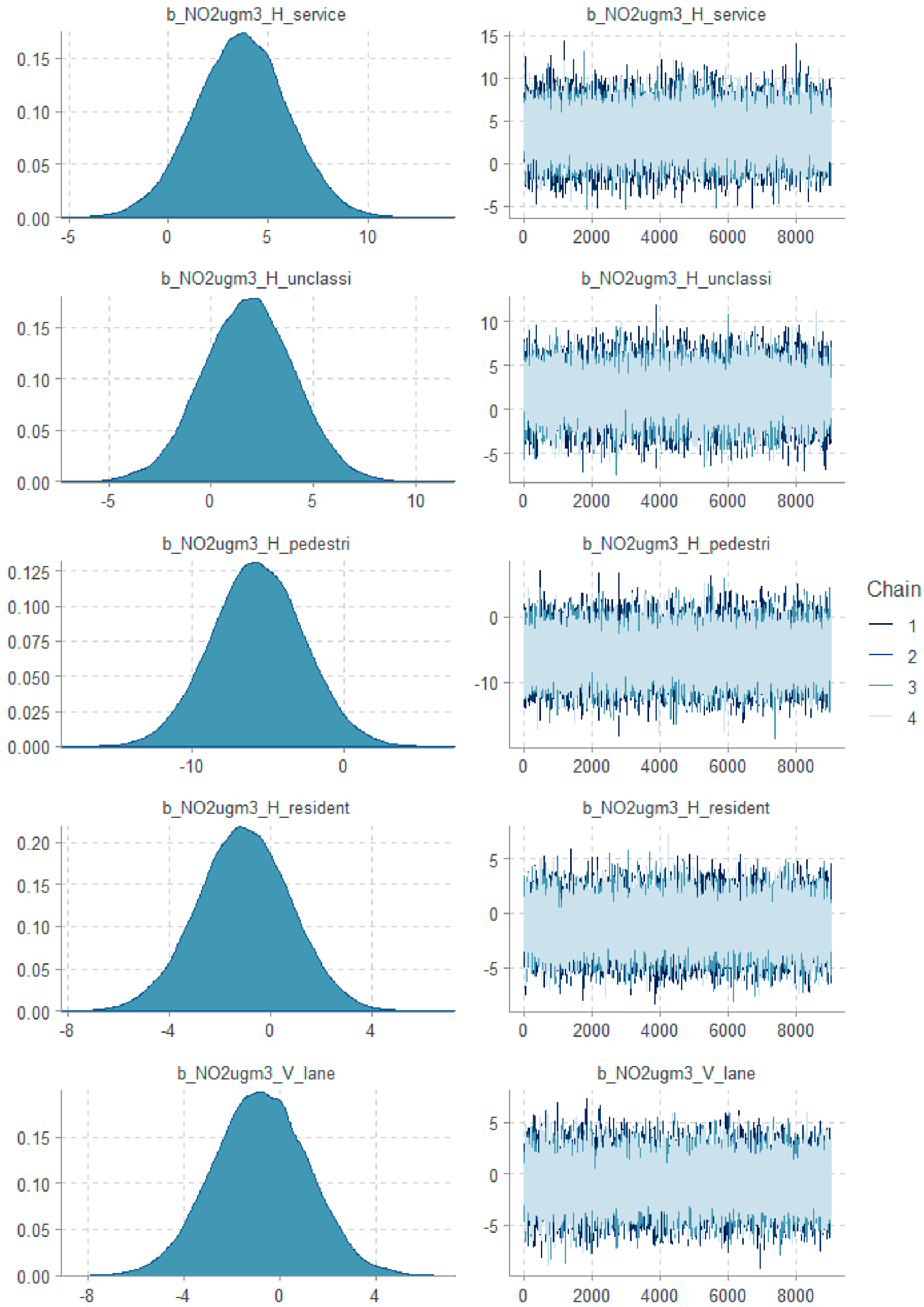
Figure S1. Visual description of the fetch index.

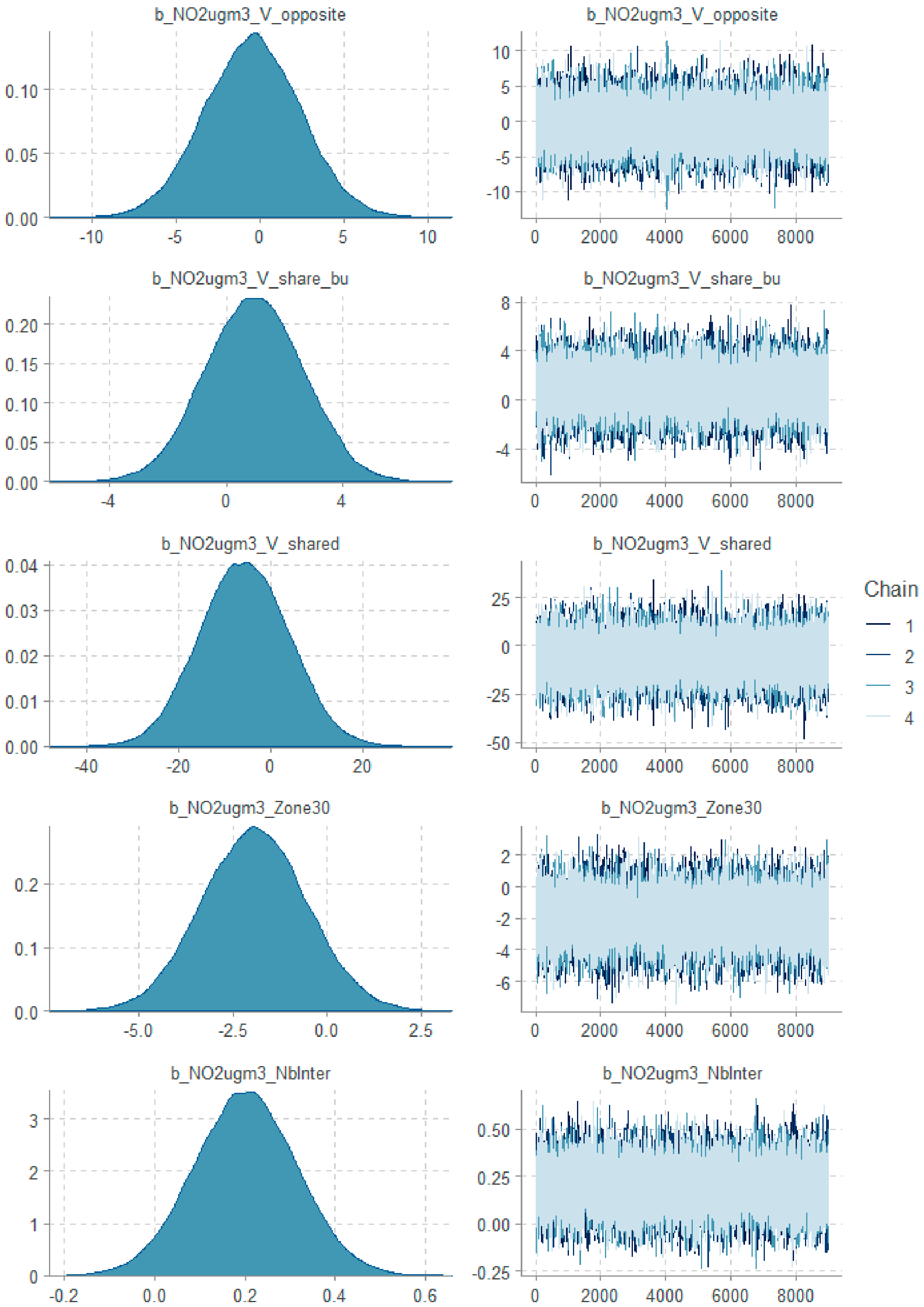
The inverse direction of the blowing wind, at point *p*, a cone is drawn with a length *d* (250m in the article), and an angle *a* (30 degrees in the article). The value of the index is the part of the cone's area that a line coming from *p* can cover in the cone without crossing a building's footprint. This represents the potential exposure to wind. As an example, in the first case of the figure above, the value of the index is 1 because no buildings impact the potential wind exposure.

