

# Temporal and spatial variations of PM<sub>2.5</sub> and PM<sub>10</sub> concentrations in Mongolia

Erdenebayar Munkhtsetseg<sup>1,2</sup>, Atsushi Shimizu<sup>3</sup>

<sup>1</sup>National University of Mongolia (NUM), Mongolia,

<sup>2</sup>Kanazawa University, Japan,

<sup>3</sup>National Institute for Environmental Studies (NIES), Japan,

---

Corresponding author: Atsushi Shimizu, [shimizua@nies.go.jp](mailto:shimizua@nies.go.jp)

**Abstract**

PM2.5 and PM10 data for the 4 distinct sites of Mongolia from 2008 to 2020 is found ....

**Plain Language Summary**

PM2.5 and PM10 data for the 4 distinct sites of Mongolia from 2008 to 2020 is found ...

**0.1 Data & Methods****1 01\_datawork**

Munkhtsetseg

Library

**2 Import the dataset and remove the duplicates**

Import the dataset from the directory of: ~/Data Input/Preprocessing data/Preprocessing data.csv, assign the dataset as object of df:

Remove the duplicates with the function of distinct(), assign the dataset as df\_01:

**2.1 Produce a table with missing data**

For date options as year, month, etc:

```
# A tibble: 35 × 9
```

```
# Groups:   Station.name [4]
```

	Station.name	Year	NA_date	NA_PM2	NA_PM10	NA_Vis	NA_WD	NA_WS	NA_OPC
	<chr>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1	Dalanzadgad	2009	8760	715	929	659	748	748	8760
2	Dalanzadgad	2010	8784	921	1086	756	787	787	8784
3	Dalanzadgad	2011	8760	2652	3309	1759	2394	2394	8760
4	Dalanzadgad	2012	5088	1074	3016	693	1412	1412	5088
5	Dalanzadgad	2013	6096	1766	1809	2479	1240	1240	6096
6	Dalanzadgad	2014	7800	843	921	6068	1482	1482	7800
7	Dalanzadgad	2015	8760	1539	1587	8115	2635	2635	8760
8	Dalanzadgad	2016	6288	1654	1613	5995	3306	3306	6288
9	Sainshand	2009	8688	376	424	423	587	587	8688
10	Sainshand	2010	8784	2557	2577	1113	1210	1210	8784

```
# 25 more rows
```

For station

```
# A tibble: 4 × 8
```

	Station.name	NA_date	NA_PM2	NA_PM10	NA_Vis	NA_WD	NA_WS	NA_OPC
	<chr>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1	Dalanzadgad	60336	11164	14270	26524	14004	14004	60336
2	Sainshand	59040	11727	11929	9320	8527	8527	59040
3	UB	76656	7879	8716	3770	4053	4053	43415
4	Zamynuud	67392	8880	10075	3444	4960	4960	67392

By percentages

```
# A tibble: 4 × 6
```

```
# Groups:   Station.name [4]
```

	Station.name	missing_PM2	missing_PM10	missing_Vis	missing_WS	missing_WD
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Dalanzadgad	25.7	19.2	44.5	24.3	24.3
2	Sainshand	20.0	19.7	15.7	14.6	14.6
3	UB	11.9	11.0	4.53	4.85	4.85
4	Zamynuud	14.4	12.7	5.49	7.44	7.44

### 3 Note that:

We use the data in the period of 2009-2018, which has been regarded as a monitoring work stabilized since 2008 when is the beginning of the monitoring. According to NIES, site maintenance was consistent up to 2018. - Sainshand site, data 2009-2015 get used; - Dalanzad site: 2009-2016. - UB: 2009-2018 - Zamyn uud: 2009-2018

### 4 Remove the spikes, and produce an extended table

Remove the spikes in the datasets, and produce the table with NA, with removed spikes; express it in a percentages. ### Remove the spikes Method 1. Mean value  $\pm$  (3-5)SD - Find Monthly mean

