STREET-LEVEL AIR POLLUTION MODELLING WITH GRAPH GAUSSIAN PROCESSES

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurately predicting air quality levels at a fine resolution is a critical task to ensuring the public's health is not at risk. In this work, we construct a graph representation of the road network in Mitcham, London, and model nitrogen dioxide levels using a Gaussian process defined on the vertices of our graph. We introduce a heteroscedastic noise process into our model to capture the complex variations that exist in nitrogen dioxide observations. Defining our model in this way offers superior predictive performance to its spatial analogue. Further, a graph representation allows us to infer the air pollution exposure that an individual would experience on a specific journey and their subsequent risk. We demonstrate this approach for the district of Mitcham, London.

1 INTRODUCTION

Exposure to poor ambient air pollution is a major public health issue and producing accurate prediction maps of pollution levels is a critical task in both the earth science and machine learning communities. In this work, we focus our attention on nitrogen dioxide (NO_2) - a pollutant that is primarily produced by the burning of fuels, such as combustion engines in cars. Disentangling the direct effect of NO_2 on one's health is challenging, however, it is estimated that 3.6 million deaths per year result from exposure to pollution created by combustion engines (Lelieveld et al., 2019). There is no single illness that causes these deaths, but short-term exposure to NO_2 has been linked to reduced lung function and increased cardiovascular risk (Strak et al., 2012; 2013).

Constructing regression models capable of interpolating air quality measurements in the absence of monitoring stations is challenging due to data scarcity and the complex underlying spatial topology that exists in urban areas. In the United Kingdom, the largest network of NO_2 sensors contains 145 stations with a disproportionately large number of these residing in major cities. With so few sensors, it is infeasible to build fine-scale models that are capable of inferring NO_2 levels at an individual street level. Furthermore, assuming the model's domain is Euclidean is a common assumption, yet unrealistic in cities and urban areas. Factors such as varying traffic speeds and congestion levels are more accurately explained using a road network (i.e., the graph approach) rather than assuming the underlying space is an unconstrained continuous space (Euclidean approach). A pertinent example of this can be visualised in Figure 1 where we see the problem encountered when computing Euclidean distances on network spaces.

We wish to design a statistical model for air quality that is well-calibrated, accounts for the underlying road network and allows us to make predictions at locations for which we do not have observations. Firstly, a well-calibrated model is essential for informing policy decisions since this highlights points at which the model can be trusted and those where further investigation is required. Secondly, our model should reflect the topology of the road network so that outputs from the model appropriately reflect constraints implied by the real world. Finally, the cost and time-intensive nature of the data collection process motivates the development of models which offer accurate predictions away from the observations. This allows practitioners to make statements based on limited data and outputs from the model can be viewed as a cheap alternative to data collection.

In this work, we construct a Gaussian process (GP) regression model that is defined on a graph representation of London's road network where a junction is represented by a vertex and a road by an edge. GPs are a well-motivated choice of model as they offer calibrated uncertainty estimates

and their flexible non-parametric form allows us to model the complex dynamics exhibited by NO_2. We parameterise our GP using a graph kernel construction, which will allow us to naturally capture London's road network structure. Such a representation allows us to use our model to assess the levels of pollution exposure experienced when taking common journeys, such as a commute or bike ride (Figure 4). Identifying roads with dangerous levels of pollution (Figure 5) enables us to identify and proactively assist individuals who are more at risk of contracting pollution-related illnesses.

2 PRELIMINARIES

Graphs A graph $\mathcal{G} = \{V, E\}$ is a set of vertices $V = \{v_1, v_2, \dots v_{n_v}\}$ and edges $E = \{(v_i, v_j) \,|\, v_i, v_j \in V \text{ and } v_i \neq v_j\}$ for $|E| = n_e$. For a weight function $w : V \times V \to \mathbb{R}$, the graph *adjacency matrix* $\mathbf{A}$ is a matrix such that $[\mathbf{A}]_{i,j}$ is equal to $w(v_i, v_j)$ if and only if $(v_i, v_j) \in E$, and 0 otherwise. To denote adjacency between two vertices v_i, v_j, we can also write $v_i \sim v_j$. The *degree matrix* $\mathbf{D}$ of a graph is a diagonal matrix such that $[\mathbf{D}]_{i,i} = \sum_{j \neq i} A_{i,j}$.