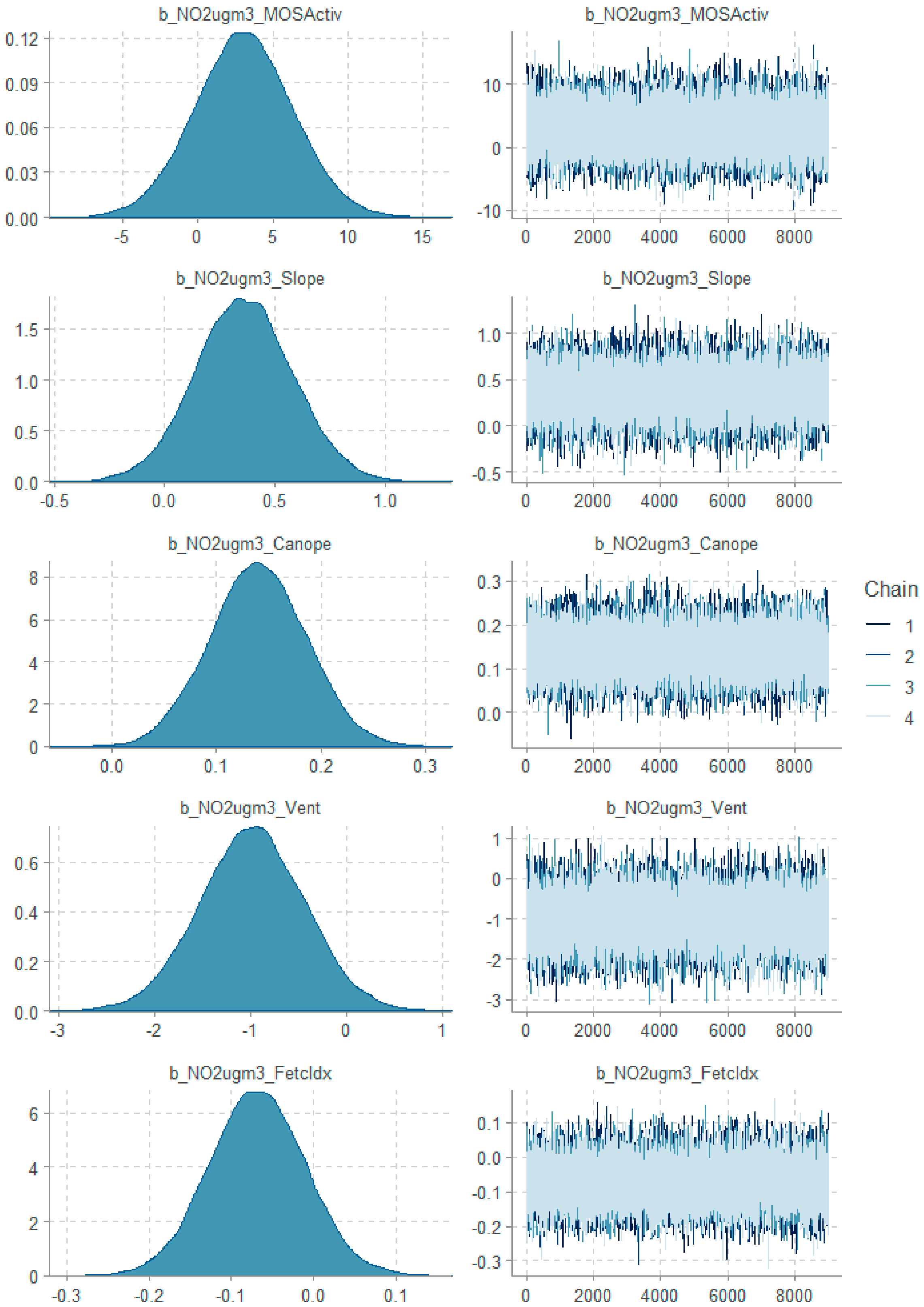
The value was calculated along segments every 10 meters and then averaged for the full segment.

3. Posterior Distributions

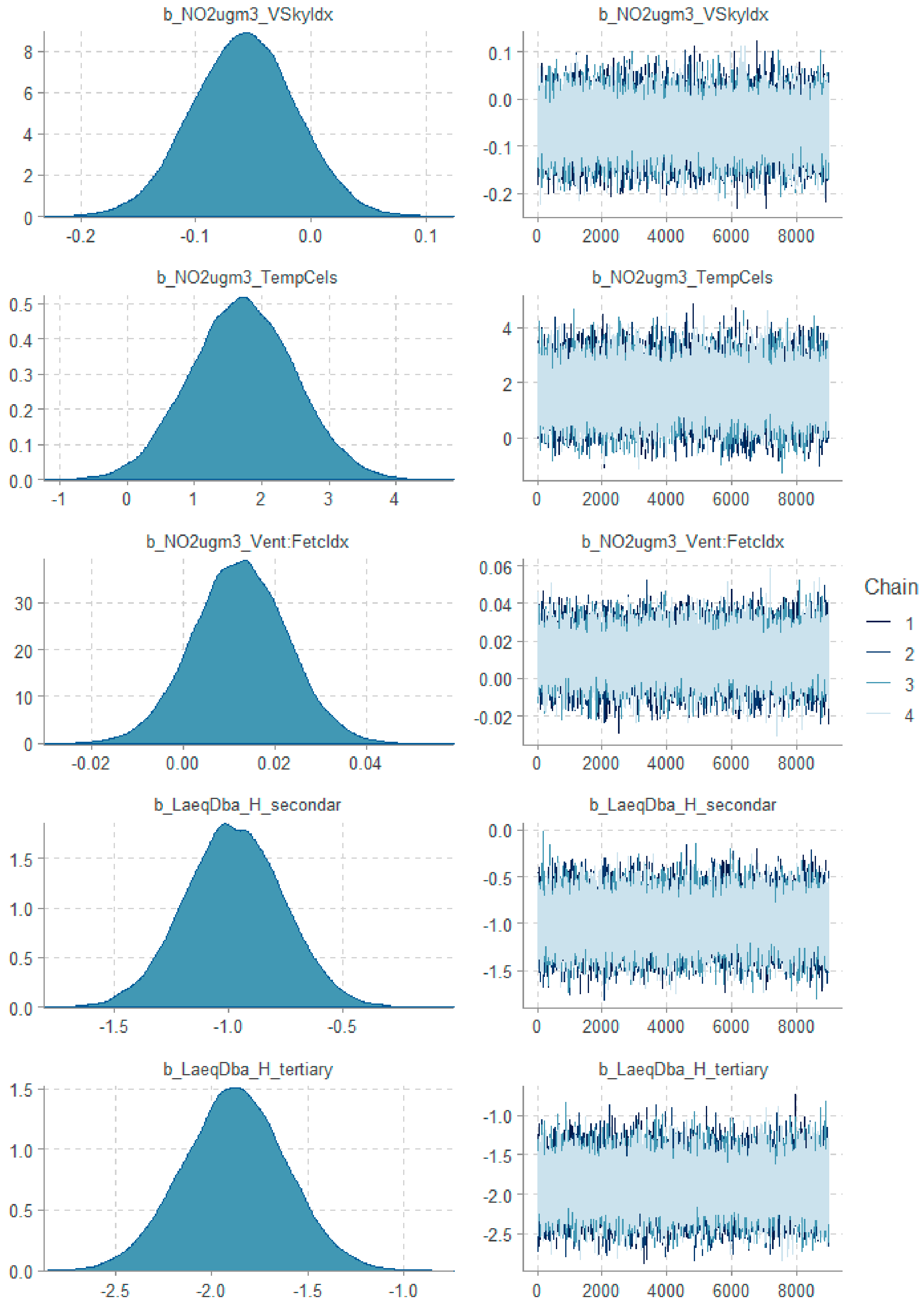
The posterior distribution of each parameter of the model are presented in Figure S2. The are obtained with four chains during 10,000 iterations where the first 1000 were used as a warmup for sampling realized with a No-U-Turn Sampler (NUTS).