# A tibble: 1,798 × 12

	Year	Month	Day	Hour	PM2	PM10	Visibility	WD	WS	OPC	Station.name
	<int>	<int>	<int>	<int>	<chr>	<chr>	<int>	<int>	<dbl>	<int>	<chr>
1	2009	1	2	17	Outl...	0.29	3622	141	0.524	NA	UB
2	2009	1	3	12	Outl...	0.446	2399	109	0.117	NA	UB
3	2009	1	3	13	Outl...	0.288	1347	17	0.492	NA	UB
4	2009	1	3	14	Outl...	0.504	1241	12	0.829	NA	UB
5	2009	1	3	15	Outl...	0.478	1341	11	0.39	NA	UB
6	2009	1	3	16	Outl...	0.449	2945	136	0.123	NA	UB
7	2009	1	3	18	Outl...	0.341	1436	13	0.742	NA	UB
8	2009	1	3	19	Outl...	0.397	1847	13	0.453	NA	UB
9	2009	1	3	20	Outl...	0.297	3359	22	0.462	NA	UB
10	2009	1	4	2	Outl...	0.311	3167	96	0.759	NA	UB

# 1,788 more rows  
# 1 more variable: Date <chr>

# A tibble: 4,014 × 12

	Year	Month	Day	Hour	PM2	PM10	Visibility	WD	WS	OPC	Station.name
	<int>	<int>	<int>	<int>	<chr>	<chr>	<int>	<int>	<dbl>	<int>	<chr>
1	2009	1	3	15	Outl...	0.292	3444	119	0.856	NA	Dalanzadgad
2	2009	1	5	13	Outl...	0.419	1383	260	1.7	NA	Dalanzadgad
3	2009	1	5	14	Outl...	0.415	1072	266	1.84	NA	Dalanzadgad
4	2009	1	5	15	Outl...	0.466	1099	261	0.83	NA	Dalanzadgad
5	2009	1	5	16	Outl...	0.509	1814	260	0.788	NA	Dalanzadgad
6	2009	1	6	0	Outl...	0.547	744	248	1.23	NA	Dalanzadgad
7	2009	1	6	1	Outl...	0.728	1093	277	0.738	NA	Dalanzadgad
8	2009	1	6	2	Outl...	0.597	1723	0	1.62	NA	Dalanzadgad
9	2009	1	6	3	Outl...	0.33	8186	95	1.1	NA	Dalanzadgad
10	2009	1	6	11	Outl...	0.39	1150	258	1.48	NA	Dalanzadgad

# 4,004 more rows  
# 1 more variable: Date <chr>

#### 4.1 Save dataset in folder: 01\_data\_raw

### 5 Tidy data

#### 5.1 Fill the missing data

Method 1. Fill the gap Method 2. Relationship equation Method 3. Look-up table

#### 5.2 Save dataset in folder: 02\_data\_tidy

Source: [01\\_datawork](#)

#### 5.3 Introduction

Source: [Article Notebook](#)

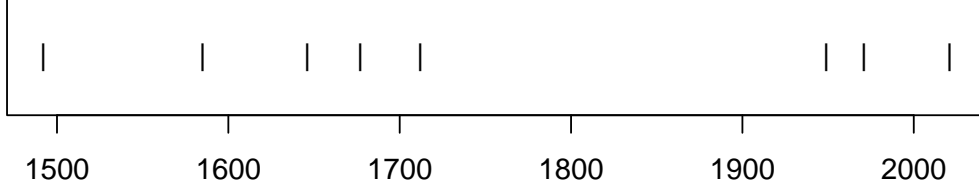


Figure 1: Timeline of recent earthquakes on La Palma

Source: [Article Notebook](#)

Source: [Article Notebook](#)

Based on data up to and including 1971, eruptions on La Palma happen every 79.8 years on average.

