applied sciences

MDPI

Article

Random Forest Spatial-Temporal and State Space Models to Assess the Impact of Bushfire-Induced Aerosol Events on Ozone Depletion in Australia

Irene Hudson [1,*,†], Phillip Pedro-Suvorov [2,†] and Servet Kocak [3]

[1] Mathematical Sciences, School of Science, Royal Melbourne Institute of Technology (RMIT), Melbourne, VIC 3000, Australia
[2] Department of Mathematical Sciences, Royal Melbourne Institute of Technology (RMIT), Melbourne, VIC 3000, Australia; s3722617@student.rmit.edu.au
[3] School of Computer Science and Information Technology, Royal Melbourne Institute of Technology (RMIT), Melbourne, VIC 3000, Australia; servet.kocak@rmit.edu.au
* Correspondence: irene.hudson@rmit.edu.au
† These authors contributed equally to this work.

Abstract: Serious concerns exist that the increasing frequency of fires may delay the recovery of ozone given increasing temperatures due to climate change. Australian bushfires from September 2019 to February 2020 were catastrophic. A random forest spatial-temporal (RF sp) analysis using satellite data to detect an association between Australian bushfires and stratosphere ozone on the local depletion of ozone in the vicinity of fires in three regions of Australia (Pacific Ocean, Victoria, NSW) has shown a significant reduction in ozone attributable to aerosols from fires. By intervention analysis, increases in aerosols in all three regions were shown to have a significant and ongoing impact 1–5 days later on reducing ozone ($p < 0.0001$). Intervention analysis also gave similar periods of aerosol exceedance to those found by Hidden Markov models (HMMs). HMMs established a significant and quantifiable decline in ozone due to bushfire-induced aerosols, with significant lags of 10–25 days between times of aerosol exceedance and subsequent ozone level decline in all three regions.

Keywords: ozone; aerosol; Australian bushfires; spatial-temporal random forests; state space models; hidden Markov models; multiple change-point analysis; intervention analysis

Citation: Hudson, I.; Pedro-Suvorov, P.; Kocak, S. Random Forest Spatial-Temporal and State Space Models to Assess the Impact of Bushfire-Induced Aerosol Events on Ozone Depletion in Australia. *Appl. Sci.* **2024**, *14*, 9825. https://doi.org/10.3390/app14219825

Academic Editor: Rui Sun

Received: 31 July 2024
Revised: 6 October 2024
Accepted: 11 October 2024
Published: 28 October 2024

1. Introduction

The recent 2019–2020 bushfire season was one of the worst fire seasons ever recorded in Australia. Human death tolls were less than the previous Black Saturday bushfires in early 2009, where 173 people died; in fact, in the 2019–2020 bushfires, 34 confirmed deaths were recorded. The fires, nonetheless, had devastating immediate effects on the environment and the economy. Over 5900 buildings were destroyed and many billions of dollars of property were damaged across a vast area, with estimates ranging from 24.3 million hectares to 33.8 million hectares, making it the most costly and damaging bushfire season on record [1,2]. Environmentally, over one billion mammals, birds and reptiles were killed and a further estimated 2 billion were displaced. In fact, the World Wildlife Fund puts estimates of deaths of animals at nearly 3 billion [3]. It has been suggested that entire species may have become extinct because of these fires [4], and many threatened species are now facing extinction [5–7]. These are the readily measurable and observable effects of bushfires.

Climate change [8] predictions anticipate increases both in the number and intensity of bushfires in the future. It is widely known the overall effect of CO_2 on the warming of our climate but there is also evidence that aerosols from fires can also have an effect on the

ozone layer. The two reports [9,10], alongside many other papers, present the evidence for this. Accordingly, it is of vital importance to understand the relationship between bushfires-induced emissions of aerosol and other pollutants they release into the atmosphere on the one hand, and ozone depletion on the other. As illustrated in Figure 1 visible light is essentially unaffected by the ozone layer, whereas with harmful UV light the ozone layer scatters and absorbs it, significantly reducing the amount that reaches the surface.

There are other, less easily observed effects. In particular, fires are a major source of pollutants originating from the land; specifically, particulate matter in the form of aerosols and greenhouse gases (mainly CO_2) released into the atmosphere after the combustion of organic materials, i.e., plants. This plant matter is called biomass and when burned it releases aerosols [11].

An image illustrating that aerosols are composed of a highly diverse mixture of particles ranging in size from individual tiny particles no more than 10 nm in diameter to large clustering aggregates containing hundreds of particles up to 2 µm wide, is displayed in Figure 2.

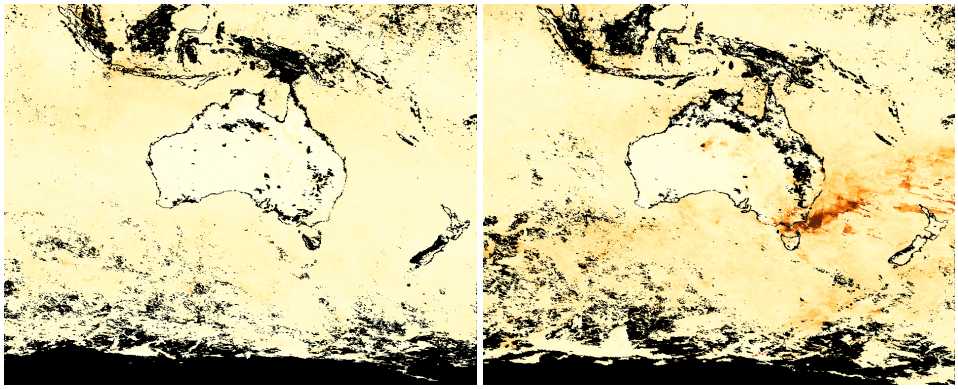

Figure 1. (**left**) NASA measurements of aerosols optical thickness over April 2019. (**right**) NASA measurements of aerosols' optical thickness over December 2020.

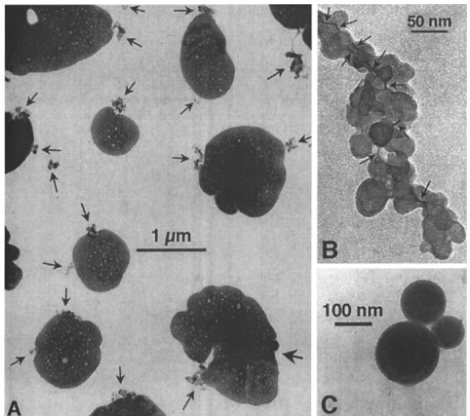

Figure 2. (**A**) Ammonium sulfate particles containing soot (marked by the small arrows) and fly-ash spheres (marked by the bold arrow in the lower-right corner). (**B**) In a typical branching soot aggregate; the arrows point to a carbon film that connects individual spherules within the aggregate. (**C**) Fly-ash spheres. Source: [12].

The aim of this study is to understand and quantify the spatiotemporal relationship between bushfires and ozone levels in the atmosphere. To this end we undertake statistical analysis, using modern empirical and spatiotemporal statistical techniques of bushfire and ozone layer data to discover and measure correlations both geographically and over time between the ozone and bushfire-induced aerosol datasets. We aim to analytically study and provide evidence for or against the hypothesis that bushfires result in a localized effect on the ozone layer in the vicinity of the bushfire or downwind of it.

2. Research Area

Australia has regular bushfires throughout mid-to-late summer, and often earlier in the season. Depending on global meteorological conditions, some seasons are worse than others. A particularly bad season was the summer of 2019–2020. Devastating bushfires spread through parts of New South Wales, Victoria and South Australia during the months of December 2019 to February 2020. The 2019/2020 bushfire season was devastating and destructive, releasing millions of tonnes of CO_2, and creating much damage to the environment, property, wildlife, flora and fauna. These recent 2019–2020 'Black Summer' bushfires altogether put a serious burden on health systems, resulting in the loss of a considerable number of lives; 34 in total as well as around 3500 hospitalizations estimated [13] nationwide. A devastating effect on the environment and economy was also recorded [14] and millions of tonnes of CO_2 were released [2] as a result of these fires. Climate change [8,15,16] predictions anticipate increases both in the number and intensity of bushfires in the future.

Specifically in relation to Australia consider the left panel of Figure 1 which shows constant and relatively low aerosol optical thickness levels over the month of April in 2019 when there were no bushfires in Australia. The right panel depicts the aerosol optical thickness throughout December 2020 during the Black Summer bushfires. It shows high aerosol optical thickness levels which are localized to the south-east of Victoria and New South Wales where the majority of fires occurred. Figure 1 suggests that there is a relationship between bushfires and aerosols.

Evidence substantiating these claims in the IPCC report concerning this increased intensity and its consequences are presented in several papers. They show that the smoke released from bushfires during Australia's Black Summer (2019–2020) could lead to persistent stratospheric warming [15], and this could result in large disturbances in the composition of the stratosphere [17], and serve as a potential predictor of the consequences of future extreme bushfire events [16].

This study aims to undertake a statistical analysis using two different tree-based techniques from geospatial statistics of bushfire and stratospheric ozone layer data to discover and measure correlations, both geographically and over time, between these datasets. The specific tree-based method that is used for this analysis will be random forests for spatial-temporal data. We shall also model changing states over the 181-day period covering September to February of ozone with respect to aerosol using Hidden Markov models [18], where the variables analyzed are total sum ozone (TSO) and total sum aerosol (TSA) which are defined in Section 3.1 on data. Our aim is that implementing these methodologies will help establish whether and how ozone levels at high altitudes are affected by bushfires.

It is expected that the distributional properties of the total cumulative sum of ozone/aerosol values contained within a localized region must change over time. It is unlikely that on days without bushfires, there will be a rise in local aerosol but on days with bushfire, it is highly probable to observe a rise in local aerosol. The studies cited above, [15,19,20], have established a relationship between aerosol and ozone; thus, changes in aerosol levels would also result in changes in ozone but not necessarily in the same locality and possibly with a time delay.

Several papers [15,21] examine the connections between specific bushfires (2017 western Canadian wildfires and the Australian 2019–2020 Black Summer bushfires) and the subsequent depletion of ozone levels in the vicinities of the Antarctic and Arctic. These papers, however, *do not localize the effect bushfires have on the ozone levels*. For example, it is known that ozone depletion tends to accumulate in the Antarctic and Arctic regions [22] where we would expect depletions to be seen, but we would also anticipate localized depletion closer to the fire even possibly within a shorter time scale and the focus is to test this in our study. Statistical evidence indicates that bushfires have been increasing in frequency in Victoria since 1995 [23] and also nationally.

The Australian bushfires of 2019–2020 severely affected the air quality of the eastern half of Australia. Vast plumes of smoke were projected high into the atmosphere and were

transported around the world, even reaching across the Pacific Ocean into the continent of South America [24]. Measurements taken from satellites have recorded that the smoke in these plumes has caused noticeable changes in the composition of the upper atmosphere and, more importantly, a decline in stratospheric levels of ozone [20,25]. But the mechanism for how smoke from the fire might lead to, or exacerbate, ozone depletion is still uncertain. Solomon et al. [19] suggested that the particulate matter of this smoke does contribute to the destruction of stratospheric ozone contrary to the atmospheric warming hypothesis suggested by Yu et al. [15,21].

The mechanism for ozone depletion from bushfires is depicted in Figure 3, taken from [26]. Briefly, oxygenated organic compounds of the type found in aerosols from bushfires can absorb hydrochloric acid and carry it into the stratosphere. Here, ultraviolet light from the sun breaks down the hydrochloric acid to release atomic chlorine, which, in turn, breaks down the ozone molecules.