b_NO2ugm3_H_service
b_NO2ugm3_H_unclassi
b_NO2ugm3_H_pedestri
b_NO2ugm3_H_resident
b_NO2ugm3_V_lane
Chain
1
2
3
4

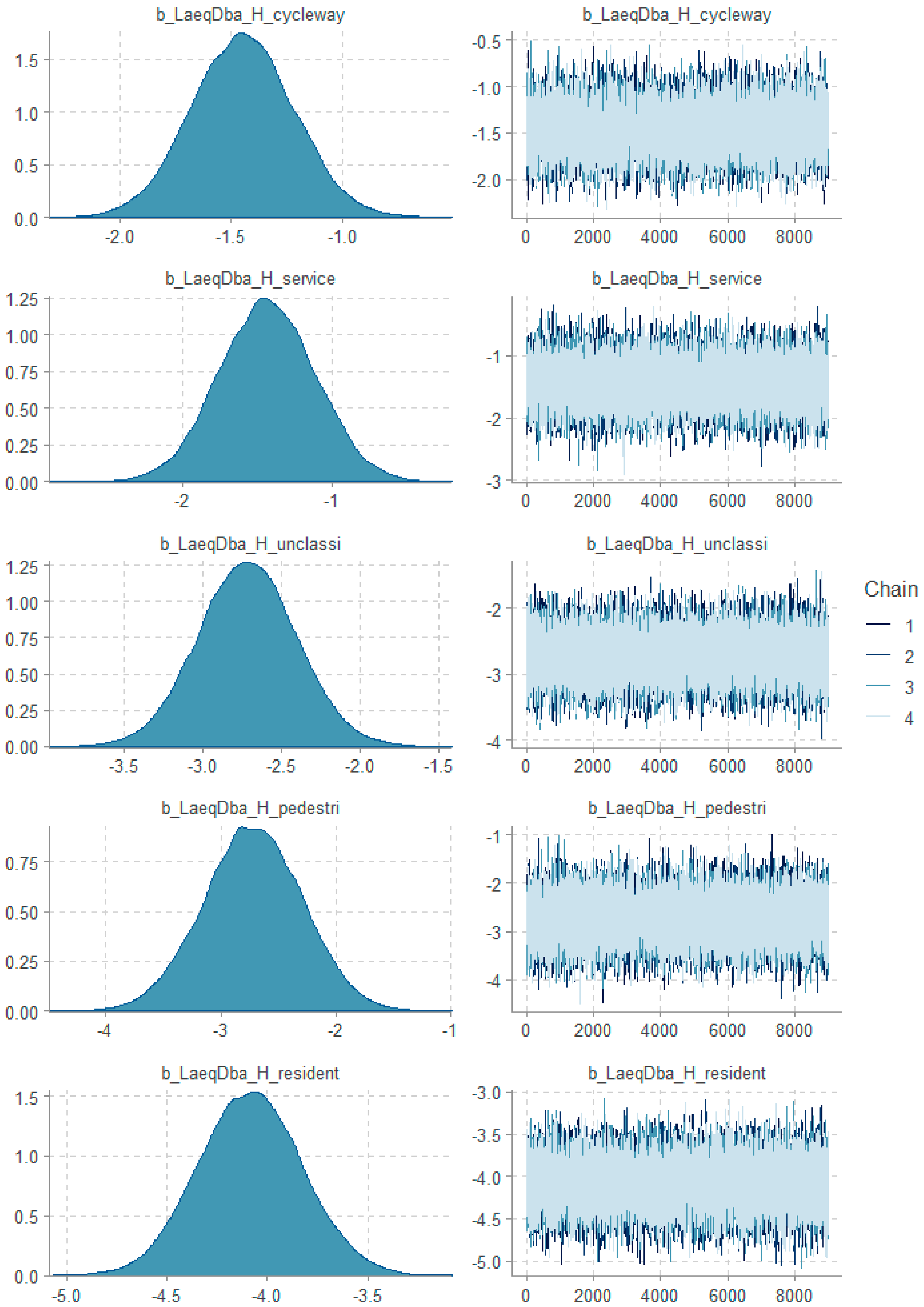

b_NO2ugm3_V_opposite
b_NO2ugm3_V_share_bu
b_NO2ugm3_V_shared
b_NO2ugm3_Zone30
b_NO2ugm3_NbInter
Chain
1
2
3
4



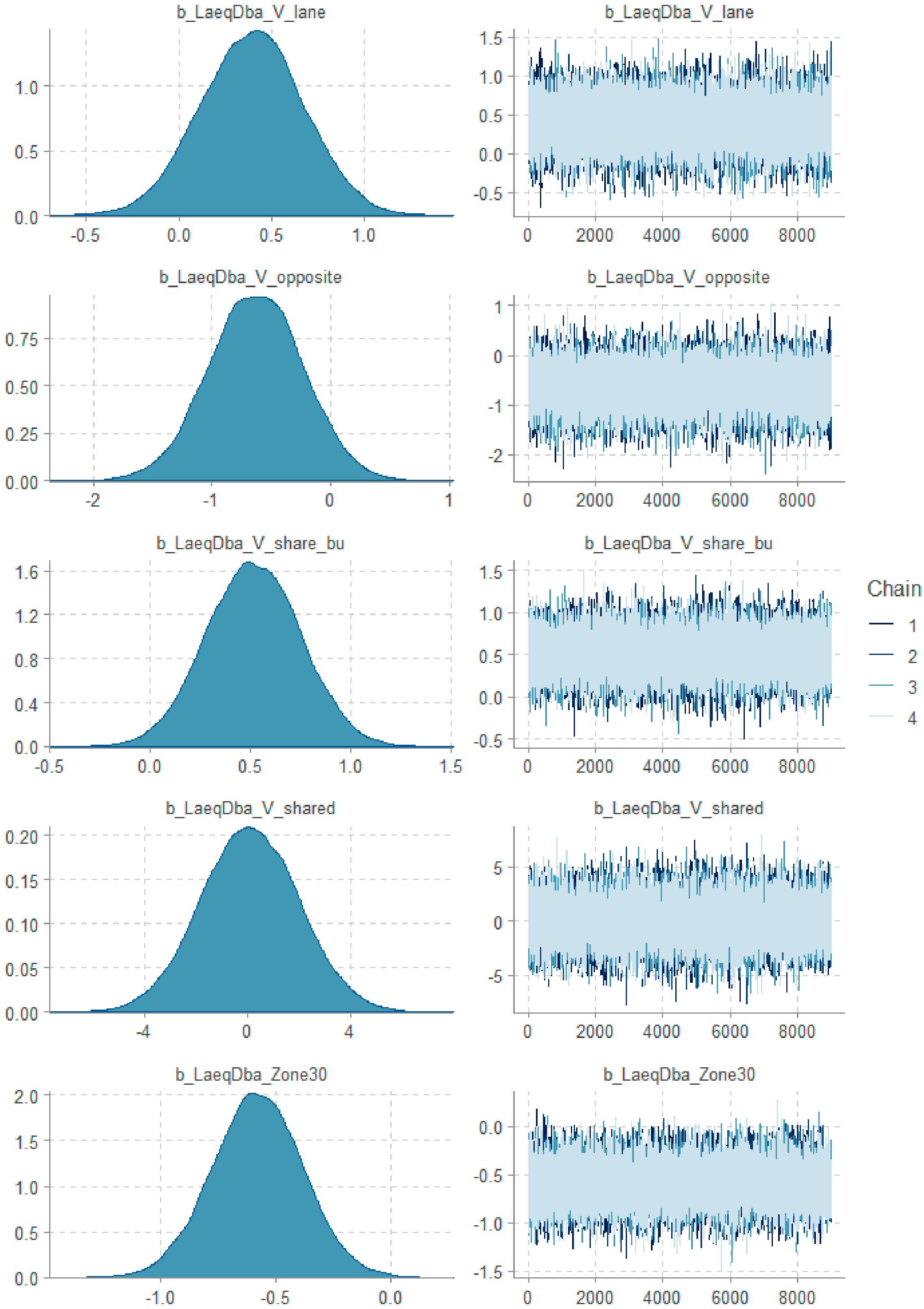
b_NO2ugm3_MOSActiv
b_NO2ugm3_Slope
b_NO2ugm3_Canope
b_NO2ugm3_Vent
b_NO2ugm3_FetcIdx
Chain
1
2
3
4