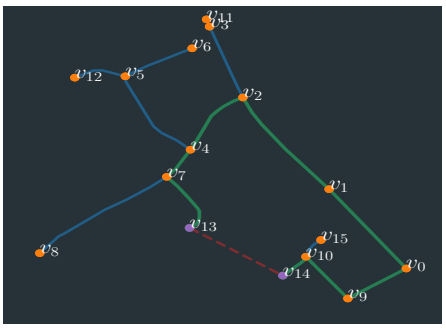


Figure 1: A comparison of graph (green) and Euclidean (red) distances when the underlying space is a network.

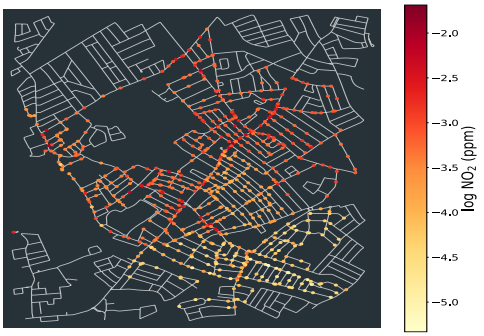


Figure 2: NO_2 observations overlayed onto Mitcham's road network. Observations have been log-scaled to aid visualisation.

The *graph Laplacian* $\mathbf{\Delta} = \mathbf{D} - \mathbf{A}$ is a fundamental concept in spectral graph theory. We can gain some intuition on the Laplacian matrix through

$$\langle \mathbf{g}, \mathbf{\Delta g} \rangle = \mathbf{g}^\top \mathbf{\Delta g} = -0.5 \sum_{v_i \sim v_j} (g_i - g_j)^2 \text{ for all } \mathbf{g} \in \mathbb{R}^{n_v} , \tag{1}$$

where we can see that the graph Laplacian quantifies how smooth a given function $\mathbf{g}$ is over the graph's vertices (Smola & Kondor, 2003). Very smooth functions will result in small $(g_i - g_j)^2$ values, whereas rough functions will change rapidly across vertices, meaning the quadratic difference will be larger. The graph Laplacian is symmetric positive-definite and therefore admits the eigendecomposition

$$\mathbf{\Delta} = \mathbf{U} \mathbf{\lambda} \mathbf{U}^\top , \tag{2}$$

where $\mathbf{\lambda}$ is the diagonal matrix of non-negative eigenvalues and $\mathbf{U}$ is the orthogonal matrix whose columns are the eigenvectors of $\mathbf{\Delta}$.

Gaussian processes regression For a set $\mathcal{X}$, a stochastic process $f : \mathcal{X} \to \mathbb{R}$ is a GP if for any finite collection $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \subset \mathcal{X}$ the random variable $\mathbf{f} := f(X)$ is distributed according to a consistent joint Gaussian distribution (Rasmussen & Williams, 2006). We write a GP, $f(\cdot) \sim \mathcal{GP}(\mu(\cdot), k(\cdot, \cdot))$ with mean function $\mu : \mathcal{X} \to \mathbb{R}$ and kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, where k gives rise to the Gram matrix $\mathbf{K_{ff}}$ such that $[\mathbf{K_{ff}}]_{i,j}$ is computed by $k(\mathbf{x}_i, \mathbf{x}_j)$. For the remainder of this work, we assume without loss of generality that $\mu(x) = 0$ for all $x \in \mathcal{X}$.

For supervised learning problems, we have n observations $\mathbf{y} = \{y_1, y_2, \dots, y_n\} \subset \mathcal{Y}$ with corresponding inputs $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \subset \mathcal{X}$ and our aim is to make predictions of the function $\mathbf{f}^\star$ at some unseen points $\mathbf{X}^\star \subset \mathcal{X}$ where $\mathbf{f}^\star = f(\mathbf{X}^\star)$. Assuming observations with Gaussian noise, our

hierarchical model can be written as

$$p(\mathbf{f} \mid \mathbf{X}) = \mathcal{N}(\mathbf{0}, \mathbf{K}_{\mathbf{ff}}) \tag{3}$$

$$p(\mathbf{y} \mid \mathbf{f}) = \mathcal{N}(\mathbf{f}, \sigma^2 \mathbf{I}_n) \tag{4}$$