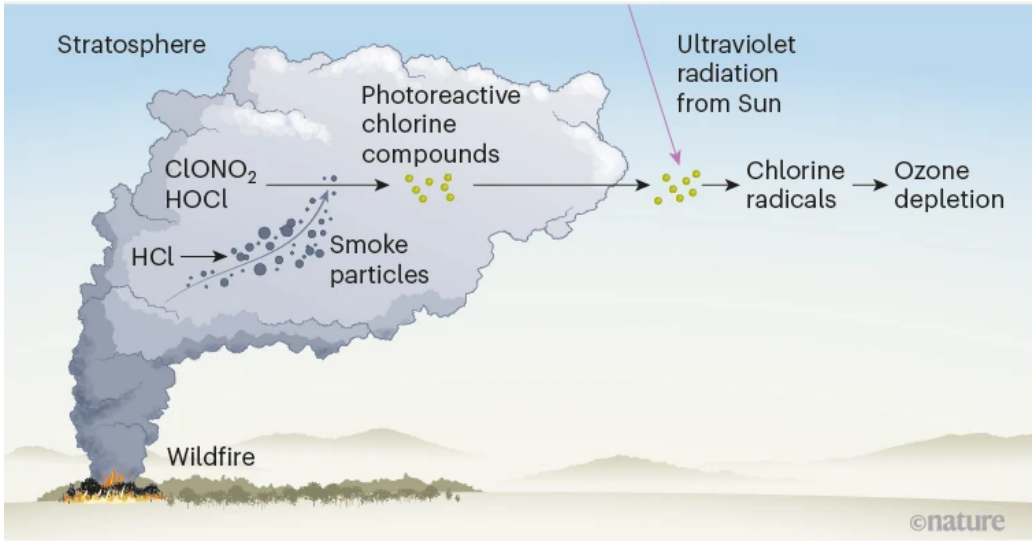

Figure 3. Pathway for smoke particles from intense bushfires to enter the stratosphere and for those same particles to, in turn, lead to ozone depletion. Source: [26].

Our study introduces new methods into this particular area of geospatial and temporal ozone level prediction. Tree-based methods, specifically, random forests have been applied in other areas of geophysical research (see [27]) but as far as we are aware, these methods have not yet been applied in any work specific to bushfire-induced ozone level change research. Furthermore, we apply State Space Models (SSMs) to the time series of Total Sum Ozone and Total Sum Aerosol columns. SSMs provide a general framework for analyzing deterministic and dynamical systems observed through a stochastic process [28]. The mathematical framework of SSMs includes latent processes and Hidden Markov models (HMMs), as applied in this study. We refer the reader to a recent study [29] that used a specific SSM, namely the Dynamic Linear Regression Model (but without accommodating so-called intervention points) to study the impact of total ozone level depletion on increased ultraviolet radiation associated with increased malignancy and cancer events [30–33].

3. Methods

3.1. Data

Daily ozone and aerosol measurements observed by the ozone mapping and profiler suite (OMPS) aboard the Suomi National Polar-orbiting Partnership (Suomi NPP) satellite [34] were retrieved using the OPeNDAP framework. `OPeNDAP` stands for *Open-source Project for a Network Data Access Protocol*. It provides a way to share data more easily across the world. Documentation for how to use `OPeNDAP` can be accessed via the following

link (https://www.opendap.org/support/user-documentation, accessed on 4 May 2022). As a satellite passes along a polar orbit it traces out a swath, with the width depending on the field of view of the measurement device.

An insufficient field of view results in coverage gaps between each swath where no data are measured. The measurements accumulated along the swaths are converted into a daily gridded map product: data that have been processed into a standard format and are accessible to a user. This is conducted using a gridding algorithm [35]. The dataset used contains measurements starting on 25 January 2012, while data for it are still being added. The spatial coverage is global, but is incomplete where coverage gaps occur, particularly for UV Aerosol index measurements because of a narrower field of view of the instrument. The ozone values are measured in Dobson units (DU), but the UV Aerosol index has no units attached. Importantly, Dobson units (DU) are a measure of the amount of a gas, usually ozone, in a vertical column descending through the atmosphere; in the case of ozone it is called *total column ozone*.

Konovalov 2011 [36] modeled the effect of bushfire emission on the near-surface concentrations of ozone. In our study, we focus only on the effect on total column ozone (i.e., ozone within a column of air from the Earth's surface to the top of the atmosphere). Figure 4, taken from https://csl.noaa.gov/assessments/ozone/2018/twentyquestions/, accessed on 30 June 2023, provides a schematic of ozone levels at different heights in the atmosphere. This particular ozone level diagram indicates how ozone varies with height in the tropics. Notice that predominantly ozone is concentrated in the stratosphere, with its distribution peaking at around 23 km.

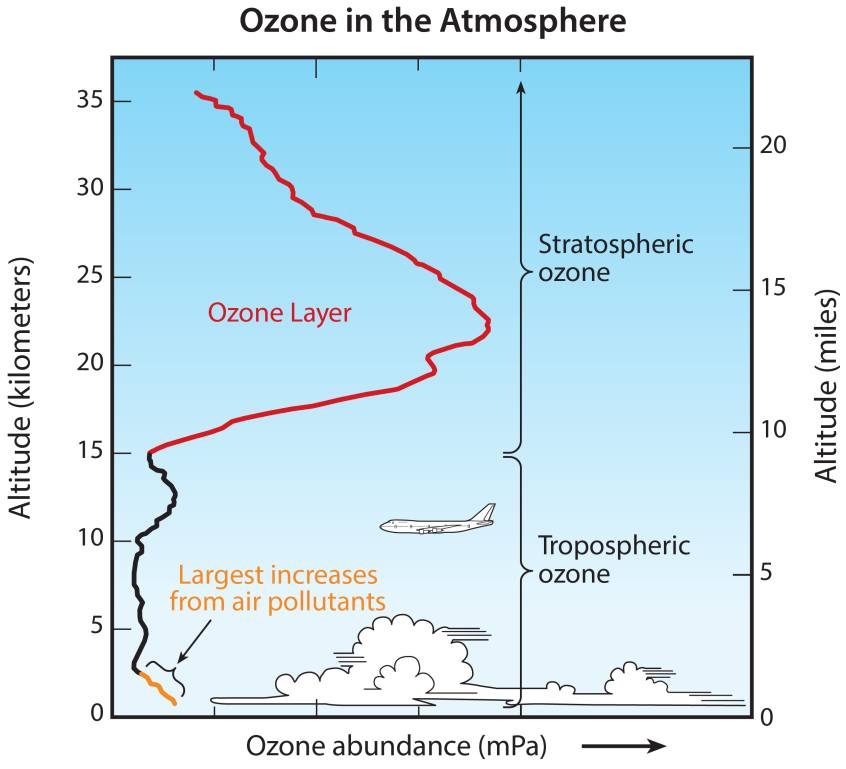

Figure 4. Schematic of ozone levels ozone within a column of air from the Earth's surface (0 km) to the top of the atmosphere (>35 km). Taken from https://csl.noaa.gov/assessments/ozone/2018/twentyquestions/, accessed on 30 June 2023.

The unit DU is defined per unit of area—the area that is occupied by the horizontal cross-section of the total column—and is the thickness (in units of 10 μm) of the layer that would be formed if the total column were subjected to standard pressure and temperature. For example, the average for a given total column of ozone is about 300 DU, equivalent to a layer with a thickness of 300 [10 μm], or 3 mm. Throughout this study, any mention of

ozone value means the total column ozone, unless explicitly specified. Ozone value data are loaded into the R programming environment using the Terra package (https://github.com/rspatial/terra, accessed on 22 May 2022).

The study area considered in this work is contained between −9.5 and −45.5 degrees latitude, and 129.5 and 167.5 degrees longitude, covering the eastern portion of Australia, including parts of SA, NT and of the Pacific Ocean; it is depicted by a checkerboard pattern in Figure 5. This area will be referred to here as the study area. All the ozone/aerosol data are from the 2019–2020 Australian bushfire period covering the dates from 1 September 2019 to 29 February 2020 and will be called the study time period. There is a total of 182 days between these two dates. On the 68th day (7 November 2019), the first sign of bushfire aerosol emissions appears and the last relatively small emission of bushfire aerosols occurs on the 159th day (6 February 2020). Total sum ozone (TSO) and total sum aerosol (TSA) values are also computed. These total sum values are equal to the sum of all cell values contained within the rectangular region.

We have defined three regions, located on or off the east coast of Australia, and they will be referred to as Region 1, Region 2 and Region 3. The longitude (lon) and latitude (lat) coordinates of their respective left/right and top/bottom edges, and a description of their location on the map of these rectangular regions are as follows:

- Region 1—lon 160.5–165.5 and lat −29.5–−32.5, Pacific Ocean approximately 500 km east of the midpoint between Sydney and Brisbane.
- Region 2—lon 146.5–151.5 and lat −36.5–−39.5, Coast of Victoria.
- Region 3—lon 148.5–153.5 and lat −31.5–−34.5, Coast of NSW.

The three regions are displayed in Figure 5.

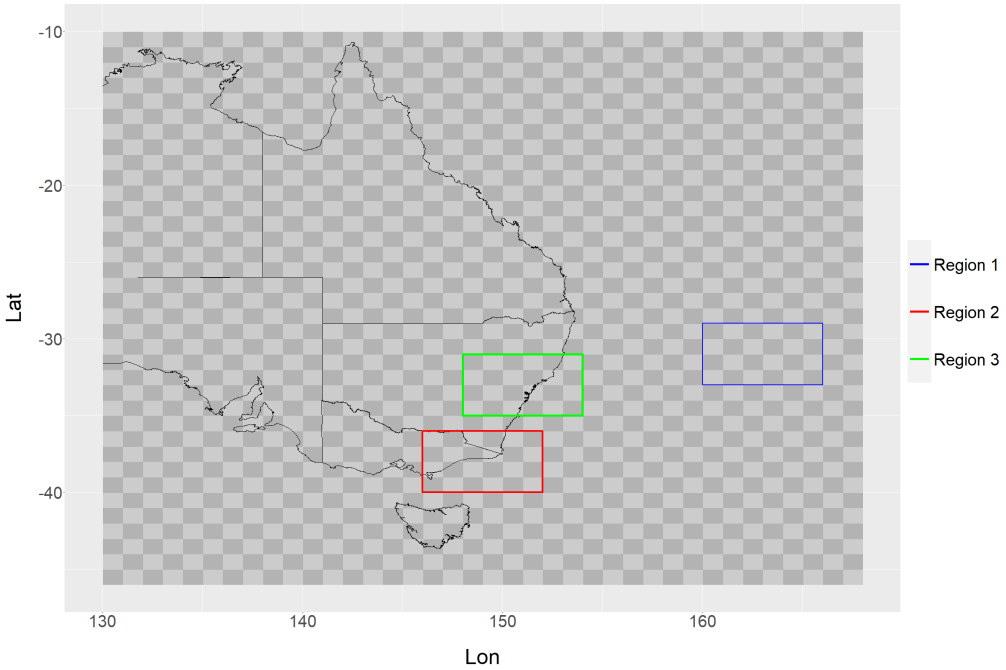

Figure 5. Outline of the map of Australia showing the locations of the three regions. The regions in order from top to bottom: Region 1 (Blue), Region 2 (Red) and Region 3 (Green). The checkerboard shading pattern identifies the study area.

The choice of these three regions was guided by the following considerations:

1. The location of the source of the majority of aerosol emissions; which were associated with the bushfires reported in Victoria and New South Wales (NSW).
2. The average wind direction during the time period; in a north-easterly direction towards Region 3.

3. These choices could potentially simplify the task of finding the correct day offset (the time delay or day offset between the release of aerosol and the resultant changes in ozone).

The location of Region 3 (NSW) was selected because, on a number of days, it had an unusually high level of aerosol, and hence had a higher chance for interactions between aerosol and ozone to occur and improved the ability of the methods to more accurately detect that relationship. We are also interested in the time delay, or day offset, between the release of aerosol and the resultant changes in ozone. Whatever the number of days offset may be, say 3 days, aerosol disperses in the atmosphere and is transported by the wind during these 3 days. From the visual inspection of the aerosol plot (see Figure 6), it is observed that aerosol ends up spreading over a significant area around where Region 3 is located.