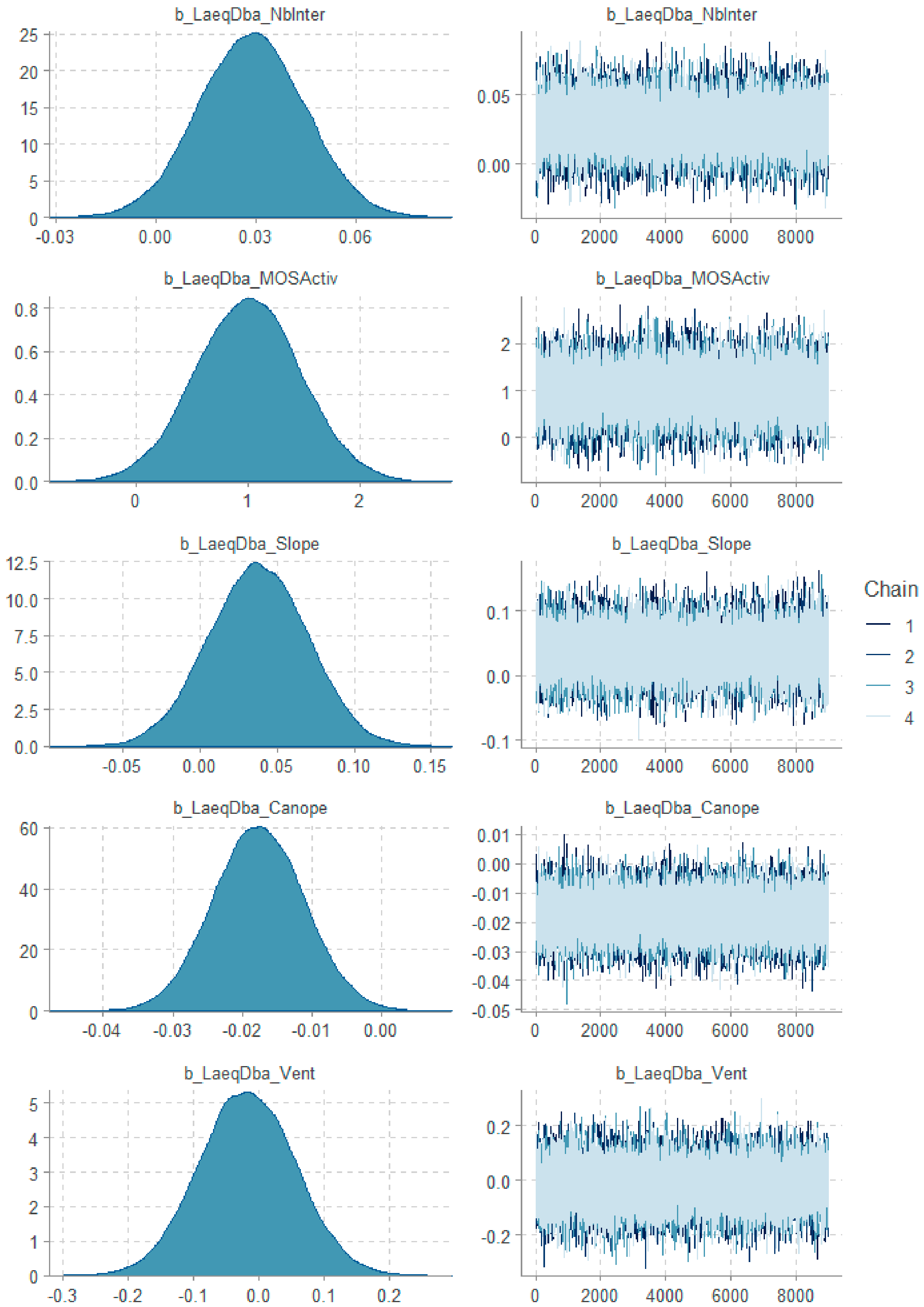


b_NO2ugm3_VSkyIdx
b_NO2ugm3_TempCels
b_NO2ugm3_Vent:FetcIdx
b_LaeqDba_H_secondar
b_LaeqDba_H_tertiary
Chain
1
2
3
4