$$p(\mathbf{f}^\star \mid \mathbf{y}, \mathbf{X}, \mathbf{X}^\star) = \mathcal{N}(\boldsymbol{\mu}_{\mid \mathbf{y}}, \Sigma_{\mid \mathbf{y}}), \tag{5}$$

where

$$\boldsymbol{\mu}_{\mid \mathbf{y}} = \mathbf{K}_{*\mathbf{f}} \left(\mathbf{K}_{\mathbf{ff}} + \sigma_n \mathbf{I}_n\right)^{-1} \mathbf{y}^\top, \quad \Sigma_{\mid \mathbf{y}} = \mathbf{K}_{**} - \mathbf{K}_{*\mathbf{f}} \left(\mathbf{K}_{\mathbf{ff}} + \sigma_n^2 \mathbf{I}_n\right)^{-1} \mathbf{K}_{\mathbf{f}*}.$$

3 Case Study

Data We use NO_2 measurements from the Breathe London Mobile study where pollution measurement sensors were fixed to two Google street view cars (Hasenkopf et al., 2015). NO_2 levels were measured every 1-10 seconds as the cars drove around London. 600 journeys were made from Autumn 2018 to Autumn 2019 when the study ended. We use data from 6-7.30 AM on December 18th, 2018 in the suburb of Mitcham, which provides us with 695 measurements (Figure 2).

Graph representation To represent the roads of Mitcham as a graph, we use data from OpenStreetMap (Bennett, 2010) and let each junction be a vertex v. Two vertices are connected by an edge e in the graph if a road connects the corresponding junction pair. Intuitively, an adjacent pair of vertices are *more similar* the closer they are to one another. For this reason, we weight the edges in our graph by the inverse distance between the adjacent vertex pair. We next augment the graph with the NO_2 measurements. For each measurement, we identify the nearest edge, splice the edge at the measurement's coordinate, and reconstruct the graph with the measurement becoming a new vertex and two new edges being created to connect the measurement vertex to its adjacent junctions. We calculate length using the measurement's distance from its adjacent junction pair. We depict this process in Figure 3.

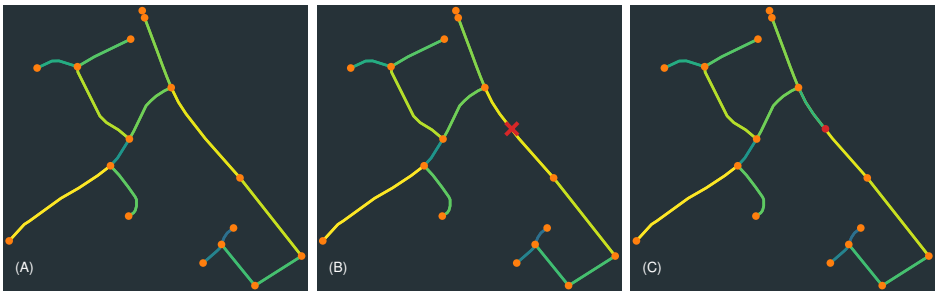
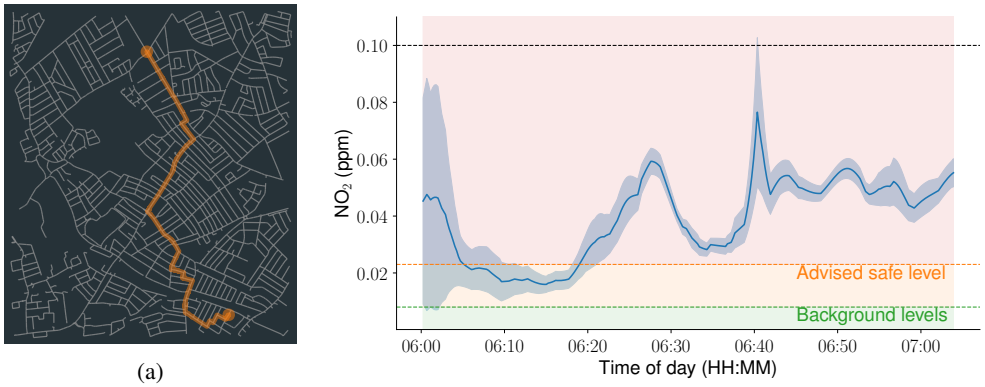


Figure 3: Schematic describing the inclusion of NO_2 measurements into a road network. With an underlying road network (A), vertices are junctions and edges are roads that connect two junctions. The colour of an edge denotes the edge's length. Measurements are made at points on the road (red cross, (B)) and the nearest edge is spliced at this point. We then reconstruct the edge as a pair of edges with length proportional (C).

