Bayesian Analysis of PM2.5 Emissions: A Spatial Hierarchical Model Comparing Developed and Developing Nations

Peter Jin, student #: 45472420
Zhenglin Wu, student #: 26455675

STAT447C: Bayesian Statistics
The University of British Columbia
April 12, 2025

Core inference problem:

Is the emission of PM2.5 explicitly related to whether this country is developed or developing?

1. Introduction and Problem Formulation

Atmospheric contamination of fine particulate matter (PM2.5) is widely accepted as a disease-causing factor in cardiovascular diseases, respiratory infections, and neoplasia. The World Health Organization states that over 90% of the world's population resides where PM2.5 is higher than safe exposure levels (Shaddick et al., 2018). Accurate measurement of concentrations of PM2.5 in nations and description of the trend variations in pollution by socioeconomic level are crucial to providing information on climate policy and public health interventions.

Fine particulate matter is particularly hazardous to health as it can travel deeply within tissue in the lungs and potentially into the blood via particles with diameters smaller than 2.5 micrometers. Particulates of such nature emanate from a wide range of sources, both human-made, like emissions from vehicles and industries, as well as constructions and natural disasters, including desert dust and wildfires. Such complexity of processes involved in the source and transportation of PM2.5 warrants the employment of elaborate statistical tools to provide for space dependency as well as the hierarchical nature of data.

This research investigates whether the trajectory of PM2.5 emissions systematically differs between developed (Annex I) and developing (Non-Annex I) countries under the United Nations Framework Convention on Climate Change (UNFCCC). Specifically, we seek to establish the following: Is PM2.5 emission directly connected to whether or not a country is developed or developing? In order to do so, we develop a hierarchical Bayesian model in the spatial framework that allows for regional and country-level variation, time trend, and intergroup differences in the pattern of pollution.

Our hypothesis would address the primary question of climate justice: Are industrial nations that have been causing greater world pollution following different emission-cutback trends from developing nations? Knowing those trends would be important so as to write equitable climate agreements and intervention programs respecting the differing abilities and obligations of different countries.

Our study meets the project topic of "spatial models" and uses a real global dataset of PM10 emissions between 1970 and 2015 as a proxy for PM2.5. Our project integrates Bayesian inference, hierarchical, and spatial modeling to provide coherent, interpretable results regarding world air pollution inequality.

2. Literature Review

Bayesian hierarchical models have proved useful in representing spatial heterogeneity as well as multi-source environmental data integration (Shaddick et al., 2018). Data Integration Model for Air Quality (DIMAQ) is a global-scale hierarchical integration that pairs ground monitoring data with satellite and chemical transport model outputs. These models account for observation uncertainty and regional heterogeneity. Shaddick et al. (2018) illustrate that this approach provides better estimates of PM2.5 concentrations than do single-source approaches. Their model works especially well in cases where there are sparse-density monitoring networks.

Remotely sensed surrogates for PM2.5, such as aerosol optical depth (AOD), can, however, have unignorable spatially correlated biases with respect to actual surface-level particulate matter. Paciorek and Liu (2009) argue that AOD is an uncertain surrogate in some cases, particularly in land use, meteorology, and emissions inventories. This thus convinces us to use direct emission inventories as our base data source. In their analysis, they found that the AOD-PM2.5 relationship is highly region- and season-dependent and hence challenging to perform conversions of ground-level concentrations to satellite observations directly.

Methodologically, by including packages like spTimer and rstan, we may have a hierarchical spatio-temporal model. Sahu and Bakar (2015) present Bayesian autoregressive models derived from Gaussian processes for forecasting air pollution, observing their capability to model structured and fixed spatial random effects. Their approach relies on three types of spatio-temporal models with different means of handling temporal dependence and spatial correlation.

Finally, policy analysis also emphasizes that the Annex I countries have a deeper moral and historical responsibility to cut emissions. Höhne et al. (2014) compare effort-sharing strategies and observe that richer regions are often expected to achieve more rigorous emission cuts relative to developing ones. This places our model specification in perspective to incorporate interaction terms between year and Annex status. Their equity analysis of climate policy postulates provides the conceptual foundation for differentiating among Annex I and non-Annex I countries in our model, with most being based on the assumption that developing countries to make more drastic emission reductions.