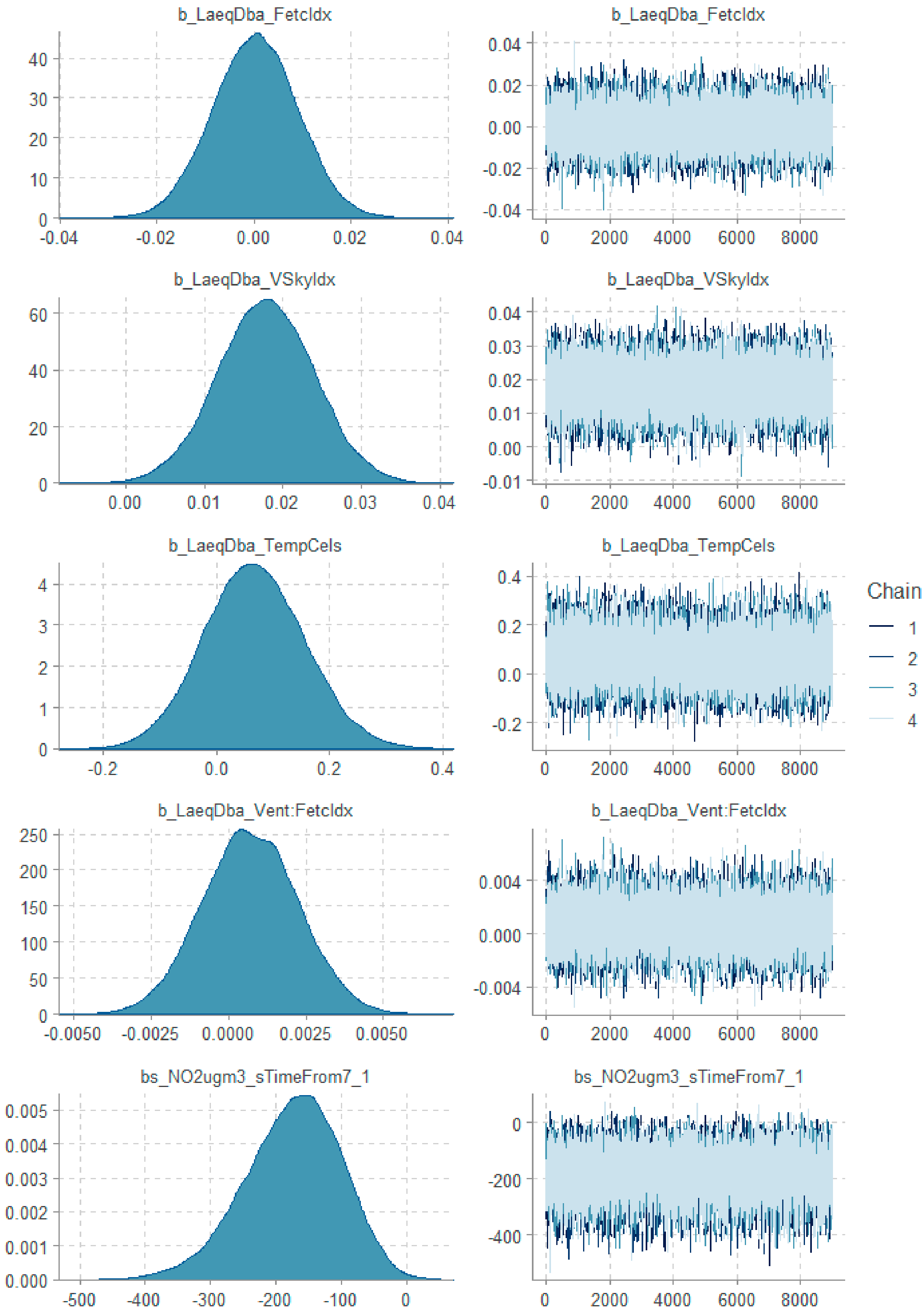


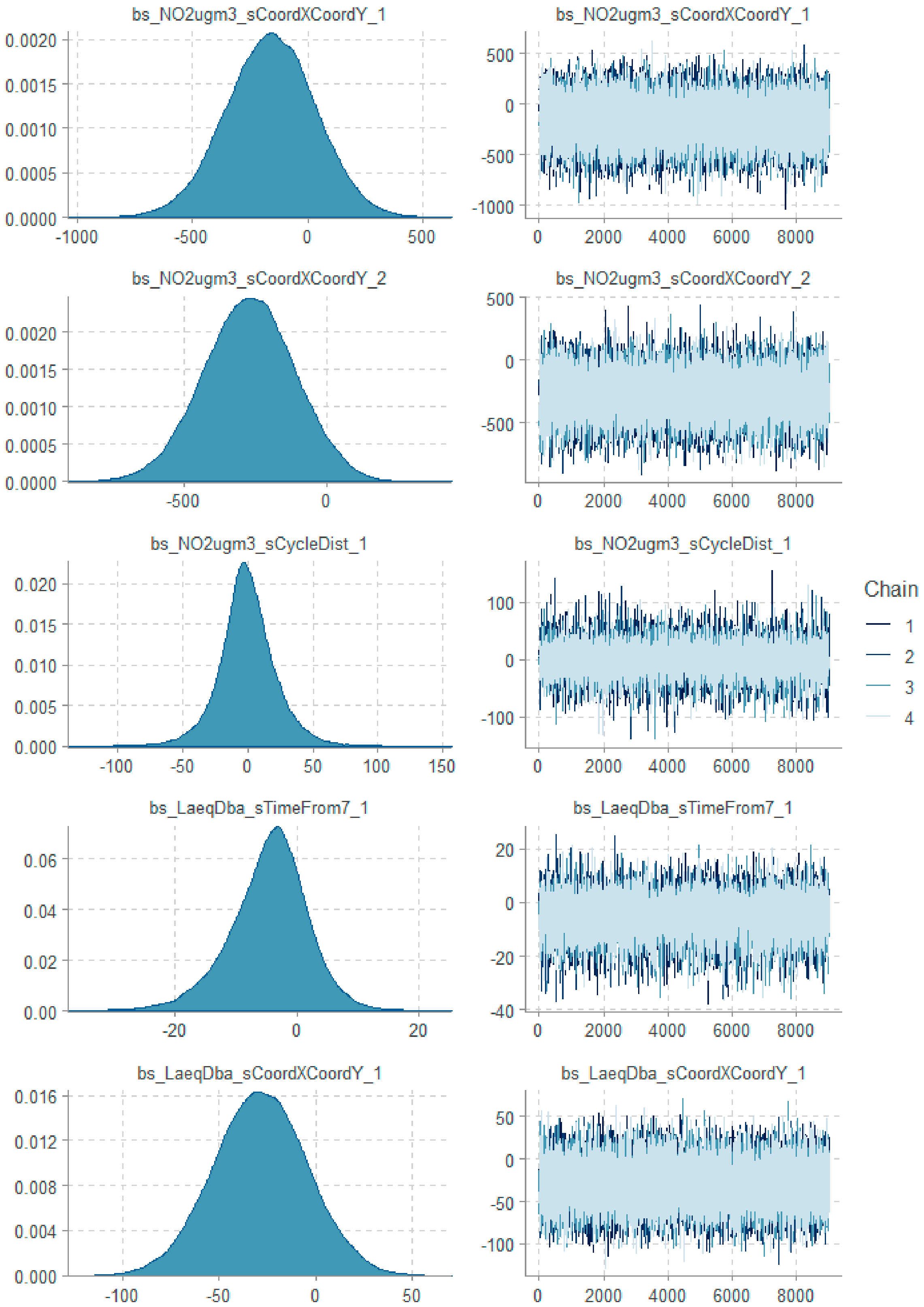
b_LaeqDba_H_cycleway
b_LaeqDba_H_service
b_LaeqDba_H_unclassi
b_LaeqDba_H_pedestri
b_LaeqDba_H_resident
Chain
1
2
3
4

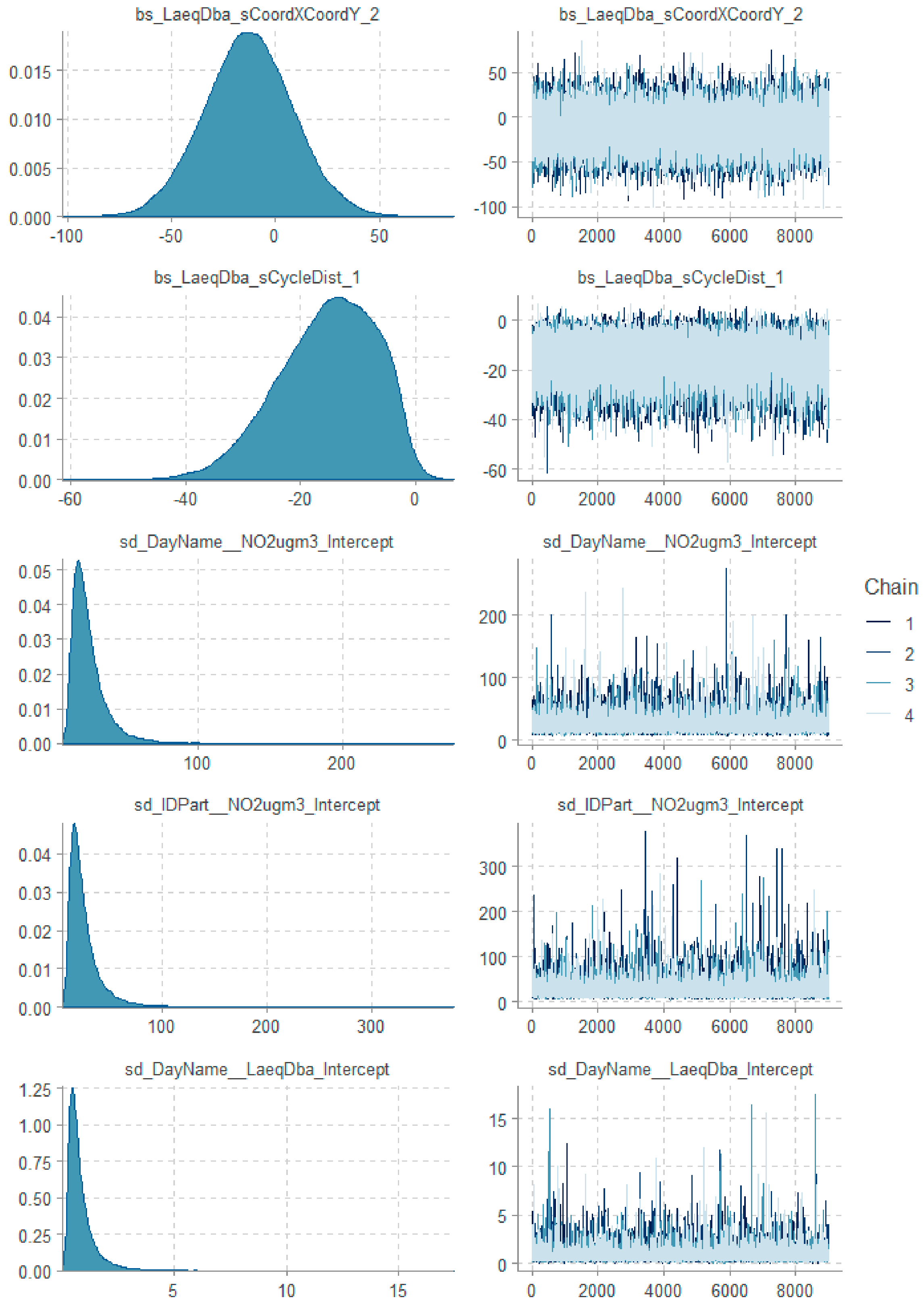

b_LaeqDba_V_lane
b_LaeqDba_V_opposite
b_LaeqDba_V_share_bu
b_LaeqDba_V_shared
b_LaeqDba_Zone30
Chain
1
2
3
4