Graph Gaussian process Equipped with a graph representation, we now seek to define a GP on the graph's vertices. We use the Matérn kernel $k : V \times V \to \mathbb{R}$ given in Borovitskiy et al. (2021). For a smoothness parameter $\nu \in \mathbb{R}_{>0}$ and lengthscale $\ell \in \mathbb{R}_{>0}$, the GP prior can be written as

$$\begin{aligned} p(\mathbf{f} \mid \mathbf{X}) &= \mathcal{N}\left(\mathbf{0}, \left(\frac{2\nu}{\ell^2} + \boldsymbol{\Delta}\right)^{-\nu}\right) \\ &= \mathcal{N}\left(\mathbf{0}, \mathbf{U}\left(\frac{2\nu}{\ell^2} + \operatorname{diag}(\boldsymbol{\lambda})\right)^{-\nu} \mathbf{U}^\top\right), \end{aligned} \tag{6}$$

using the eigendecomposition of the Laplacian matrix given in Equation (2).



(a) (b)

Figure 4: The NO_2 exposure experienced whilst walking the journey plotted in (a) plotted against time of day. The shading in (b) indicates the safeness of a given level (World Health Organization, 1997). In (b), we can see the model's predictive mean and standard deviation.

Handling heteroscedastic noise A homoscedastic noise model was assumed in Equation (4). However, the NO_2 values that we seek to model have a non-constant variance. To account for this, we replace the scalar noise term by a noise vector $\boldsymbol{\sigma^2} = \{\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2\}$. To learn each component of $\boldsymbol{\sigma^2}$, we parameterise it by a second GP $r : V \to \mathbb{R}^n_{>0}$ where positivity on $\boldsymbol{\sigma^2}$ is ensured by modelling $\log(\boldsymbol{\sigma^2})$. Under this new parameterisation, we replace the noise model from Equation (4) with

$$p(\mathbf{y} \mid \mathbf{f}, \mathbf{r}) = \mathcal{N}(\mathbf{f}, \exp(\mathbf{r})^2 \mathbf{I}_n) \,.$$

Introducing $\exp(r(V))$ makes inference for our model intractable. To resolve this, we lower bound the marginal log-likelihood using Jensen's inequality and learn a variational approximation to the posterior following Lázaro-Gredilla & Titsias (2011).

Inference procedure With each observation pair (v_i, y_i) corresponding to an individual vertex and corresponding NO_2 value, we assess our model's performance by partitioning the dataset into a training and testing set. We hold back 137 vertices for testing, and fit our model to the remaining 548 vertices. To validate our model, we compare it against its Euclidean and homoscedastic analogues. We construct a Euclidean representation of the data by computing the coordinate location of each vertex, making the input space $\mathbb{R}^2$. A Euclidean Matérn kernel is then fit to the Euclidean data and we compare these models both with and without the heteroscedastic noise term.

R^2 coefficients measure the amount of variation in the data that is explained by the model. A perfect model will attain an R^2 score of 1, whilst predicting the data's mean $\bar{y} = n^{-1} \sum_{i=1}^n y_i$ everywhere scores of 0. The predictive posterior density evaluated on test data quantifies how well the model explains unseen data points. We compute each of these metrics (Table 1) and find that the optimal model is one with heteroscedastic noise term and a graph kernel. We further validate our model through an analysis of residuals Appendix A.1 and report our experimental setup in Appendix A.2.

Table 1: R^2 coefficients and predictive posterior density scores on a held-back test set for the four models considered in Section 3. For both metrics, a higher score is better.