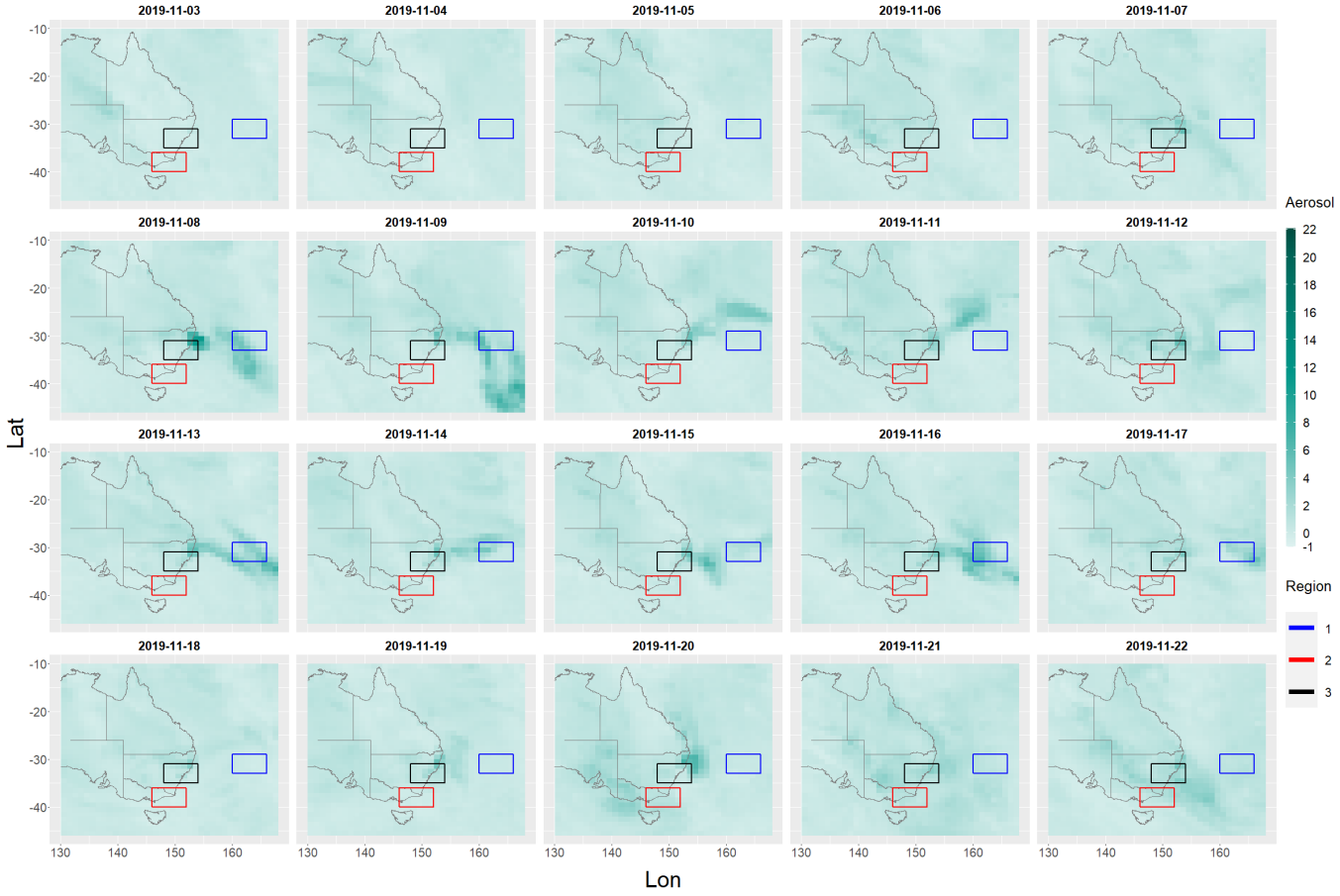

Figure 6. Observed from data of daily gridded aerosol values plot of entire study area over the 20 days from 3 November 2019.

For better causal modeling we have included additional parameters, the aerosol for each grid cell within the study area, the cumulative number of days since 3 November 2019 (`cdate`) and the day of the year (`doy`) over the dates of the study time period.

From the visual inspection of the Ozone plot (see Figure 7), it is observed that the ozone levels diminish, and these areas of decrease are observed over all the regions over the 20 days post November 3. Also, as a comparison, Figures 8 and 9 display the time series profiles for Total Sum Ozone and Total Sum Aerosol, respectively, over time and all 3 regions.

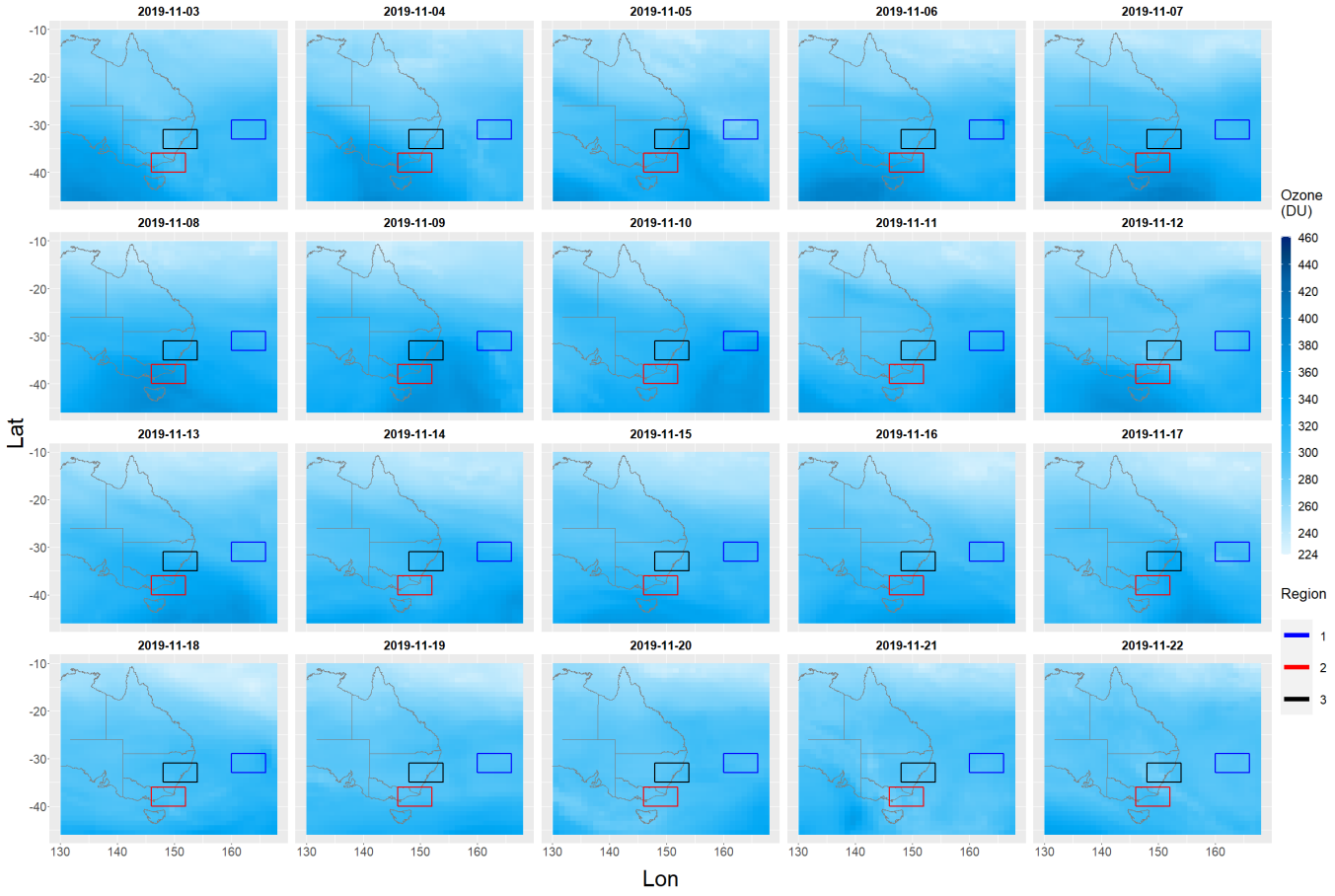

Figure 7. Observed from data of daily gridded ozone values plot of entire study area over the 20 days from 3 November 2019.

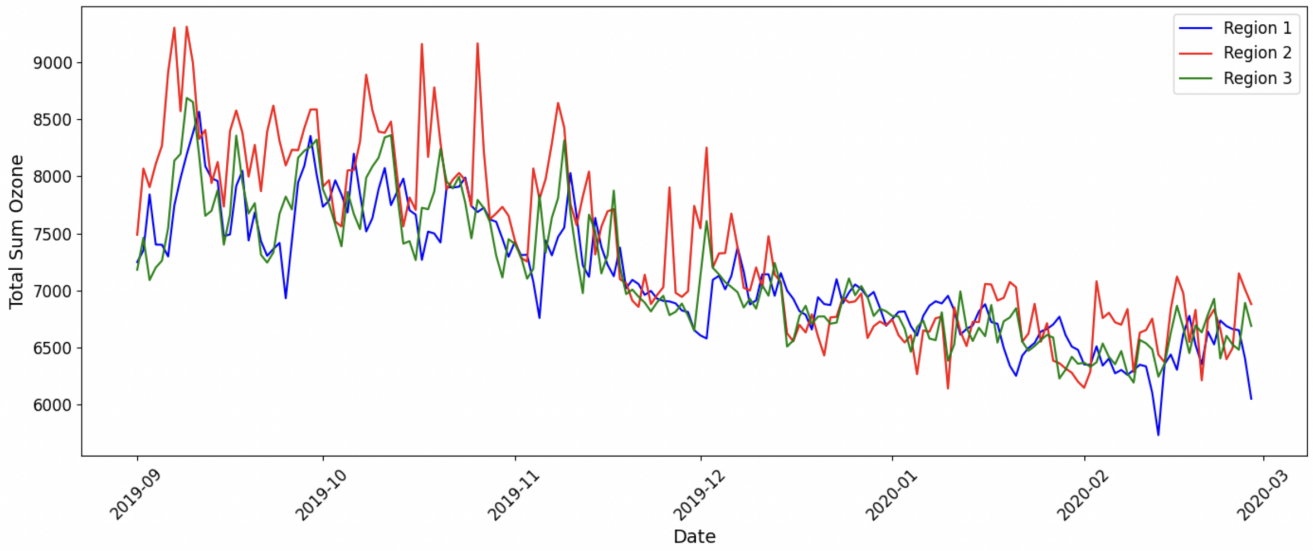

Figure 8. Time series plot of Total Sum Ozone for each region.

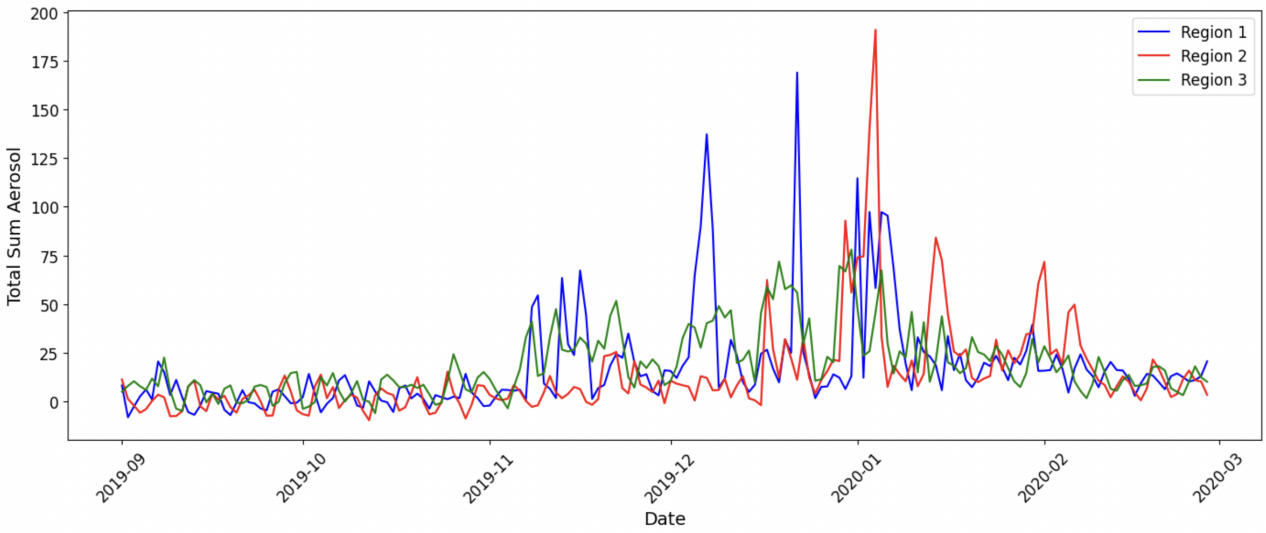

Figure 9. Time series plot of Total Sum Aerosol for each region.

3.2. Methodology

Our methods include random forests, a technique that has been widely adopted by the machine learning community, and proven effective in a large variety of classification and regression problems where physical models are limited, and more recently has been used successfully in modeling variables of a spatial and spatiotemporal nature. Also, we analytically determine change-points by Binary Segmentation [37] and change-points derived as times of change of states in ozone by HMM [18] analyses to further assist the interpretation of our spatiotemporal maps predicted from `RFsp`.