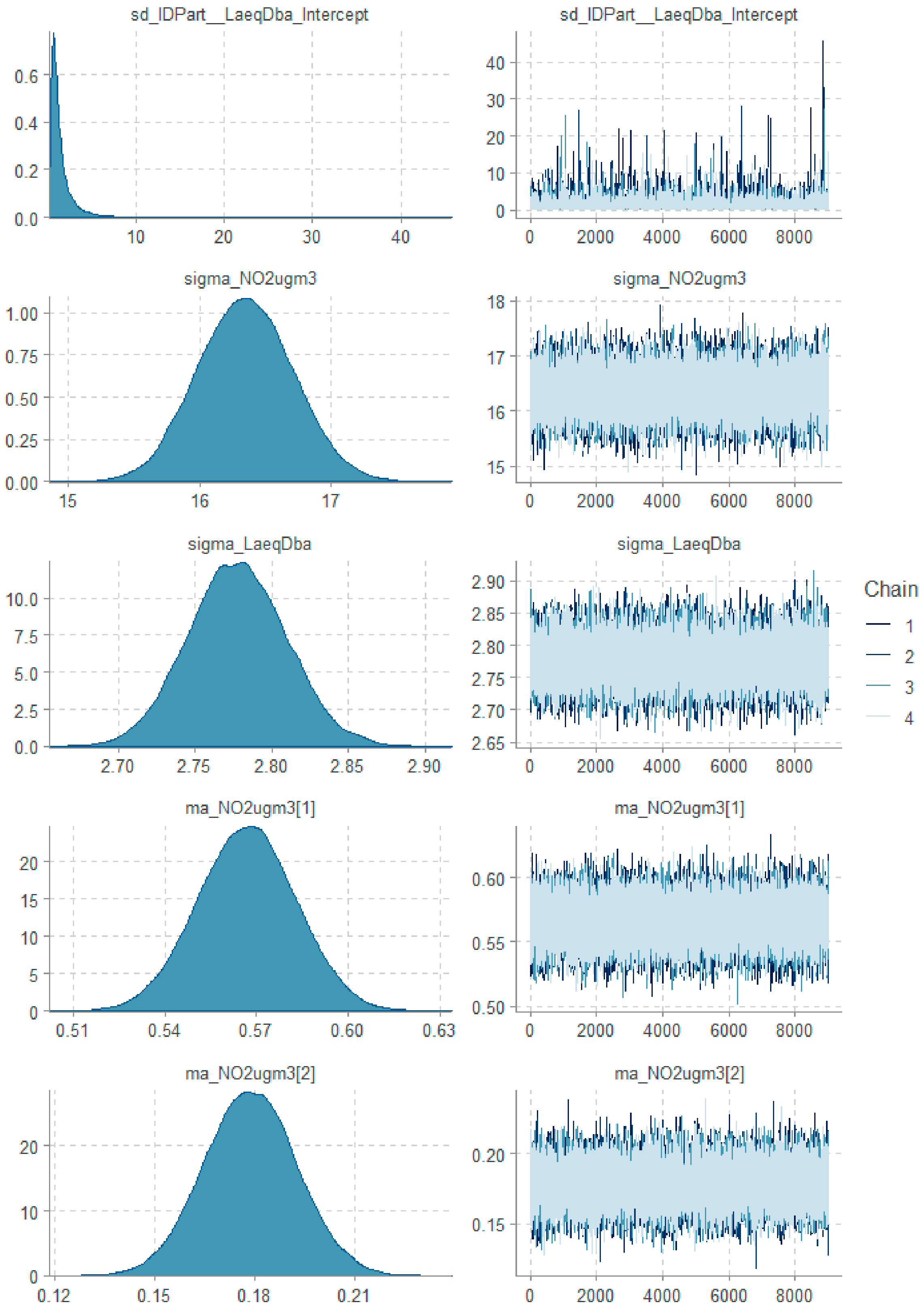
b_LaeqDba_NbInter
b_LaeqDba_MOSActiv
b_LaeqDba_Slope
b_LaeqDba_Canope
b_LaeqDba_Vent
Chain
1
2
3
4

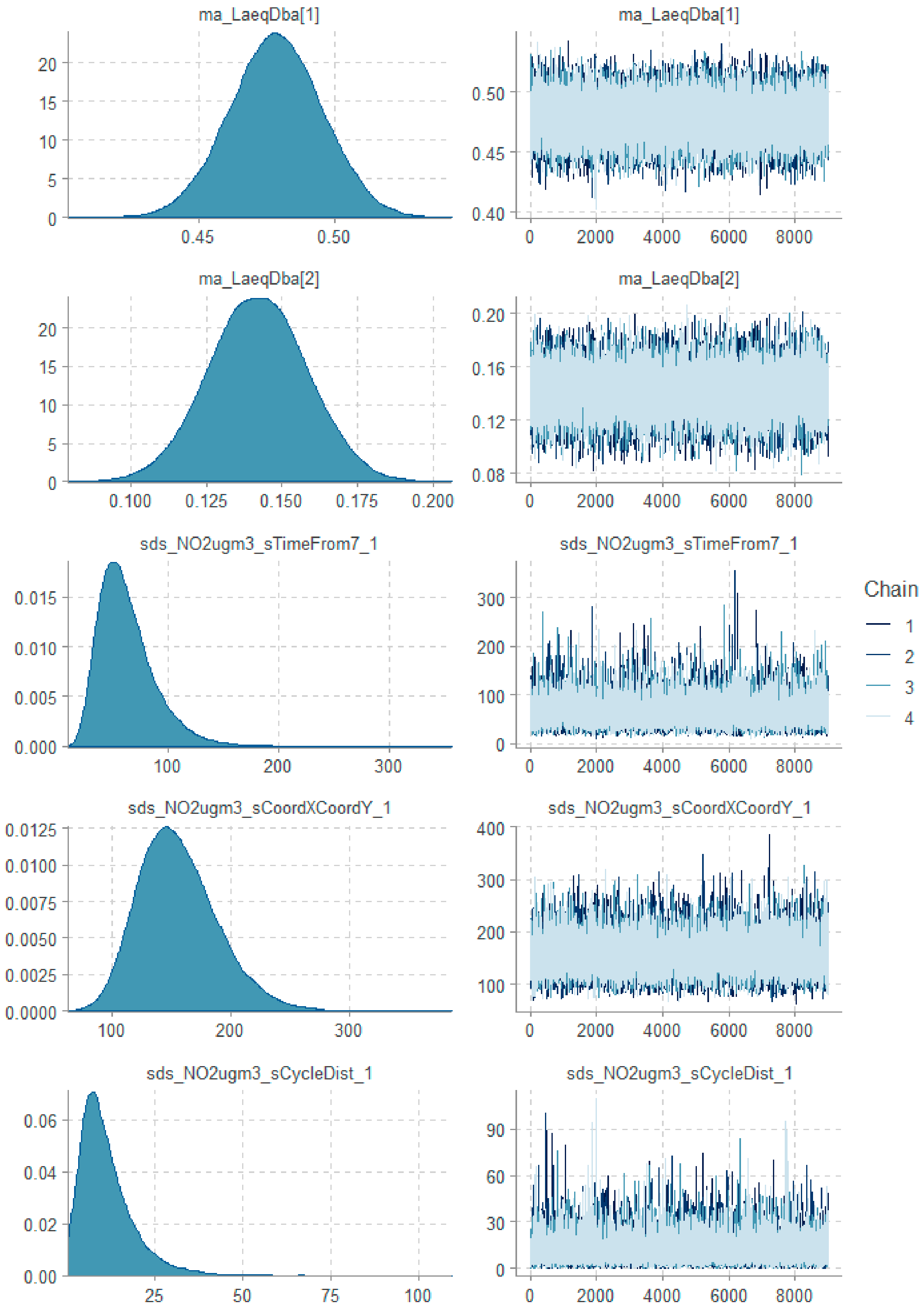

b_LaeqDba_FetcIdx
b_LaeqDba_VSkyIdx
b_LaeqDba_TempCels
b_LaeqDba_Vent:FetcIdx
bs_NO2ugm3_sTimeFrom7_1
Chain
1
2
3
4


bs_NO2ugm3_sCoordXCoordY_1
bs_NO2ugm3_sCoordXCoordY_2
bs_NO2ugm3_sCycleDist_1
bs_LaeqDba_sTimeFrom7_1
bs_LaeqDba_sCoordXCoordY_1
Chain
1
2
3
4