3. Data and Preprocessing

Our data provide annual sums of PM10 emissions for 224 countries between 1970 and 2015, disaggregated by IPCC Annex status and world region. We employ PM10 as a proxy for PM2.5 using standard conversion practice (Brauer et al., 2015). While PM2.5 is technically a subset of PM10 (particles with diameters less than 10 micrometers), evidence suggests that PM2.5 typically accounts for 50-70% of PM10 mass, though the ratio varies by location and source of emissions. For our purposes, we employ a conservative conversion factor consistent with the literature, with the understanding that this introduces some measurement error that will be captured in our model's error term.

The original wide-format dataset was reshaped into a long format using R, with each row representing a country-year pair.

Variables include:

- **pm25**: Estimated PM2.5 level for a country-year
- **year**: Calendar year
- **Annex**: IPCC development classification (Annex I vs. Non-Annex I)

- **country_id**: Unique index for each country
- **region_id**: Unique index for each world region

Exploratory data analysis of similar emissions between and within Annex groups revealed strong heterogeneity in emission trends. Visual inspection also confirmed the spatial clustering of emissions, as similar paths were observed among neighboring countries and supported our decision to include spatial structure within the model.

The complete preprocessing and modeling scripts are available in our GitHub repository: https://github.com/Hawkxiang114/STAT-447-Term-Project.git.

4. Model Construction and Inference

We constructed a hierarchical Bayesian model in a spatial structure, implemented in Stan. We aim to estimate the effect of year, Annex status, and their interaction on PM2.5 emissions while accounting for regional and country-specific random effects.

Model construction:

Prior:

- $\alpha \sim N(0, 5)$, A global intercept (baseline $PM_{2.5}$ when year and annex status are zero).
- $\beta_{year} \sim N(0, 2)$, Effect of time (year) on $PM_{2.5}$ for non-Annex countries.
- $\beta_{annex} \sim N(0, 2)$, Baseline difference between Annex I and non-Annex I countries.
- $\beta_{interaction} \sim N(0, 2)$, Interaction effect (how Annex status modifies the time trend).
- $\gamma_j \sim N(0, \sigma^2 country)$, Country-specific deviation for country j (nested within regions).
- $\delta k \sim N(0, \sigma^2 region)$, Region-specific deviation for region k.
- $\sigma_{country} \sim Exp(1)$, Standard deviation of country-level effects
- $\sigma_{region} \sim Exp(1)$, Standard deviation of region-level effects
- $\sigma \sim Exp(1)$, Prior to observation noise

The parameters γj and δk are hierarchical structured variables, which make them random, unlike the fixed priors. Our prior specifications are a compromise between strong domain experience regarding informative priors and allowing data to speak through somewhat vague distributions. The normal prior over fixed effects is around zero, with scales set to accommodate a plausible effect size for standardized data. For variance parameters, we used exponential priors with rate 1, which gives most prior weight to small values but is weakly informative.

Likelihood:

- $yi \sim N(\mu i, \sigma^2)$
- $\mu i = \alpha + \beta year \cdot yeari + \beta annex \cdot annexi + \beta interaction \cdot (yeari \cdot annexi) + \gamma j + \delta k$

The model was written in Stan using non-centered parameterization on the hierarchical terms to optimize sampling. This reparametrizes the random effects such that funnel-shaped posteriors obstructive to Hamiltonian Monte Carlo sampling are avoided. Instead of directly sampling γj and δk, we sample from a standard normal variate and then multiply it by the relevant standard deviations:

$\gamma_j = \sigma_{country} \cdot \tilde{\gamma}_j$, where $\tilde{\gamma}_j \sim N(0, 1)$

$\delta_k = \sigma_{region} \cdot \tilde{\delta}_k$, where $\tilde{\delta}_k \sim N(0, 1)$

We optimized the model afterward with rstan on 4 chains of 2000 iterations (1000 warm-up). R-hat for all our major parameters was near 1, and effective sample sizes >1000; diagnostics of convergence show good mixing and convergence of chains. Monte Carlo standard error averaged less than 0.01 for most major parameters of interest under study and confirms satisfactory precision within our posterior estimates.