The aim is to further elucidate the temporal and possibly lagged effects between levels induced by bushfires and the depletion of ozone. We anticipate that this approach will provide evidence for or against the hypothesis that bushfires can result in a localized change in the ozone layer in the vicinity of the bushfire or downwind of it.

3.3. Random Forest for Spatial Data

Random Forests [38] is a method for classification and regression based on the idea of bootstrapping, that is, randomly selecting with the replacement of multiple samples from the data, applying a classifier/regression model, and then averaging over the samples—this is called *bagging*. In a generic context, if $x_1, x_2, \ldots, x_n$ are input measurements (independent variable, which could be multi-dimensional) and $y_1, y_2, \ldots, y_n$ are corresponding outcomes (dependent variable), then a subset S of these data of size B is chosen randomly and a classification or regression tree t_S is computed, based on this subset S. The result of the bagging is obtained by averaging t_S over all such random selections. We emphasize that this random selection is conducted with replacement. This methodology is known to reduce variance while not increasing bias.

Recently there has been interest in applying RFs to geospatial temporal statistical problems [27]. One important issue to overcome is that, in many ways, the standard random forests approach is indifferent to the spatial nature of the data, specifically in relation to what geographical location it represents and its attributes at each location. Hengl et al. [27] extend the basic random forests model to overcome this defect. This was achieved by introducing a measure of distance between observation points into a random forest model. That is, for each point/location ℓ, with an accompanying observed value Y_ℓ, a covariate made up of the distances to all other locations is given. In practice, in the model this amounts to the distance matrix between all pairs of locations with a Y_ℓ. These are referred to as buffer distances.

Random Forests for spatial data (RFsp) as proposed by Hengl et al. [27] describes the generic equation for an `RFsp` system, as follows, where

$$Z_\ell = f(X_{G_\ell}, X_{R_\ell}, X_{P_\ell}), \tag{1}$$

where Z_ℓ is the predicted value at location ℓ, X_{G_ℓ} contains the distances from location ℓ to ℓ_j for all $j \in \{1, \ldots, N\}$, X_{R_ℓ} and X_{P_ℓ} represent possible additional covariates; two types—surface reflectance and process based (e.g., wetness index).

The function f is learned by the random forest, thereby there is no covariance function used. Hengl et al. [27] combine the methods of random forests with multiple linear regression and kriging [39] to handle the spatial variation issues.

As delineated in Hengl et al. [27] the RF is a data-driven statistical method. The final predictions are the average of predictions of individual trees [38,40,41]. Statistical issues of RF estimation have been detailed recently in [41,42].

We also refer the reader to a recent study by Wong et al. [43] focused on an applied comparison of four 'fast' geostatistical modeling methods and the software available to implement them. Along with Spatial Random Forests (SpRF) three other methods were Integrated Nested Laplace Approximation (INLA), tree boosting with Gaussian processes and mixed effect models (GPBoost)) and Fixed Rank Kriging (FRK). This application involved estimating malaria prevalence on two different spatial scales, at the country and continent-scale, with the performance of the four methods on these data based on accuracy, computation time, and ease of implementation.

As noted by Hengl et al. [27], amongst the suite of machine learning algorithms (MLA), the RF has been confirmed to be efficient for spatial predictions. Additionally, RF is available in R through several packages such as `randomForest` [44] or the computationally faster alternative `ranger` [45], which is chosen in this study (see the results; Section 4.1). Complete benchmarking of the prediction efficiency is documented in R code and is available via the GitHub repository at https://github.com/thengl/GeoMLA, accessed on 30 August 2023.

A fundamental concept of much of the work on geospatial statistics is a **Gaussian process**. These are assignments of Gaussian random variables X_ℓ (which can be multi-dimensional) to locations ℓ with a covariance structure, typically describing how X_ℓ relates to $X_{\ell'}$ for ℓ' a neighbour of ℓ [39,46]. In more complicated situations the "neighbouring locations" can be ascribed by a graph, as in the case of SPAM trees [47].

Kriging is a technique to utilize knowledge about spatial autocorrelation in modeling and prediction [48]. Kriging [39] uses the Gaussian process model to estimate "missing data" at locations where there are no measurements from locations using the covariance structure. If measurements of a random variable Z_ℓ are known at locations $\ell_1, \ell_2, \ldots, \ell_m$, then the missing data at a location ℓ can be inferred from these by a formula like

$$\hat{Z}_\ell = \sum_{k=1}^{m} w(\ell, \ell_k) Z_{\ell+k}, \tag{2}$$

where $w(\ell, \ell_k)$ is a weight obtained from the covariance matrices connecting ℓ with ℓ_k.

Kriging requires a rigid statistical structure for this "interpolation" of missing data. The Random Forest spatial prediction method of Hengl et al. [27], however, provides an alternative that avoids such rigid assumptions. We also refer the reader to Fixed Rank Kriging (FrK) [49,50]. Fixed Rank Kriging is kriging with the class of non-stationary covariance functions, which is a relaxation of the assumptions of Ordinary Kriging.

Implementation of Generalization of `RFsp` to Spatial-Temporal Data

In the previous section, we showed the generic equation of an `RFsp` (Equation (1)), and we use this to formulate the model in this study. Following from Equation (1), the `RFsp` model is

$$Z_\ell = f(X_{G_\ell}, A_\ell, T_\ell), \tag{3}$$

where ℓ is now extended to a spatiotemporal domain, by adding a time dimension, $\ell \in \underline{R}^2 \times [0, \infty)$. We define A_ℓ as the aerosol value at the spatiotemporal location ℓ ,T_ℓ is the time covariate, namely, the cumulative number of days since 1 January 1970 (`cdate`) and day of the year (`doy`) over the dates of the study time period. X_{G_ℓ} denotes the distance matrix; the distances between all pairs of locations ℓ within the defined study area; that is, the ij entry is the distance between the ith and jth locations.

We follow the steps below in order to calculate the distances for all pairs of grid cells, and then the resultant *design matrix* is used in the `RFsp` algorithm (implemented in the `ranger` package).

Procedural steps via the `ranger` package using Equations (1)–(3):

1. Create an $n \times n$ matrix where n is the number of cells in the grid and each row of the matrix corresponds to an individual cell. Each cell is dependent, not only spatially with respect to surrounding cells, but also dependent on the amount of aerosols contained in the cells.
2. For each row of the matrix, all of those aerosol values from each cell are included. This step adds n columns to the design matrix.
3. Next, we repeat the above process for each day of the study period.
4. Also included in the model are `cdate` and `doy` for the number of days of observation, which we denote by r.
5. As there are n cells of data for each day, all `cdate` and `doy` values are repeated n times, then all together we have a $(r \times n) \times (2n + 2)$ matrix.

These steps create a large matrix, which is the design matrix.

Quantile prediction intervals are used as a measure of the prediction error in the ozone predictions obtained from the application of `RFsp`. These are defined (see [51]) in terms of the cumulative conditional distribution function $F(y|x)$ of the predictor $Y = y$ conditioned on the covariate $X = x$, where Y is a real-valued random variable. Then, the α-quantile $Q_\alpha(x)$ of Y conditioned on $X = x$ is defined to satisfy

$$P(Y < Q_\alpha) = \alpha; \tag{4}$$

that is,

$$Q_\alpha(x) = \inf\{y : F(y|X = x) \geq \alpha\}.$$

Now, the *βth quantile prediction interval* expressed in percentages of the total distribution, is the interval

$$[Q_\alpha(x),\ Q_{\alpha'}(x)], \tag{5}$$

where

$$\alpha = 0.5 \times 1 - \frac{\beta}{100}, \qquad \alpha' = 1 - \alpha.$$

In this study, we perform quantile regression forests to obtain an estimate of these conditional quantiles. Estimates of a lower $\beta = 68.2\%$ ($\alpha = 0.159$) quantile, a middle $\alpha = 0.5$ quantile (a quantile which is symmetric around the mean containing half of the distribution), and an upper $(\alpha = 0.841)$ quantile are used.

Models based on the `RFsp` generic framework for spatial and spatiotemporal prediction [27] as a machine learning algorithm (MLA) discussed in Section 3.3 will be tested for model accuracy, given by R^2 and mean square error (MSE) using five-fold cross-validation [52]; where cross-validation is central to many statistical algorithms and workflows, in particular, for predictive modeling frameworks in ecology [52,53].

3.4. Change Point Detection

Analytically, one approach for the determination of change-points (CPs) is by Binary Segmentation. In addition, CPs can be determined as stochastic points of changing states using Hidden Markov Models (see Section 3.5 on Methods for HMM). We use both of these

methods to assist the interpretation of the ozone spatiotemporal maps predicted from `RFsp`. Note, that HMMs can also provide a further complementary understanding of the temporal and the lagged effect between elevated aerosol levels induced by bushfires and depletion of ozone levels.

Change-point detection is a pivotal area in the statistical analysis of time series. A single change-point detection problem is detailed in [54,55], assuming the model is Gaussian. Various methods are available also for the detection of multiple change-points. Three of them are available in the R package `changepoint`. We briefly describe only one of the three methods here, namely Binary Segmentation. Binary Segmentation was first described in the paper [56] by Edwards and Cavalli-Sforza in 1965.

Implementation of the Change Point Analysis

We have used the R package `changepoint` in our study, refer to [37]. The `changepoint` package implements the methods described above for both single change-point and multiple change-point detection. A standard approach to such problems is a *generalized likelihood ratio statistic*. The method is described in [57]. This also serves as a reference for the R package `changepoint` (https://cran.r-project.org/web/packages/changepoint/index.html, accessed on 14 June 2022). The tests vary according to the families of distributions being analyzed and the specific parameters being tested.

Several other R packages are available for change-point detection, including `strucchange` (https://cran.r-project.org/package=strucchange, accessed on 25 June 2022), `cumSeg` (https://cran.r-project.org/package=cumSeg, accessed on 14 June 2022), and `stepR` (https://cran.r-project.org/package=stepR, accessed on 15 June 2022). These are discussed in detail and their methodologies are compared both algorithmically and in terms of results against specific problems in [58].

3.5. Hidden Markov Model Analytic Approach—A Discrete Valued State Process

The Hidden Markov Model is a probabilistic model proposed by Baum and Petrie [59] to describe the statistical properties of random processes with parameters. A Hidden Markov Model (HMM) generates a hidden sequence of states from an observable sequence of observations. Mixture and hidden Markov models are statistical models that are useful when an observed stochastic, dynamic system over time occupies a number of distinct "regimes" or unobserved (hidden) states [60].

HMMs are widely used in a variety of fields, including environmental science, artificial intelligence, biology, and psychology [61]. HMMs have been shown to be useful specifically in weather prediction, robotics, detecting protein homologies and computer vision [60,62]. Hidden Markov Models (HMMs) are also extensively used in applications, such as natural language processing (NLP) and finance [63]. Likewise, a recent utilization of HMMs has been in earthquake prediction [64] and in computational biology, where an HMM may be used to model the evolution of a protein or DNA sequence, and observations generally are a sequence of amino acids or nucleotides [65]. In speech recognition, for example, an HMM can be used to model the underlying sounds or phonemes that generate the speech signal, and the observations generally are the features extracted from the speech signal [62].

Hidden Markov Models can be viewed as a generalization of two types of simpler models. Firstly, HMMs are a generalization of simple Markov models, where the generalization involves additional modeling of the error structure. Secondly, Hidden Markov models can be viewed as an extension of mixture models, to model transitions between states over time. Here, the generalization involves additional modeling of the sequential nature of the data. In addition, recent advances in HMM theory [18,60] allow the easy incorporation of time-dependent covariates into the model. This means that both the transition and the emission probabilities can depend on a set of covariates, so that they can vary over time and thereby the traditional and unrealistic stationarity assumptions need not be made.

HMM model sequences/series are either derived from continuous or discrete probability distributions. HMMs are related to state space and Gaussian mixture models

(GMMs) [66] where HMMs aim to estimate the state(s) which have given rise to the observations. The states are unknown or 'hidden' and HMMs attempt to estimate the states similar to an unsupervised clustering procedure. Hidden Markov Models can classify observable sequences into unobservable state sequences. State transitions refer to the random change in states of the Markov process in discrete time. The Markov model thus follows the concept of a memory-less property, i.e., the transition from one state to another state depends only on the present state [67].