Studies of the magma systems feeding the volcano, such as Marrero et al. (2019), have proposed that there are two main magma reservoirs feeding the Cumbre Vieja volcano; one in the mantle (30-40km depth) which charges and in turn feeds a shallower crustal reservoir (10-20km depth).

Eight eruptions have been recorded since the late 1400s (Figure 1).

Data and methods are discussed in Section 0.1.

Let  $x$  denote the number of eruptions in a year. Then,  $x$  can be modeled by a Poisson distribution

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (1)$$

where  $\lambda$  is the rate of eruptions per year. Using Equation 1, the probability of an eruption in the next  $t$  years can be calculated.

Table 1: Recent historic eruptions on La Palma

Name	Year
Current	2021
Teneguía	1971
Nambroque	1949
El Charco	1712
Volcán San Antonio	1677
Volcán San Martin	1646
Tajuya near El Paso	1585
Montaña Quemada	1492

Table 1 summarises the eruptions recorded since the colonization of the islands by Europeans in the late 1400s.

La Palma is one of the west most islands in the Volcanic Archipelago of the Canary Islands (Figure 2).

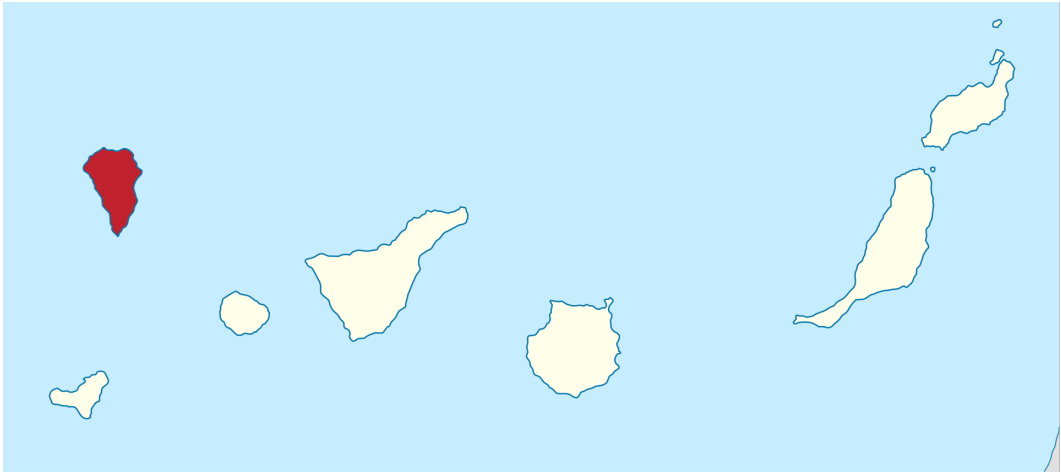


Figure 2: Map of La Palma

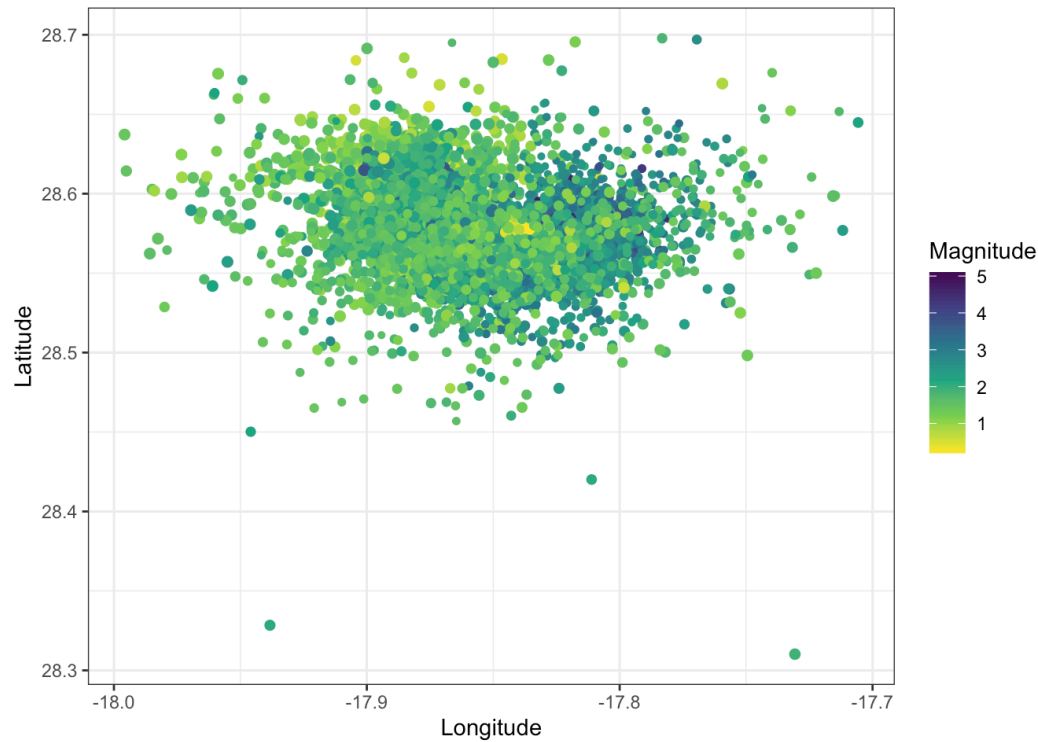


Figure 3: Locations of earthquakes on La Palma since 2017

120 Source: [Explore Earthquakes](#)

121 kk

122 **6 Explore Earthquakes**  
123 Munkhtsetseg

## Library

**7 Import the dataset and remove the duplicates**

Import the dataset from the directory of: ~/Data Input/Preprocessing data/Preprocessing data.csv, assign the dataset as object of df:

Remove the duplicates with the function of distinct(), assign the dataset as df\_01:

**7.1 Produce a table with missing data**

For date options as year, month, etc:

```