In order to confirm our model specification, we also estimated a reduced model without spatial random effects and compared models with leave-one-out cross-validation (LOO-CV). The hierarchical spatial model worked much better (difference in expected log predictive density > 300 with standard error < 30), and it confirmed that in this case, the spatial structure modeling was significant.

5. Results and Evaluation

Posterior summaries show a statistically significant time trend in PM2.5 emissions. The major findings from our model include:

1. **Global intercept (alpha)**: The posterior mean for the intercept is 2.78 (95% credible interval: [-6.55, 12.29]), indicating substantial uncertainty in the baseline PM2.5 level when other predictors are at zero.
2. **Global time trend (beta_year)**: The posterior mean for the year effect is 3.85 (95% credible interval: [3.46, 4.23]), indicating a significant positive trend in PM2.5 emissions over time for non-Annex countries when controlling for other factors.
3. **Annex status effect (beta_annex)**: The baseline difference between Annex I and non-Annex I countries has a posterior mean of 0.11 (95% credible interval: [-3.79, 4.10]), suggesting no clear evidence of a difference between developed and developing nations at the reference year when accounting for other factors.
4. **Interaction effect (beta_interaction)**: The interaction term was strongly negative with a posterior mean of -6.36 (95% credible interval: [-7.21, 5.51]), implying Annex I countries have reduced emissions at a substantially faster rate over time compared to Non-Annex I nations. This is an ideal answer to our core inference question, proving that emission trajectories differ by development status.

Trace plots for all the parameters showed well-mixed chains with no sign of ill-convergence. The effective sample sizes were substantial for all the parameters (ranging from 2132 to 8229), and this suggests credible posterior estimates. The R-hat values of 1 in all parameters indicate superb convergence of the Markov chains.

Posterior predictive checks also indicated a good fit of the model with close similarity of simulated distributions of data and observed emissions. Sensitivity analysis through adjusting prior scales also produced consistent inference about the critical interaction parameter, attesting that our inferences for differential emission paths between developed and developing nations are not sensitive.

6. Methodological Significance and Creativity

Our study contributes to the science of climate accountability by merging temporal modeling, spatial hierarchical effects, and policy categories into a single probabilistic model. Though inspired by such large-scale models as DIMAQ (Shaddick et al., 2018), our approach remains tractable and interpretable and is suitable for inferential, rather than predictive, ends.

This hierarchical structure enables our model to have several advantages: it enables partial pooling that balances local heterogeneities and the risk of overfitting; it propagates uncertainty from spatial level to spatial level; and it integrates several spatial scales to reflect the multi-scale character of environmental processes.

Our model incorporates Rao-Blackwellization for greater MCMC speed, time trend * Annex status interaction for estimation of our research question directly, and hierarchical nesting that is still sensitive to geographic pattern while allowing regional patterns that force country-level estimation. This method, used as a synthesis of a policy-relevant environmental question, is an effective application of Bayesian spatial model concepts.

7. Discussion and Limitations

Our analysis provides substantial support that emission trends of PM2.5 differ systematically between developed (Annex I) and developing (non-Annex I) countries, where developed nations signify more acute emission reductions over time. This finding aligns with the anticipation that more responsible historic nations with greater technology capability would lead the way in pollution controls (Höhne et al., 2014).

However, several limitations persist:

1. **Measurement uncertainty due to proxy use:** Using PM10 as a proxy for PM2.5 introduces measurement error. Even though we condition on observation noise in our model, PM2.5 measurements would be a stronger inference.
2. **Linear time trends**: Perhaps our linear year trend assumption is too simplistic to capture regimes or policy-switching countries' emission dynamics. More advanced models will incorporate changepoint detection or non-parametric time trends in order to represent more complicated temporal dynamics.
3. **Omitted variables**: Our model does not have covariates like GDP, population density, and industrialization levels that can account for some of the differences observed. The inclusion of these variables may help determine if the Annex effect still holds after accounting for economic development.
4. **Simplified spatial structure**: We use discrete random effects for regions and nations rather than continuous spatial processes. While computationally easier, the simplification may miss finer-scale patterns of spatial correlation that Paciorek and Liu (2009) highlight.
5. **Limited satellite integration**: Unlike the DIMAQ model (Shaddick et al., 2018), we do not merge satellite data with ground-level observations, which can increase coverage in regions with sparse monitoring networks.