The HMM algorithm involves defining the state space, observation space, and the parameters of the state transition probabilities and observation likelihoods. The training of the model generally uses the Baum–Welch algorithm or the forward-backward algorithm and decoding the most likely sequence of hidden states is performed by the Viterbi algorithm [68], as well as evaluating the performance of the model. The Baum–Welch algorithm is a version of the Estimation Maximization (EM) algorithm [60].

To date, HMM modeling has been used to analyze ozone levels by several authors. Two examples of such research, specifically in predicting the maximum 8 h average ozone concentration, are Zhang et al. [69] who use a Hidden Markov Model for measurements assumed to come from a Gamma distribution, rather than a Gaussian distribution. The other study is by Sun et al. [70] who use a supervised HMM combined with generalized linear models.

Implementation of HMM Analysis

Here, we investigate the possible effects of bushfires-induced emissions of aerosols on the levels of ozone in the total column. This is achieved by observing and analyzing changes in states rather than by analyzing changes in actual observed values (e.g., ozone values), as in traditional time series analysis [71].

We use the `depmixS4` package in R [18]. DepmixS4 specifies and fits hidden Markov models, optimized with the EM algorithm. Models can also fit multiple sets of observations and covariates (e.g., aerosol levels). The performance of the HMM can be evaluated using various metrics, such as AIC or BIC when using R `depmixS4` by Visser et al. [18], or accuracy, precision, recall, or F1 score when using Python `hmmlearn`—a set of algorithms for unsupervised learning and inference of HMMs.

In modeling we add covariates to the HMM models as follows: Let S denote the states, and $N = T$. The joint likelihood of observations $O_{1:T}$ and latent state $S_{1:T} = (S_1, \ldots, S_T)$, with model parameters θ and covariates $z_{1:T} = (z_1, \ldots, z_T)$, can be written as:

$$P(O_{1:T}, S_{1:T} \mid \theta, z_{1:T}) = \pi_i(z_1) b_{S_1}(O_1 \mid z_1) \prod_{t=1}^{T-1} a_{ij}(z_t) b_{S_t}(O_{t+1} \mid z_{t+1}). \quad (6)$$

Note, the following details:

1. S_t is an element of the set of n latent states.
2. $\pi(zi) = P(S = i|z_1)$, denotes the probability of state i at time $t = 1$ with covariate z_1
3. $a_{ij}(z_t) = P(S_{t+1} = j|S_t = i, z_t)$, provides the probability of a transition from state i to state j with covariate z_t.
4. b_{S_t} is a vector of observation densities $b_j^k(z_t) = P(O_t^k|S_t = j, zt)$, namely, the conditional densities of observations O_t^k associated with latent class/state j and covariates $z_t, j = 1, \ldots, n; k = 1, \ldots, m$.
5. Specific analytic steps to model daily changes in ozone and aerosol levels by HMMs are as follows:
   (a) Compare the days when the ozone state changes with respect to the observed daily aerosol levels.
   (b) Compare the trends across the three geographical regions (i.e., Regions 1, 2 and 3).

Figure 10 part (B) shows our paradigm for HMM modeling which entails adding covariates (e.g., aerosol levels over time).

For modeling ozone, with and without aerosol as a covariate, we use function `depmax` to create an S4 object of class depmix, which needs the `fit` function to fit and to optimise the parameters. The output of the `fit` function comprises a transition probability matrix, mean intercept and standard deviation table. The probabilities are the estimated probabilities of transitions between states [18].

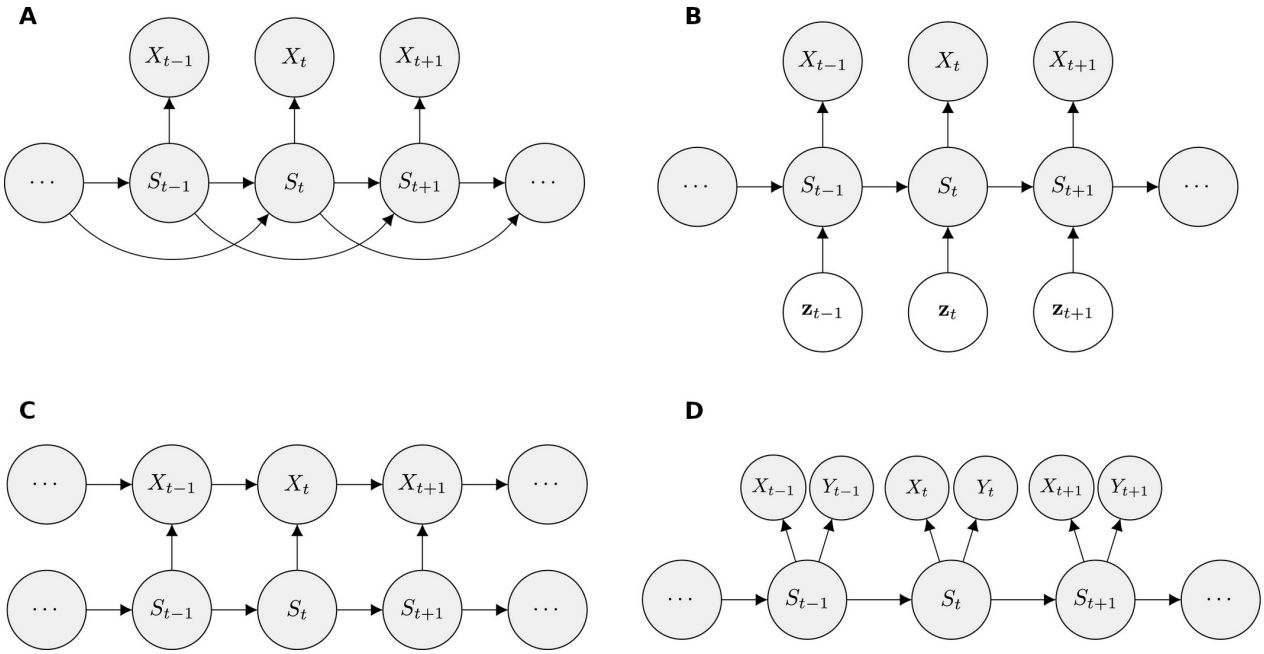

Figure 10. Graphical models associated with extensions to the basic HMM. (**A**) state sequence with a memory order of 2. (**B**) influence of covariate $z_1, \ldots, z_T$ on state dynamics. (**C**) observations depending on both states and previous observations. (**D**) bivariate observation sequence, conditionally independent given the states. Source: [72].

3.6. A State Space Model (SSM) Incorporating an Intervention Analysis

Whilst HMMs apply to scenarios where the state process is discrete-valued, we also apply an SSM allowing for interventions. Here, the state process is continuous-valued and interventions are related to times/events of aerosol exceedances or declines. Intervention analysis, initially proposed by Box and Tiao in 1975 [73], provides a structured approach for evaluating the influence of an intervention on a given time series. An intervention is considered to be an event that brings about a shift in the time series, affecting not only its average behavior, but also potentially altering characteristics such as the conditional variance and the properties of serial correlation [74].

For a time series Y_t, this general intervention model is given by

$$Y_t = m_t + N_t, \tag{7}$$

where m_t is the change in the mean function and N_t is modeled as another process, such as a seasonal or harmonic process or a trend model. The sequence N_t denotes the segment without any intervention and is known as the inherent or unaffected process. This process may exhibit features of stationarity or non-stationarity, seasonality, or lack thereof.

Assume the moment in time, T, marks the occurrence of the intervention. Prior to T, m_t is presumed to be consistently zero. The set $Y_t, t < T$ is termed as pre-intervention data and is used to define the model for the unaffected process N_t.

To delineate the impact of the intervention on the mean function the step function can be determined as:

$$S_t^{(T)} = \begin{cases} 1, & \text{if } t \geq T \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Hence the values of $S_t^{(T)}$ will be 0 during the pre-intervention period and 1 in the post-intervention period. Then we define the pulse function as follows:

$$P_t^{(T)} = S_t^{(T)} - S_{t-1}^{(T)} \quad (9)$$

The pulse function above is equal to 1 at the time point T and 0 elsewhere, indicating the point where the intervention occurred.

Implementation of the Intervention Analysis

The intervention analysis is applied in this study with the dynlm package in R [75]. The dynlm package is a powerful tool for conducting dynamic linear modeling, and is particularly well-suited for time series analysis and intervention studies. The package primarily focuses on fitting dynamic linear models while preserving the characteristics of time series data, with a single main function [76]. It provides a flexible framework for modeling relationships in data that evolve over time, allowing for lagged effects and dynamic dependencies. This package's capabilities accommodate time-dependent variables and interventions. By incorporating lagged terms, the temporal dependencies can be captured. Furthermore, the inclusion of intervention variables enables the assessment of the impact of specific events or interventions on the time series under investigation.

4. Results

This section presents the results of our analyses, using the range of techniques discussed in the Methods Section 3. The data analyzed have been described in Section 3.1. We have applied various techniques to the data across the entire range of cells of the whole area of study, and across the summed totals of ozone and aerosol over each of the three specified regions (R1, R2, R3).

4.1. RFsp and Quantile Prediction Intervals Results

For the data across the entire area, we have applied the `RFsp` method [27]. It has been trained on the data described in Section 3.1 on the ozone and aerosol data. This produces a spatiotemporal model (see Section 3.3). We present the output of the model when test data are passed into it and examine the statistical performance of the model using some common tests that are widely applied in analyses where RF models are used; specifically, we have applied k-fold cross-validation (CV), and out-of-bag (OOB), in each case yielding an R^2 and an MSE.

The RFsp methodology applied to all $(r \times n)$ data rows in the design matrix produces a model predicting ozone level given all r predictor variables. Predicted values from the `RFsp` algorithm are presented for gridded ozone in Figures 11–13, for the 20 consecutive days starting on 3 November 2019, 1 December 2019 and 20 January 2020, respectively. These predictions should be compared to the actual ozone level data shown in Figure 7 in Section 3.1. For ease of comparison, note that the figures such as Figure 7 should be compared with Figure 11. The other two regions and their appropriate dates produce similar comparisons; we refrain from showing these.

As well as predicted values, corresponding data relevant to quantile prediction intervals covering the same set of days are presented in Figures 14–16. The choice of each starting date in the above-mentioned figures, together with the 20 consecutive days thereafter, allows for effective coverage across most of the change-points in the data, as discussed below. We investigate these change-points for both total sum ozone and total sum aerosol time-series separately for each of the three regions.

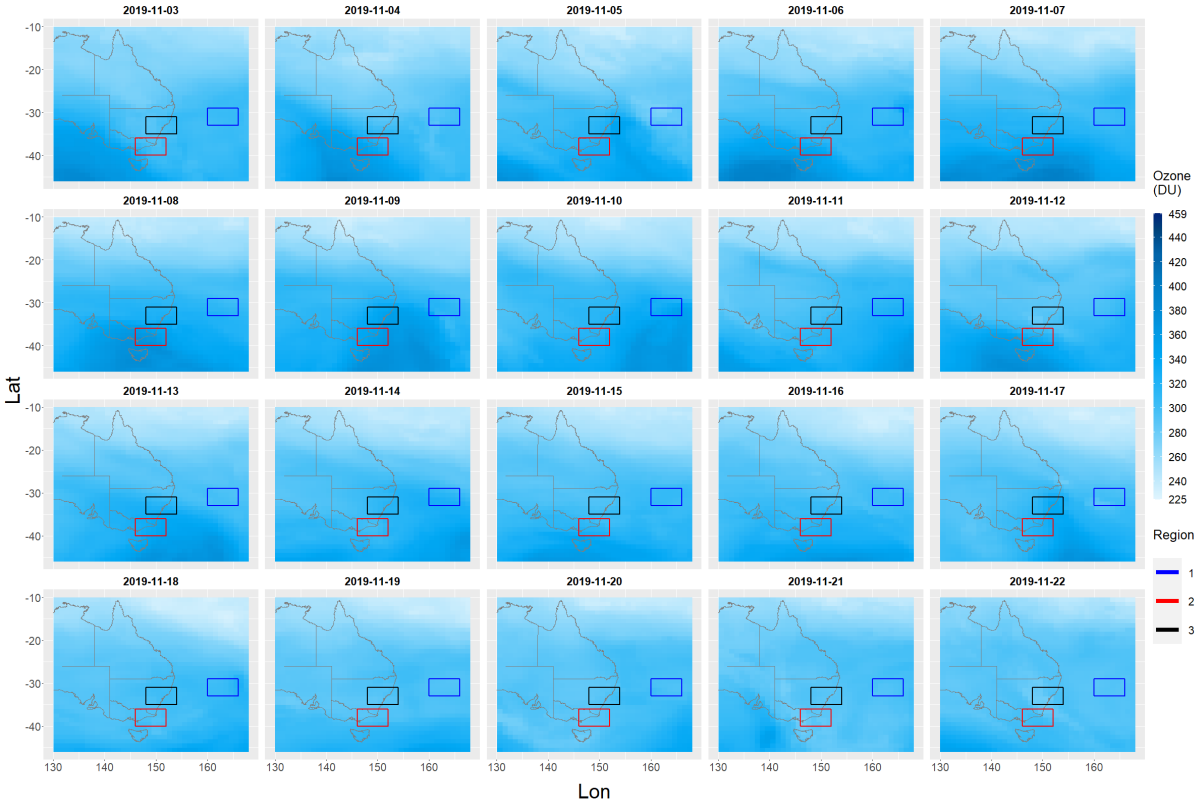

Figure 11. Spatial-temporal predictions of daily gridded ozone values over 20 days from 3 November 2019.