Kernel Domain	Noise Model	R^2	Predictive likelihood
Euclidean ($\mathbb{R}^2$)	Homoscedastic	0.636	0.511
	Heteroscedastic	0.716	0.672
Graph (V)	Homoscedastic	0.688	0.521
	Heteroscedastic	**0.722**	**0.724**

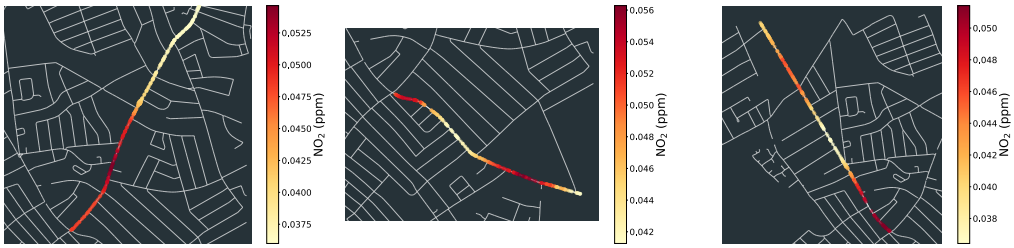
The improvement in model diagnostic measures (Table 1) alone is enough to justify using a graph kernel. However, we would like to explicitly acknowledge that the computational cost incurred

when fitting a graph kernel scales equally to a Euclidean kernel. Furthermore, the choice of kernel domain is a modelling decision, and in this work, a graph kernel is more appropriate as it more accurately reflects the fact that observations are collected whilst a car moves along the roads of Mitcham. Selecting a Euclidean kernel would implicitly imply the car could drive anywhere within Mitcham, an assumption that is untrue due to the presence of parks, buildings and other constructions that cars are not permitted to move over.

Exposure levels Here we seek to explore the NO_2 levels that an individual would be exposed to during a *typical* walking journey. To answer this question, we consider the journey plotted in orange in the left panel of Figure 4 and infer from our fitted GP model the NO_2 levels that would have been experienced at 6 AM. The World Health Organisation consider NO_2 levels of 0.023 parts per million (ppm) to be unsafe (World Health Organization, 1997). The prediction in Figure 4 we can see that a pedestrian would spend 51 minutes of a 64-minute walk at levels greater than this threshold. The significance of this should not be underplayed as the European commission consider more than 18 exceedances of 0.1ppm on an hourly average in a single year to be a violation of EU law.

Areas of greatest exposure The fine-scale resolution our model provides allows us to precisely identify streets the streets whose NO_2 exposure is greatest. To do this, we evaluate the GP's predictive mean at every street for which we have at least one observation. We then compute the mean NO_2 value for each street and identify the three streets with the largest mean NO_2 value. NO_2 predictions at vertices that are adjacent to two or more streets are repeated once per street. We identify Tooting Bec Road, Trinity Road and Balham High Road as the three streets with the largest mean NO_2 values.

To further investigate the NO_2 exposure experienced along these roads, we simulate 250 additional vertices for each of the three streets mentioned above and include the vertices into our graph in an identical manner to the process described in Figure 3. We refit our GP to the expanded graph and plot the GP's predictive mean at the vertices associated with each of the three above streets. In Figure 5, we can see that despite the average exposure being large on these streets, there are areas more exposed to high NO_2 levels than others. Such analyses are invaluable from a policy-making perspective as they enable us to better understand which people are more at risk of pollution-related illnesses.



(a) Balham High Road (b) Tooting Bec Road (c) Trinity Road

Figure 5: The three streets in Mitcham with the largest mean NO_2. In each panel we visualise the predictive mean along the respective street where we can see regions with larger NO_2 values.

4 Conclusions

We have presented a heteroscedastic graph GP that enables fine-scale inference of NO_2 exposure at a street level to be made. We have demonstrated the efficacy of this model in two pertinent examples (Figure 4 and Figure 5) that highlight the utility provided by this work. Whilst this work has focused on air pollution, the modelling approach is generalisable to other areas within the earth sciences. Examples of this are in hydrology where we seek to model the dispersion of chemicals and sewage in river networks, and ice sheet modelling where fractures make a Euclidean assumption invalid. Within the machine learning community, future work could consider higher-order interaction networks (Pinder et al., 2021) or augmenting the graph kernel with a temporal dimension (Nikitin et al., 2021) to model NO_2 temporally on road networks. We envision that as data collection technologies improve and we are supplied with more frequent observations at a finer granularity, approaches such as the one demonstrated in this article will become a core methodology for the earth science community.

REFERENCES

Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. TensorFlow: A system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pp. 265–283, 2016.

Jonathan Bennett. *OpenStreetMap*. Packt Publishing Ltd, 2010.

Geoff Boeing. OSMnx: A Python package to work with graph-theoretic OpenStreetMap street networks. *Journal of Open Source Software*, 2(12), 2017.