Future studies can keep adding the model to capture non-linear patterns with time or include covariates such as GDP and population density. In addition, fusion of data from multiple sources, as in Shaddick et al. (2018), can enhance validity and coverage in global

PM2.5 estimation. Finally, examining other spatial covariance structures with the methodology proposed by Sahu and Bakar (2015) will also better fit and provide additional information concerning dispersion patterns of contaminants.

8. Contributions

This was a team project completed by Peter Jin and Zhenglin Wu. Zhenglin completed all the report writing and edited small portions of the R code. Peter completed the majority of the R and Stan coding, model specification, posterior sampling, and figure generation. Both members submitted their codes and documents to the shared GitHub repository.

Appendix

R Script:

```r
library(rstan)
library(dplyr)
library(tidyr)

df <- read.csv("v50_PM10_1970_2015 (1)(TOTALS BY COUNTRY).csv") %>%
  # Remove all auto-generated empty columns(weird problem)
  select(-matches("^X(\\.?\\d+)?$")) %>%
  # Convert and pivot
  mutate(across(starts_with("Y_"), as.numeric)) %>%
  #reshaping the data into a new,long format for easier future processing
  pivot_longer(
    cols = starts_with("Y_"), # Columns convert to pivots
    names_to = "year", # Create a new column containing the years
    values_to = "pm25", # Create a new column containing the pm2.5 values
    values_drop_na = TRUE # Remove NA pm2.5 rows
  ) %>%
  # Convert "Y_1990" → 1990
  mutate(year = as.numeric(gsub("Y_", "", year)))

stan_data <- list(
  N = nrow(df), # Total observations
  pm25 = df$pm25,
  year = df$year - min(df$year), # Centered year (1970-2015 becomes 0-45)
  annex = ifelse(df$IPCC.Annex == "Annex_I", 1, 0), # Binary annex status
  country_id = as.numeric(factor(df$ISO_A3)), # Unique country IDs
  region_id = as.numeric(factor(df$World.Region)), # Unique region IDs
  J = length(unique(df$ISO_A3)), # Number of countries
  K = length(unique(df$World.Region)) # Number of regions
)


spatial_model <- stan(
  file = "spatial_model_rao.stan",
  data = stan_data,
  chains = 4,
  iter = 2000,
  seed = 1
)

print(spatial_model, pars = c("alpha", "beta_year", "beta_annex", "beta_interaction"))
```