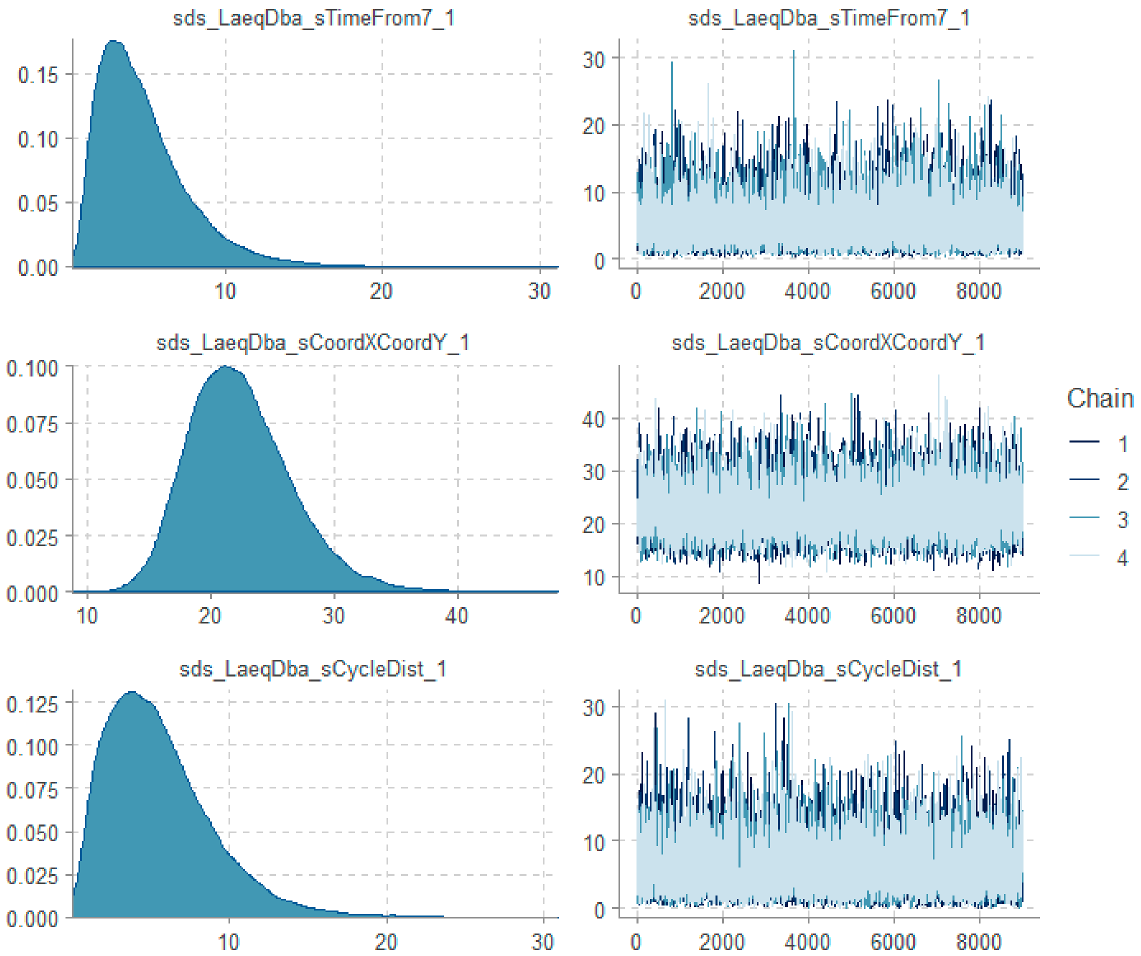


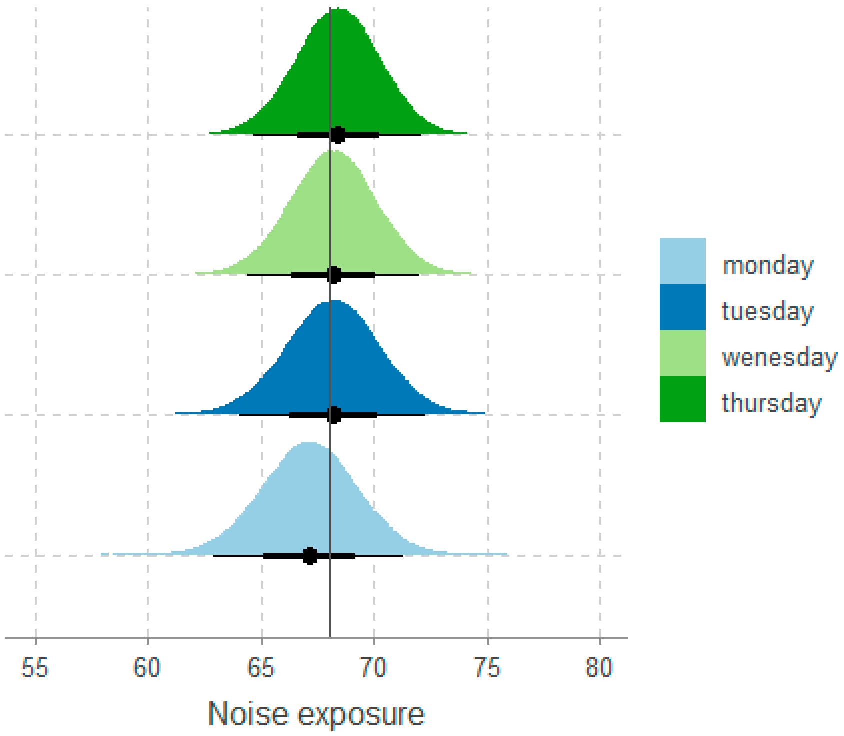
bs_LaeqDba_sCoordXCoordY_2
bs_LaeqDba_sCycleDist_1
sd_DayName__NO2ugm3_Intercept
sd_IDPart__NO2ugm3_Intercept
sd_DayName__LaeqDba_Intercept
Chain
1
2
3
4

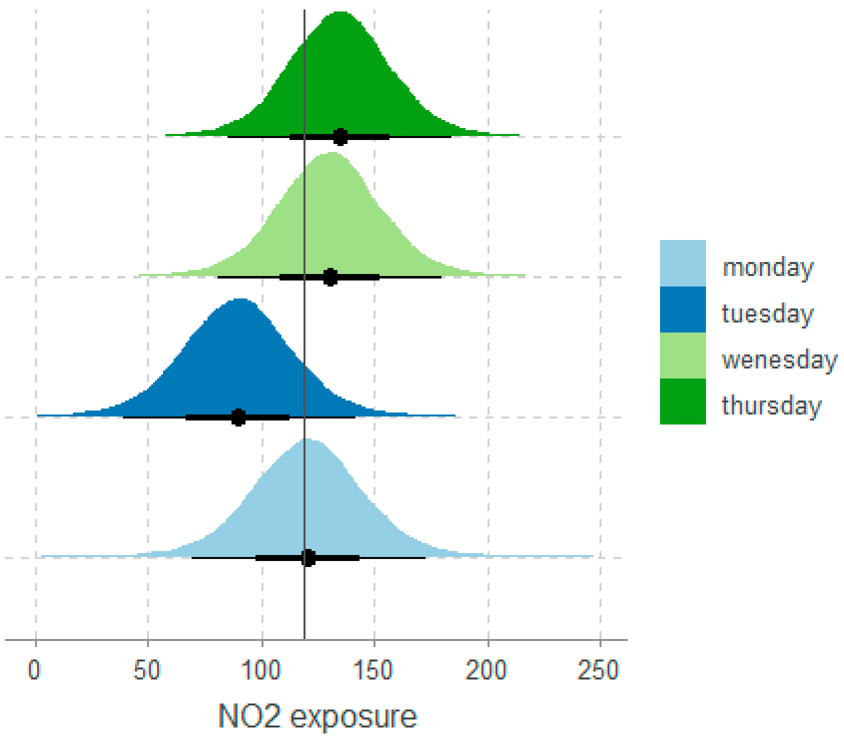

sd_IDPart__LaeqDba_Intercept
sigma_NO2ugm3
sigma_LaeqDba
ma_NO2ugm3[1]
ma_NO2ugm3[2]
Chain
1
2
3
4