Viacheslav Borovitskiy, Iskander Azangulov, Alexander Terenin, Peter Mostowsky, Marc Deisenroth, and Nicolas Durrande. Matérn Gaussian processes on graphs. In *International Conference on Artificial Intelligence and Statistics*, pp. 2593–2601. PMLR, 2021.

Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using NetworkX. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

Charles R Harris, K Jarrod Millman, Stéfan J Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, et al. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020.

Christa A Hasenkopf, JC Flasher, Olaf Veerman, and Helen Langley DeWitt. Openaq: A platform to aggregate and freely share global air quality data. In *AGU Fall Meeting Abstracts*, volume 2015, pp. A31D–0097, 2015.

John D Hunter. Matplotlib: A 2D graphics environment. *Computing in science & engineering*, 9(03): 90–95, 2007.

Kelsey Jordahl, Joris Van den Bossche, Martin Fleischmann, Jacob Wasserman, James McBride, Jeffrey Gerard, Jeff Tratner, Matthew Perry, Adrian Garcia Badaracco, Carson Farmer, Geir Arne Hjelle, Alan D. Snow, Micah Cochran, Sean Gillies, Lucas Culbertson, Matt Bartos, Nick Eubank, maxalbert, Aleksey Bilogur, Sergio Rey, Christopher Ren, Dani Arribas-Bel, Leah Wasser, Levi John Wolf, Martin Journois, Joshua Wilson, Adam Greenhall, Chris Holdgraf, Filipe, and François Leblanc. geopandas/geopandas: v0.8.1. Zenodo, July 2020. doi: 10.5281/zenodo.3946761. URL https://doi.org/10.5281/zenodo.3946761.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Miguel Lázaro-Gredilla and Michalis K Titsias. Variational heteroscedastic gaussian process regression. In *ICML*, 2011.

J Lelieveld, K Klingmüller, A Pozzer, RT Burnett, A Haines, and V Ramanathan. Effects of fossil fuel and total anthropogenic emission removal on public health and climate. *Proceedings of the National Academy of Sciences*, 116(15):7192–7197, 2019.

Alexander G de G Matthews, Mark Van Der Wilk, Tom Nickson, Keisuke Fujii, Alexis Boukouvalas, Pablo León-Villagrá, Zoubin Ghahramani, and James Hensman. GPFlow: A Gaussian process library using TensorFlow. *Journal of Machine Learning Research*, 18(40):1–6, 2017.

Alexander Nikitin, ST John, Arno Solin, and Samuel Kaski. Non-separable spatio-temporal graph kernels via spdes. *arXiv preprint arXiv:2111.08524*, 2021.

Thomas Pinder, Kathryn Turnbull, Christopher Nemeth, and David Leslie. Gaussian processes on hypergraphs. *arXiv preprint arXiv:2106.01982*, 2021.

Carl Edward Rasmussen and Christopher K Williams. *Gaussian processes for machine learning*. Number 3. MIT press Cambridge, MA, 2006.

Hugh Salimbeni, Stefanos Eleftheriadis, and James Hensman. Natural gradients in practice: Non-conjugate variational inference in gaussian process models. In *International Conference on Artificial Intelligence and Statistics*, pp. 689–697. PMLR, 2018.

Alexander J Smola and Risi Kondor. Kernels and regularization on graphs. In *Learning theory and kernel machines*, pp. 144–158. Springer, 2003.

Maciej Strak, Nicole AH Janssen, Krystal J Godri, Ilse Gosens, Ian S Mudway, Flemming R Cassee, Erik Lebret, Frank J Kelly, Roy M Harrison, Bert Brunekreef, et al. Respiratory health effects of airborne particulate matter: the role of particle size, composition, and oxidative potential—the raptes project. *Environmental health perspectives*, 120(8):1183–1189, 2012.

Maciej Strak, Gerard Hoek, Maaike Steenhof, Evren Kilinc, Krystal J Godri, Ilse Gosens, Ian S Mudway, René van Oerle, Henri MH Spronk, Flemming R Cassee, et al. Components of ambient air pollution affect thrombin generation in healthy humans: the raptes project. *Occupational and environmental medicine*, 70(5):332–340, 2013.

World Health Organization. Nitrogen oxides environmental health criteria 188. *International programme on chemical safety*, 1997.