# A tibble: 52 × 9
# Groups:   Station.name [4]
  Station.name Year NA_date NA_PM2 NA_PM10 NA_Vis NA_WD NA_WS NA_OPC
  <chr>         <int>   <int>   <int>   <int>   <int>   <int>   <int>   <int>
1 Dalanzadgad  2008    4630    1543    1672    1463    1566    1566    4630
2 Dalanzadgad  2009    8760     715     929     659     748     748    8760
3 Dalanzadgad  2010    8784     921    1086     756     787     787    8784
4 Dalanzadgad  2011    8760    2652    3309    1759    2394    2394    8760
5 Dalanzadgad  2012    5088    1074    3016     693    1412    1412    5088
6 Dalanzadgad  2013    6096    1766    1809    2479    1240    1240    6096
7 Dalanzadgad  2014    7800     843     921    6068    1482    1482    7800
8 Dalanzadgad  2015    8760    1539    1587    8115    2635    2635    8760
9 Dalanzadgad  2016    6288    1654    1613    5995    3306    3306    6288
10 Dalanzadgad 2017    3264     36      45    3264    3264    3264    3264
# 42 more rows
```

For station

```
# A tibble: 4 × 8
  Station.name NA_date NA_PM2 NA_PM10 NA_Vis NA_WD NA_WS NA_OPC
  <chr>         <int>   <int>   <int>   <int>   <int>   <int>   <int>
1 Dalanzadgad  69454  13081  16327  32475  20058  20058  69454
2 Sainshand    101230 27588  36117  28986  13768  13768  101230
3 UB           95662  7895   8785   3775   4121   4121   62421
4 Zamynuud     99742  32281  33597  22525  5373   5373   99742
```

By percentages

```
# A tibble: 4 × 2
# Groups:   Station.name [4]
  Station.name sdq
  <chr>         <dbl>
1 Dalanzadgad  10.7
2 Sainshand    25.9
3 UB           17.9
4 Zamynuud     39.6
```

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

**8 Remove the spikes, and produce an extended table**

Remove the spikes in the datasets, and produce the table with NA, with removed spikes; express it in a percentages.