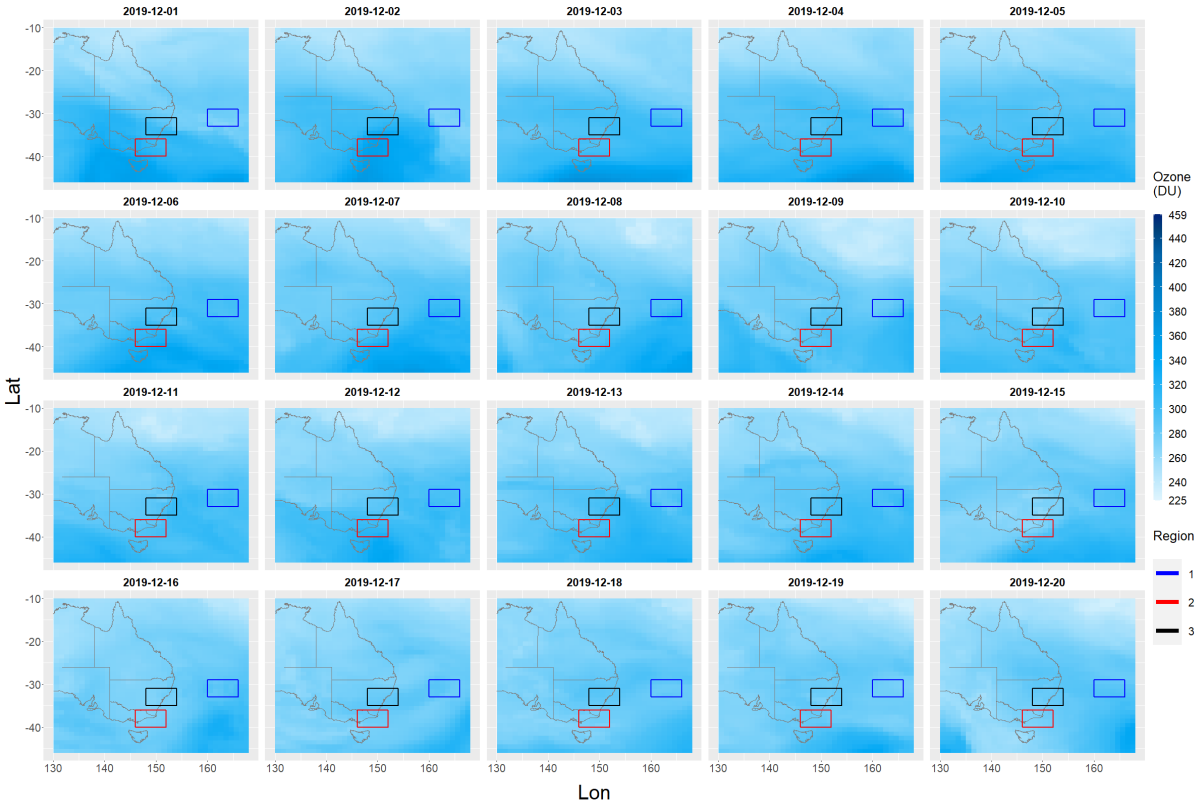

Figure 12. Spatial-temporal predictions of daily gridded ozone values over 20 days from 1 December 2019.

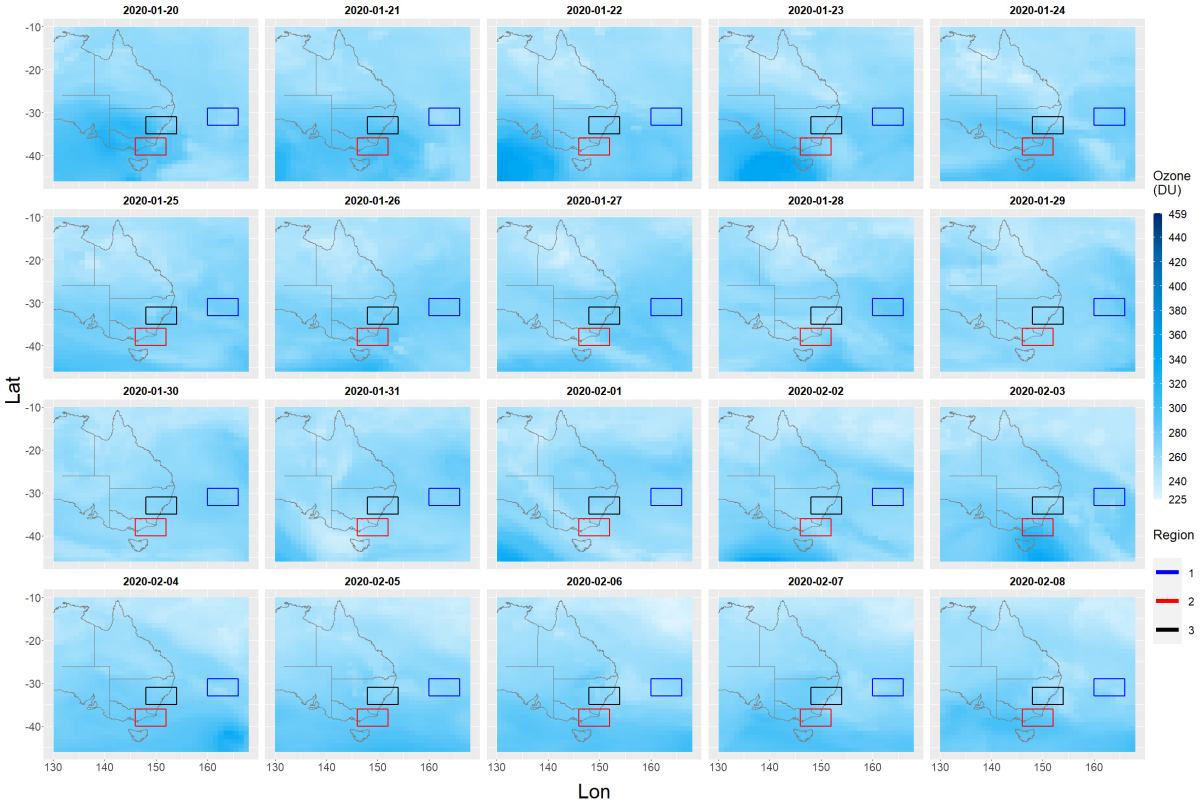

Figure 13. Spatial-temporal predictions of daily gridded ozone values over 20 days from 20 January 2020.

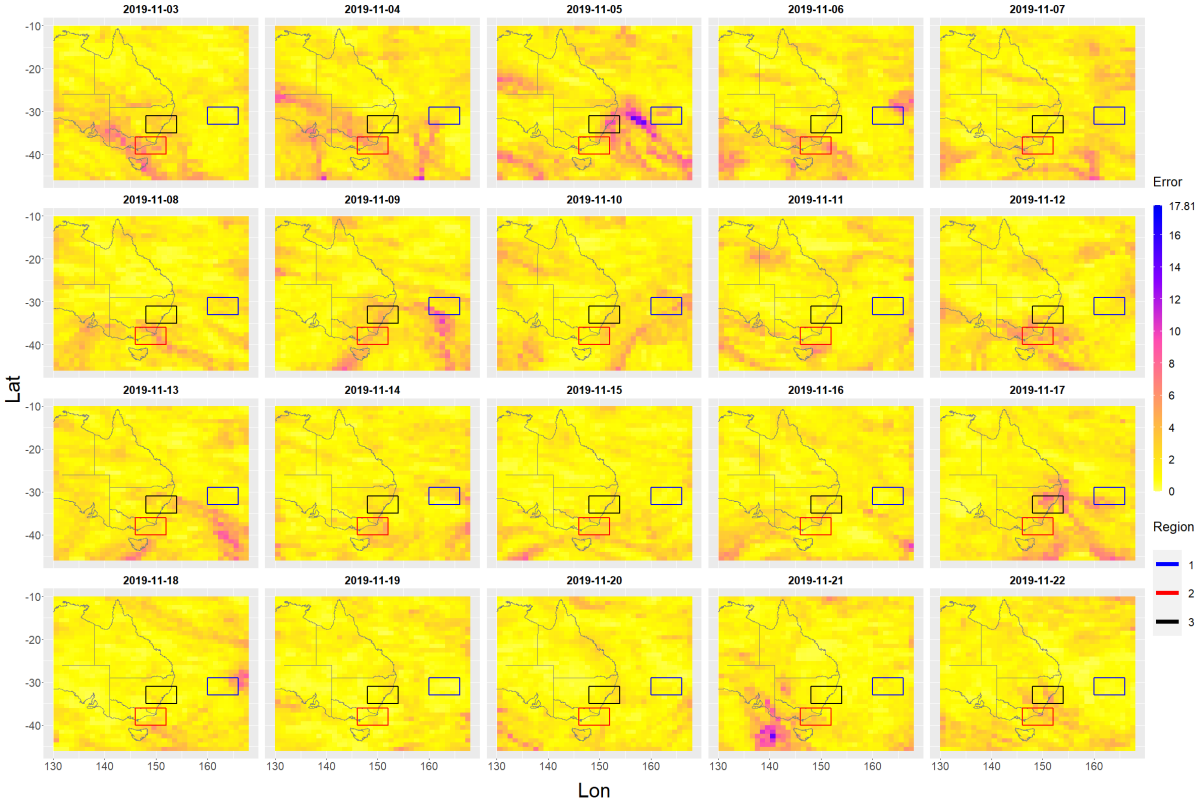

Figure 14. Quantile prediction intervals of daily gridded ozone values over 20 days from 3 November 2019.

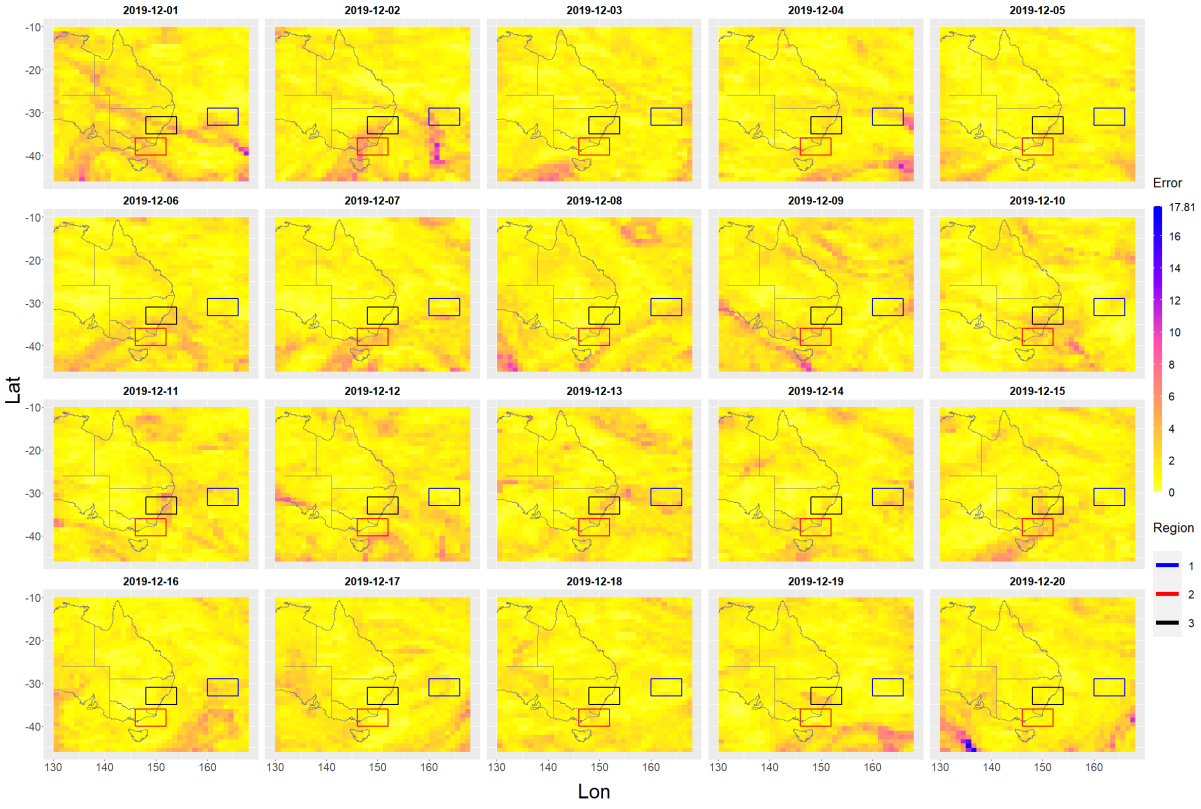

Figure 15. Quantile prediction intervals of daily gridded ozone values over 20 days from 1 December 2019.