Output:

```
SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 1).
Chain 1:
Chain 1: Gradient evaluation took 0.003628 seconds
Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 36.28 seconds.
Chain 1: Adjust your expectations accordingly!
Chain 1:
Chain 1:
Chain 1: Iteration:    1 / 2000 [  0%]  (Warmup)
Chain 1: Iteration:  200 / 2000 [ 10%]  (Warmup)
Chain 1: Iteration:  400 / 2000 [ 20%]  (Warmup)
Chain 1: Iteration:  600 / 2000 [ 30%]  (Warmup)
Chain 1: Iteration:  800 / 2000 [ 40%]  (Warmup)
Chain 1: Iteration: 1000 / 2000 [ 50%]  (Warmup)
Chain 1: Iteration: 1001 / 2000 [ 50%]  (Sampling)
Chain 1: Iteration: 1200 / 2000 [ 60%]  (Sampling)
Chain 1: Iteration: 1400 / 2000 [ 70%]  (Sampling)
Chain 1: Iteration: 1600 / 2000 [ 80%]  (Sampling)
Chain 1: Iteration: 1800 / 2000 [ 90%]  (Sampling)
Chain 1: Iteration: 2000 / 2000 [100%]  (Sampling)
Chain 1:
Chain 1:  Elapsed Time: 170.457 seconds (Warm-up)
Chain 1:                154.745 seconds (Sampling)
Chain 1:                325.202 seconds (Total)
Chain 1:
```

```
SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 2).
Chain 2: 
Chain 2: Gradient evaluation took 0.002398 seconds
Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 23.98 seconds.
Chain 2: Adjust your expectations accordingly!
Chain 2: 
Chain 2: 
Chain 2: Iteration:    1 / 2000 [  0%]  (Warmup)
Chain 2: Iteration:  200 / 2000 [ 10%]  (Warmup)
Chain 2: Iteration:  400 / 2000 [ 20%]  (Warmup)
Chain 2: Iteration:  600 / 2000 [ 30%]  (Warmup)
Chain 2: Iteration:  800 / 2000 [ 40%]  (Warmup)
Chain 2: Iteration: 1000 / 2000 [ 50%]  (Warmup)
Chain 2: Iteration: 1001 / 2000 [ 50%]  (Sampling)
Chain 2: Iteration: 1200 / 2000 [ 60%]  (Sampling)
Chain 2: Iteration: 1400 / 2000 [ 70%]  (Sampling)
Chain 2: Iteration: 1600 / 2000 [ 80%]  (Sampling)
Chain 2: Iteration: 1800 / 2000 [ 90%]  (Sampling)
Chain 2: Iteration: 2000 / 2000 [100%]  (Sampling)
Chain 2: 
Chain 2:  Elapsed Time: 174.619 seconds (Warm-up)
Chain 2:                154.768 seconds (Sampling)
Chain 2:                329.387 seconds (Total)
Chain 2: 
```

```
SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 3).
Chain 3: 
Chain 3: Gradient evaluation took 0.002626 seconds
Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 26.26 seconds.
Chain 3: Adjust your expectations accordingly!
Chain 3: 
Chain 3: 
Chain 3: Iteration:    1 / 2000 [  0%]  (Warmup)
Chain 3: Iteration:  200 / 2000 [ 10%]  (Warmup)
Chain 3: Iteration:  400 / 2000 [ 20%]  (Warmup)
Chain 3: Iteration:  600 / 2000 [ 30%]  (Warmup)
Chain 3: Iteration:  800 / 2000 [ 40%]  (Warmup)
Chain 3: Iteration: 1000 / 2000 [ 50%]  (Warmup)
Chain 3: Iteration: 1001 / 2000 [ 50%]  (Sampling)
Chain 3: Iteration: 1200 / 2000 [ 60%]  (Sampling)
Chain 3: Iteration: 1400 / 2000 [ 70%]  (Sampling)
Chain 3: Iteration: 1600 / 2000 [ 80%]  (Sampling)
Chain 3: Iteration: 1800 / 2000 [ 90%]  (Sampling)
Chain 3: Iteration: 2000 / 2000 [100%]  (Sampling)
Chain 3: 
Chain 3:  Elapsed Time: 173.46 seconds (Warm-up)
Chain 3:                151.293 seconds (Sampling)
Chain 3:                324.753 seconds (Total)
Chain 3: 
```

```
SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 4).
Chain 4: 
Chain 4: Gradient evaluation took 0.002594 seconds
Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 25.94 seconds.
Chain 4: Adjust your expectations accordingly!
Chain 4: 
Chain 4: 
Chain 4: Iteration:    1 / 2000 [  0%]  (Warmup)
Chain 4: Iteration:  200 / 2000 [ 10%]  (Warmup)
Chain 4: Iteration:  400 / 2000 [ 20%]  (Warmup)
Chain 4: Iteration:  600 / 2000 [ 30%]  (Warmup)
Chain 4: Iteration:  800 / 2000 [ 40%]  (Warmup)
Chain 4: Iteration: 1000 / 2000 [ 50%]  (Warmup)
Chain 4: Iteration: 1001 / 2000 [ 50%]  (Sampling)
Chain 4: Iteration: 1200 / 2000 [ 60%]  (Sampling)
Chain 4: Iteration: 1400 / 2000 [ 70%]  (Sampling)
Chain 4: Iteration: 1600 / 2000 [ 80%]  (Sampling)
Chain 4: Iteration: 1800 / 2000 [ 90%]  (Sampling)
Chain 4: Iteration: 2000 / 2000 [100%]  (Sampling)
Chain 4: 
Chain 4:  Elapsed Time: 184.283 seconds (Warm-up)
Chain 4:                151.143 seconds (Sampling)
Chain 4:                335.426 seconds (Total)
Chain 4: 
```

```
> print(spatial_model, pars = c("alpha", "beta_year", "beta_annex", "beta_interaction"))
Inference for Stan model: anon_model.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

                  mean se_mean   sd  2.5%   25%   50%   75% 97.5% n_eff Rhat
alpha             2.78    0.11 4.89 -6.55 -0.51  2.63  6.07 12.29  2132    1
beta_year         3.85    0.00 0.20  3.46  3.72  3.86  3.99  4.23  2583    1
beta_annex        0.11    0.02 2.00 -3.79 -1.18  0.09  1.42  4.10  8229    1
beta_interaction -6.36    0.01 0.44 -7.21 -6.67 -6.36 -6.07 -5.51  2610    1

Samples were drawn using NUTS(diag_e) at Fri Apr 11 22:00:18 2025.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
> 
```