**8.0.1 Remove the spikes Method 1. Mean value  $\pm (3-5)SD$** 

Method 2. Seasonal variations, and trend-mean

170 **8.1 Save dataset in folder: 01\_data\_raw**  
171 **9 Tidy data**  
172 **9.1 Fill the missing data**  
173 Method 1. Fill the gap Method 2. Relationship equation Method 3. Look-up table  
174 **9.2 Save dataset in folder: 02\_data\_tidy**  
175 Read a clean version of data:  
176 Create spatial plot:

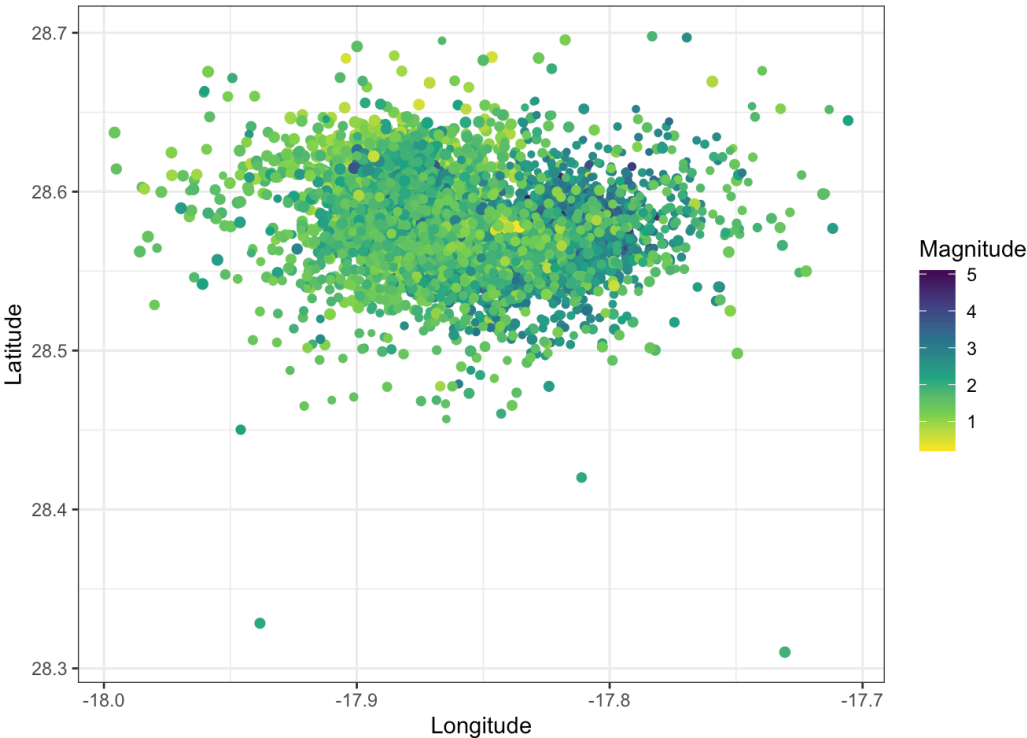


Figure 4: Locations of earthquakes on La Palma since 2017

177 Source: [Explore Earthquakes](#)  
178 Figure 4 shows the location of recent Earthquakes on La Palma.  
179 **9.3 Results**  
180 **9.4 Discussion**  
181 **9.5 Conclusions**  
182 **References**  
183 Marrero, J., García, A., Berrocoso, M., Llinares, Á., Rodríguez-Losada, A., & Ortiz,  
184 R. (2019). Strategies for the development of volcanic hazard maps in mono-  
185 genetic volcanic fields: The example of La Palma (Canary Islands). *Journal of*  
186 *Applied Volcanology*, 8. <https://doi.org/10.1186/s13617-019-0085-5>