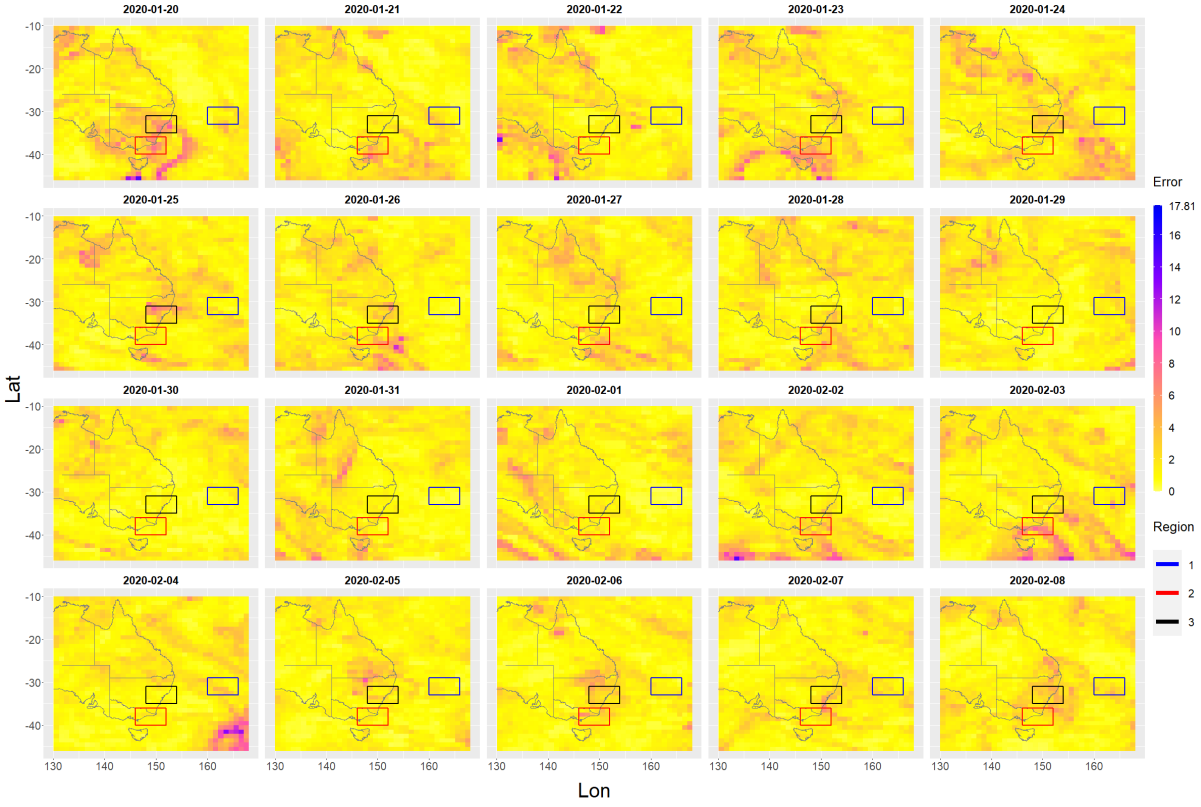

Figure 16. Quantile prediction intervals of daily gridded ozone values over 20 days from 20 January 2020.

We anticipate that these quantile prediction intervals will be wider around the change-points. The change-points in the total summed data, which are not represented in the above-mentioned figures, serve to focus our subsequent investigation into the data, specifically, in Region 1 (Ocean), Region 2 (Vic), and Region 3 (NSW).

It can be seen from Figures 11–13, that the parts of the map where the gradient of the ozone value is large in absolute terms, typically correspond to a distinctive ridge-like pattern in the quantile prediction interval plots. These ridge lines roughly follow the contours of the three-dimensional surfaces of the predicted ozone levels. These predicted ozone levels are represented by the colors in the ozone color bar (e.g., 'Ozone (DU)' legend; RHS of Figure 11). The value of the quantile prediction interval also appears to be, at least visually, closely correlated with the magnitude of these gradients of the predicted ozone value maps. In the same vein, areas where the predicted ozone values have sharp high-value peaks are associated with a higher concentration of wider quantile prediction intervals. The darker blue areas, with values ranging between 380 and 420 DU of the ozone prediction plots roughly match up with the purple areas within the range 12 to 15 on the color bar such as in Figure 14.

4.2. RFsp Error Diagnostics

We now present the outputs created after passing the design matrix into the `ranger` function in `R` and describe the results in terms of the model performance tests R^2 and MSE. The output produced by the `ranger` function is displayed in Table 1. The five-fold CV produces results that show that the `RFsp` model performs well, giving very accurate ozone predictions from aerosol measurements. The model OOB (Out Of Bag) R^2 reported in Table 1 is 0.996 and OOB mean-square-error is 4.32. The R^2 with the predictions produced during five-fold CV achieved very similar results, as did the average mean-square-error; with values of 0.995 and 5.33, respectively.

Table 1. Error diagnostics and RF hyperparameters of the `ranger` function.

Type	Regression
Number of trees	150
Sample size	248,976
Number of independent variables	2738
Mtry	52
Target node size	5
Variable importance mode	impurity
Splitrule	variance
OOB prediction error (MSE)	4.324842
R squared (OOB)	0.9960641

4.3. Quantile Prediction Intervals Examination

From Figure 14, there are 4 days that have particularly wide quantile prediction intervals occurring in and around Region 1. The days in question are the 6, 9, 17 and 18 November 2019. Similarly, binary segmentation change-points for the total sum ozone series in Region 1, were detected on 18 November 2019 which was notably 12 days after the first of these wide quantile prediction intervals, see Figure 14.

Examining Region 2 in Figure 14, wide quantile prediction intervals are seen to occur on 10 separate days: the 3rd, 4th, 6th, 7th, 8th, 9th, 12th, 14th, 17th and the 22nd of November. Similarly, in Region 3 there are 4 days of wide quantile prediction intervals, namely the 5th, 9th, 17th and 22nd of November.

In Figure 15 there were a total of 4 days, the 1, 2, 13 and 14 December with wide quantile prediction intervals inside Region 1. Analogously there were 6 days, the 1, 2, 6, 7, 12 and 15 December inside Region 2. Also, there were 3 days, the 1, 11 and 13 December inside Region 3.

In Figure 16, the 3 days with wide quantile prediction intervals are the 3rd, 4th and 8th of February. For Region 2 the 10 days: 20, 21, 23 and 26–28 January and the 2, 3, 6 and 7 February. Lastly for Region 3 the 7 days of wide quantile prediction intervals were the 20 and 24–26 January and the 6–8 February.

4.4. Binary Segmentation Results

In this section, the change-points discussed are based on binary segmentation as shown in Figures 17–19 for the three separate geographical regions. We use the acronyms TSA and TSO to abbreviate Total Sum Aerosol and Total Sum Ozone in Figures 17–19.

The binary segmentation method applied to the total sum ozone and total sum aerosol time series in Region 1 shows an interesting pattern, namely, that changes in aerosol can induce a delayed effect on ozone of between 10 and 25 days.

This can be seen by comparing the dates on which change-points occur in the total sum aerosol time series against the total sum ozone time series. The first total sum aerosol change-point occurred on 7 November 2019, 11 days before the first total sum ozone change-point occurred; specifically on 18 November 2019 (Figure 17).

This same pattern occurs at the second and third change-points of the total sum aerosol occurring on 4 December 2019 and 8 January 2020 with the perceived effect on ozone occurring 25 days later on 29 December 2019 for the second change-point and 10 days later on 18 January 2020 for the third change-point.

This lag pattern also appears in Region 2, as the first total sum aerosol change-points occur after the corresponding, second and third total sum ozone change-points in the same region. However, the second and third total sum aerosol change-points occur after the last total sum ozone change-point in Region 2 (Figure 18).

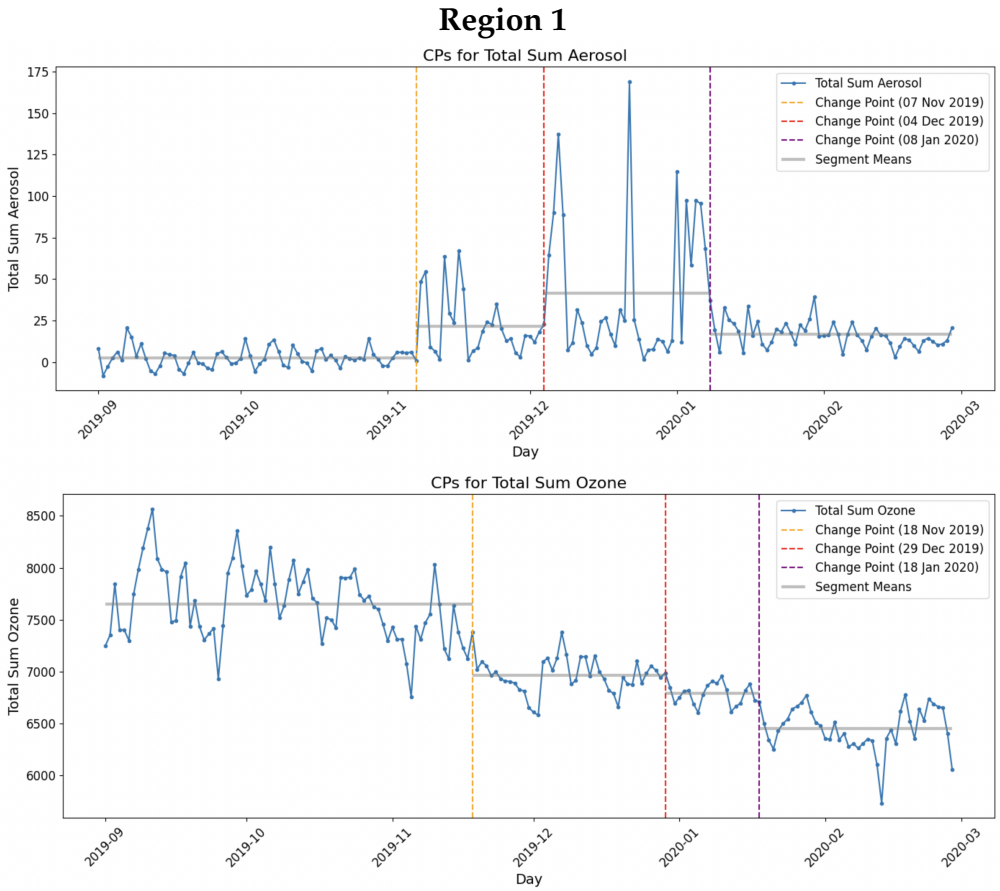

Figure 17. CPs for TSA and TSO for Region 1 (Pacific Ocean) as shown by vertical grey dashed lines.

For Region 3, a slightly different situation arises, as the binary segmentation method only detected two change-points for the total sum ozone time series. The first occurring on 18 November 2019 and the second on the 2 January 2020 (Figure 19).

The first total sum aerosol change-point occurring on 5 November 2019, has a similar lag pattern to those identified in Region 1, since the first total sum ozone change-point occurred 13 days later, on 18 November 2019. Overall, there is a lag between change-points in total sum aerosol levels and corresponding change-points in total sum ozone levels of 10 to 25 days.