<u>Spatial Model Stan:</u>

447_final_project.R* | spatial_model_rao.stan | df

Check on Save

```
data {
  int<lower=0> N; # Total observations
  vector[N] pm25; # pm2.5 measurements
  vector[N] year; # Centered years (year - 1970)
  int annex[N]; # Annex status (1 = Annex, 0 = Non-Annex)
  int<lower=0> J; # Number of unique countries
  int<lower=1, upper=J> country_id[N]; # Country IDs (1..J)
  int<lower=0> K; # Number of regions
  int<lower=1, upper=K> region_id[N]; # Region IDs (1..K)
}

parameters {
  # Fixed effect
  real alpha; # Intercept (baseline pm2.5)
  real beta_year; # Effect of time (year)
  real beta_annex; # Effect of Annex I status
  real beta_interaction; # Interaction: Year × Annex I

  # Random effects
  vector[J] z_country; # Standardized country random effects
  vector[K] z_region; # Standardized region random effects
  real<lower=0> sigma_country; # SD of country effects
  real<lower=0> sigma_region; # SD of region effects

  real<lower=0> sigma; # Residual error SD
}

```

```
transformed parameters {
  # Non-centered parameterization
  vector[J] country_effect = z_country * sigma_country;
  vector[K] region_effect = z_region * sigma_region;

  # Linear predictor
  vector[N] mu;
  for (n in 1:N) {
    mu[n] = alpha +
            beta_year * year[n] +
            beta_annex * annex[n] +
            beta_interaction * year[n] * annex[n] +
            country_effect[country_id[n]] +
            region_effect[region_id[n]];
  }
}

```

```
model {
  # Priors
  alpha ~ normal(0, 5);
  beta_year ~ normal(0, 2);
  beta_annex ~ normal(0, 2);
  beta_interaction ~ normal(0, 2);
  sigma ~ exponential(1);

  # Random effects
  z_country ~ std_normal();
  z_region ~ std_normal();
  sigma_country ~ exponential(1);
  sigma_region ~ exponential(1);

  # Likelihood
  pm25 ~ normal(mu, sigma);
}

generated quantities {
  # Posterior predictive checks
  vector[N] y_rep;
  for (n in 1:N) {
    y_rep[n] = normal_rng(mu[n], sigma);
  }
}

```

Reference

Bakar, K. S., & Sahu, S. K. (2015). spTimer: Spatio-Temporal Bayesian Modeling Using*R*. *Journal of Statistical Software*, *63*(15). https://doi.org/10.18637/jss.v063.i15

Niklas, H. Michel, d. E & Donovan, E. (2013). Regional GHG reduction targets based on effort sharing: a comparison of studies. (2015). *Climate Policy*. https://doi.org/10.1080//14693062.2014.849452

Paciorek, C. J., & Liu, Y. (2009). Limitations of Remotely Sensed Aerosol as a Spatial Proxy for Fine Particulate Matter. *Environmental Health Perspectives*, *117*(6), 904–909. https://doi.org/10.1289/ehp.0800360

Shaddick, G., Thomas, M. L., Jobling, A., Brauer, M., Donkelaar, van, Burnett, R., Chang, H., Cohen, A., Dingenen, V., Dora, C., Gumy, S., Liu, Y., Martin, R., Waller, L. A., West, J., Zidek, J. V., & Prüss-Ustün, A. (2016). Data Integration Model for Air Quality: A Hierarchical Approach to the Global Estimation of Exposures to Ambient Air Pollution. ArXiv.org. https://arxiv.org/abs/1609.00141