ma_LaeqDba[1]
ma_LaeqDba[2]
sds_NO2ugm3_sTimeFrom7_1
sds_NO2ugm3_sCoordXCoordY_1
sds_NO2ugm3_sCycleDist_1
Chain
1
2
3
4


sds_LaeqDba_sTimeFrom7_1
0.15
0.10
0.05
0.00
10
20
30
sds_LaeqDba_sTimeFrom7_1
30
20
10
0
0
2000
4000
6000
8000
sds_LaeqDba_sCoordXCoordY_1
0.100
0.075
0.050
0.025
0.000
10
20
30
40
sds_LaeqDba_sCoordXCoordY_1
40
30
20
10
Chain
1
2
3
4
sds_LaeqDba_sCycleDist_1
0.125
0.100
0.075
0.050
0.025
0.000
sds_LaeqDba_sCycleDist_1


monday
tuesday
wenesday
thursday
55
60
65
70
75
80
Noise exposure


monday
tuesday
wenesday
thursday
0
50
100
150
200
250
NO2 exposure

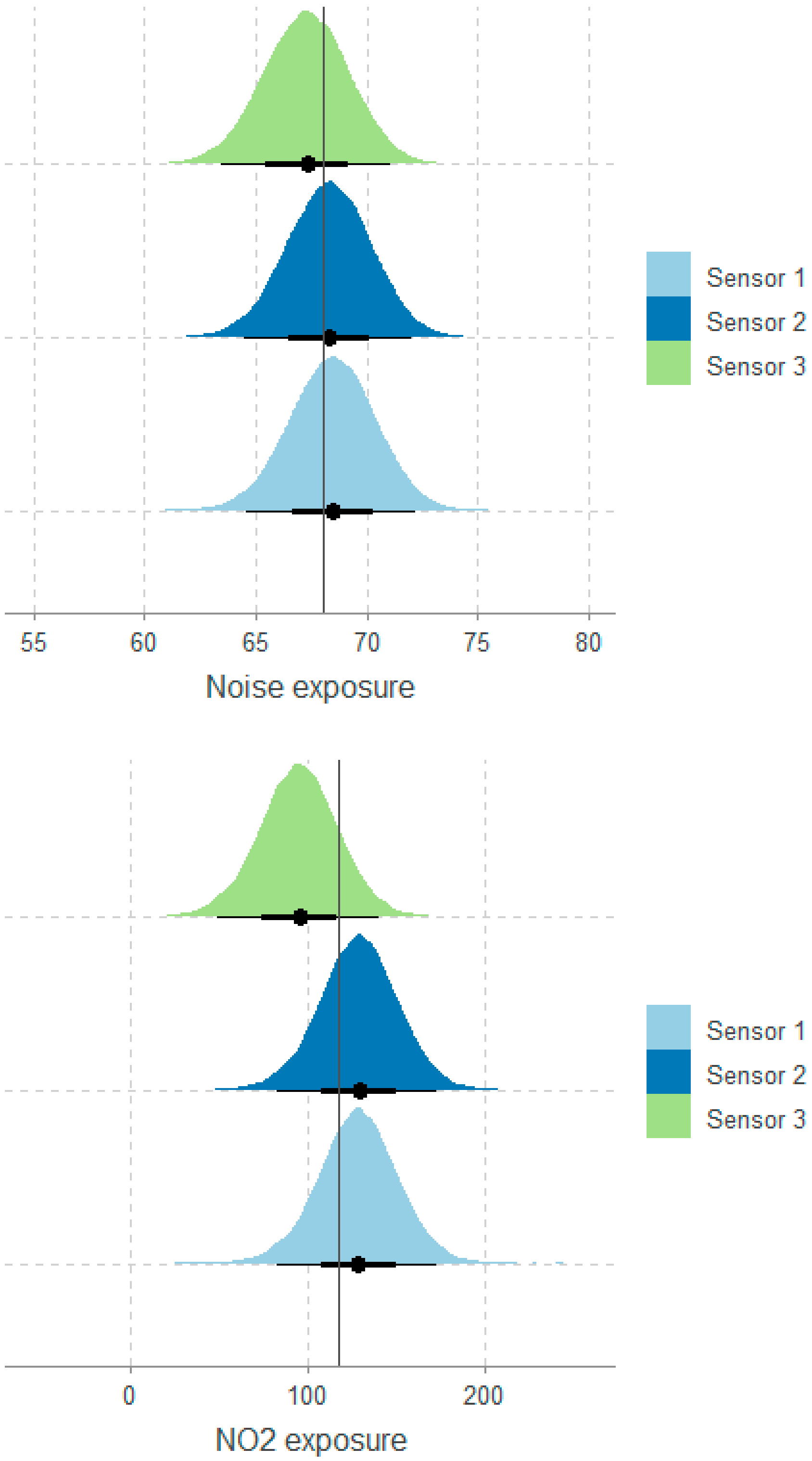


Figure S2. Posterior distributions of all parameters.

4. Graphical Posterior Predictive Checks

The graphical posterior predictive checks for the two independent variables are presented in Figure S3.

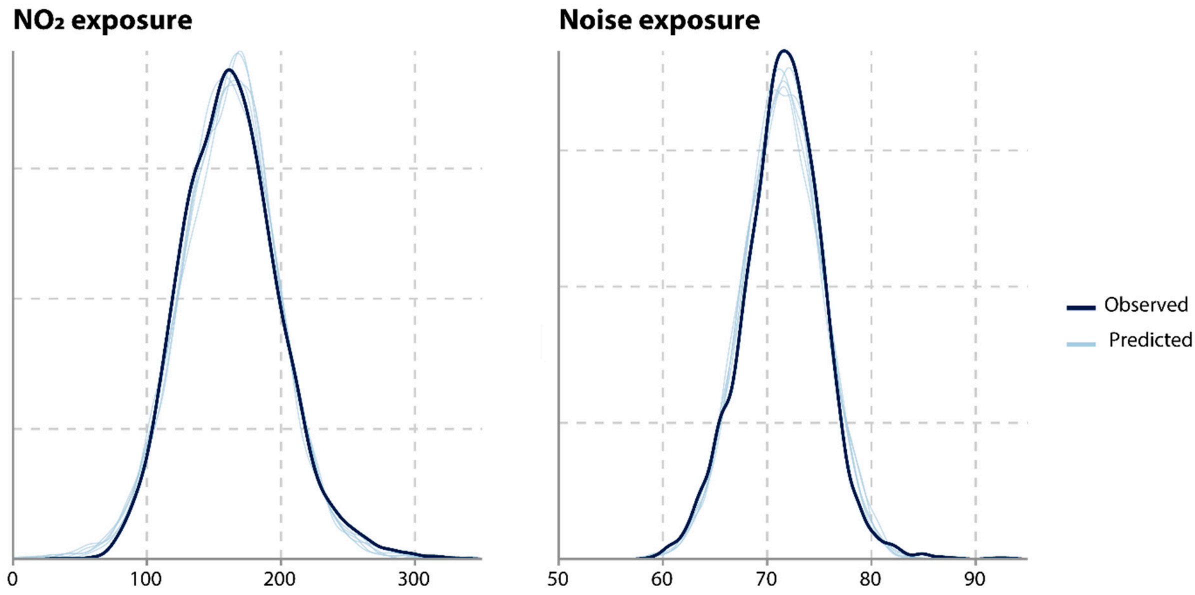


Figure S3. Graphical posterior predictive checks.

Reference

1. Gelman, A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian anal.*,**2006**, *1*(3), 515–534.