It is difficult, on the basis of this change-point analysis alone, to definitively detect a causal relationship between changes in aerosol levels and changes in ozone levels, though there is some evidence of a pattern. An alternative approach is applied in this study to statistically establish times at which the ozone states change, as determined by HMM models, with adjustment of the aerosol levels as a covariate (as discussed in Section 4.5).

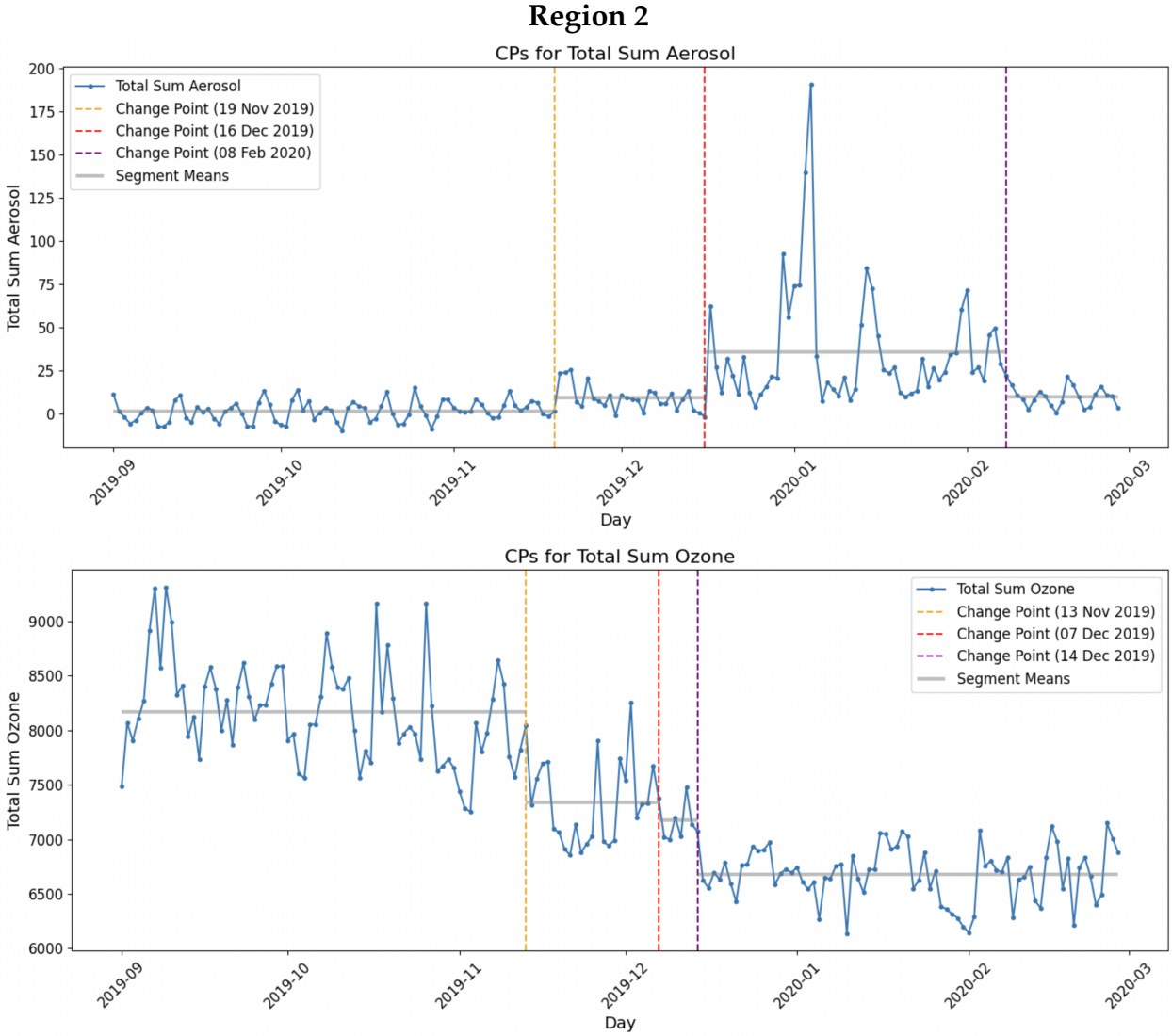

Figure 18. CPs for TSA and TSO for Region 2 (Vic) as shown by vertical grey dashed lines.

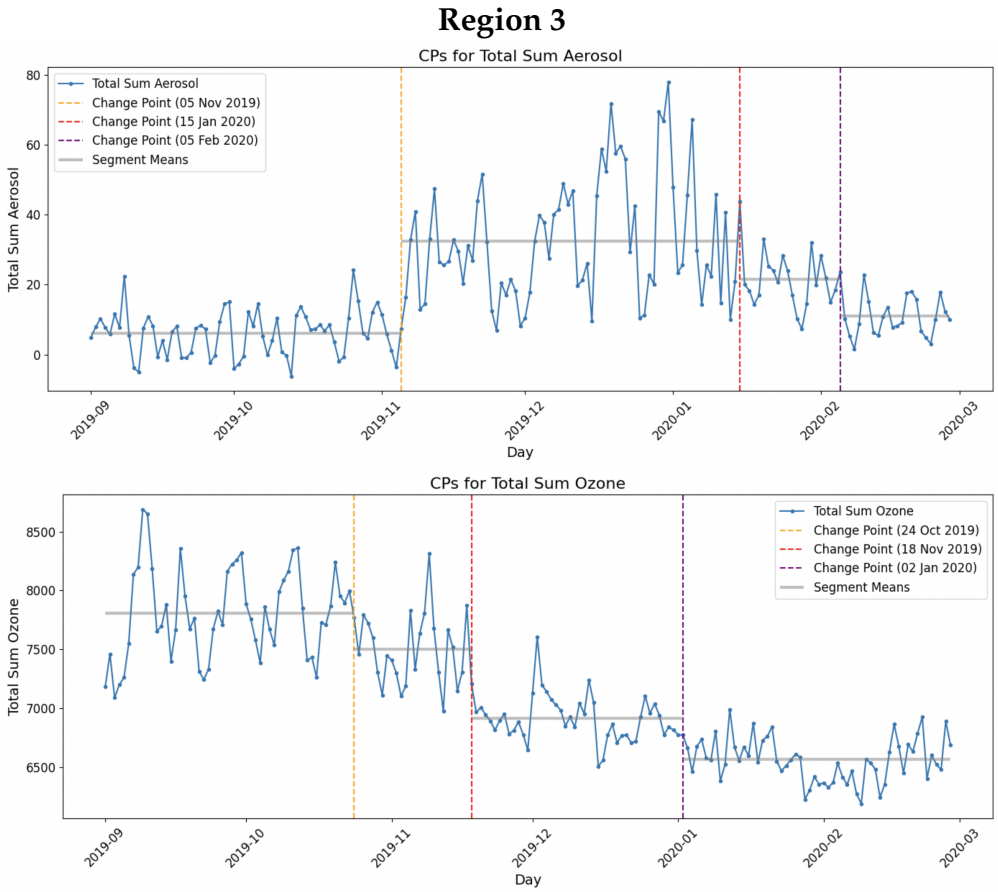

Figure 19. CPs for TSA and TSO for Region 3 (NSW) as shown by vertical grey dashed lines.

4.5. HMM Analysis Results

In this section, the results of the HMM analysis for Region 3 are presented. The HMM analyses on the total sum ozone time-series in each region are performed for two different cases *with* aerosol as a covariate in the HMM. Outputs of AIC and BIC in testing for 2, 3, 4 and five state HMM model candidates are given, with the selection of the optimal model based on the minimum AIC and BIC achieved. Also presented is the transmission matrix of the transition probabilities, and the mean and standard deviation estimates for each state means for both ozone and aerosol where an adjustment is made in the HMM for aerosol as a covariate. This section reports the results of the intervention analysis for Region 3 (NSW). The corresponding results for the optimal HMM models for Region 1 (Pacific Ocean) are given in Appendix B.1 and for Region 2 (Vic) shown in Appendix B.2.

HMM Analysis of Ozone for Region 3 with Adjustment by Aerosol as a Covariate

Tables 2–4, and Figure 20 report the HMM models when aerosol levels are added as a covariate (the so-called bivariate model) to better explain the ozone changes observed in Region 3.

Table 2. AIC and BIC criteria for HMM model selection with differing number of states for total sum ozone with total sum aerosol as a covariate in Region 3.

Model Name	AIC	BIC
HMM (with Aerosol)—with five states	3957.754	4098.487
HMM (with Aerosol)—with four States	4013.133	4112.286
HMM (with Aerosol)—with three States	4020.225	4084.195
HMM (with Aerosol)—with two States	4088.763	4123.947

Table 3. Transmission matrix of all five states from HMM for total sum ozone with total sum aerosol as a covariate in Region 3.

Transmission Matrix	State 1	State 2	State 3	State 4	State 5
State 1	0.979	0	0	0.021	0
State 2	0.043	0.957	0	0	0
State 3	0	0	0.985	0.015	0
State 4	0	0	0	0.364	0.636
State 5	0	0.133	0	0.068	0.8

Table 4. Parameter estimates of all five states from HMM for total sum ozone and total sum aerosol Region 3.

	Ozone (DU)		Aerosol	
	Mean	Std. Deviation	Mean	Std. Deviation
State 1	6563.193	197.722	15.293	7.648
State 2	6802.999	181.552	36.726	18.652
State 3	7738.528	386.539	6.113	6.335
State 4	7624.158	407.595	14.421	2.57
State 5	7327.054	295.041	32.355	6.97

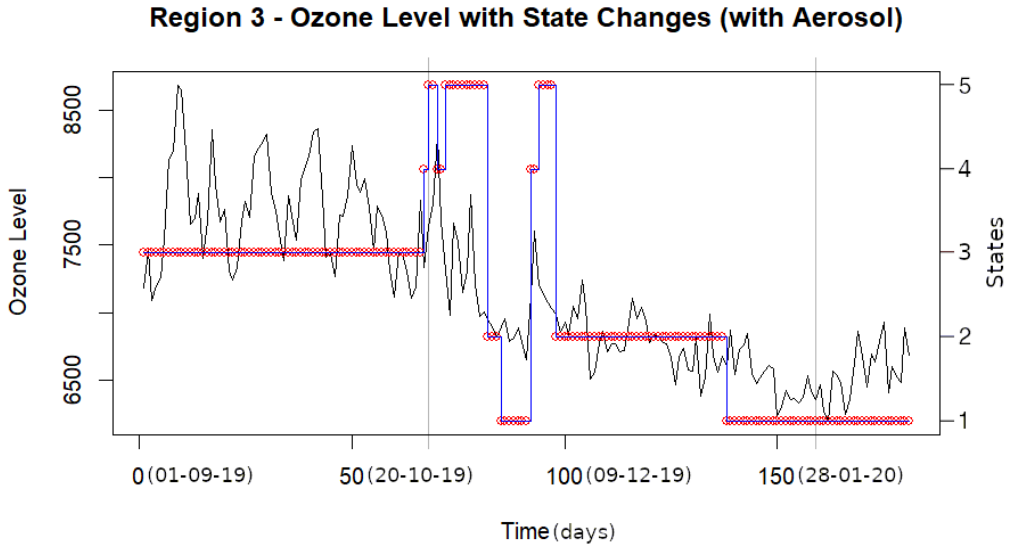

Figure 20. HMM bivariate state changes in total sum ozone time-series with respect to total sum aerosol time-series as a covariate for Region 3.

According to the intercept of the parameters of the HMM model, the intercept value of ozone of State 3 is the largest with an average mean ozone level of 7738.53 DU and lowest aerosol = 6.11, and State 1 has the smallest average state-specific mean ozone level of 6563.19 DU aligned with an increased aerosol of 15.29 (Table 4).

From the beginning of the series to day 66 (5 November 2019) the stochastic process is identified as being in State 3, with the highest mean ozone intercept parameter = 7738.53 DU, and lowest mean aerosol = 6.11. Then, after day 66 (5 November 2019) to day 82 (21 November 2019) ozone transitions between States 4 and 5, with State 5 (mean ozone intercept of 7327.05 DU, aerosol level = 32.36).

Between day 82 (21 November 2019) and day 91 (30 November 2019), ozone levels change between States 1 and 2. After day 91 (30 November 2019) to day 97 (6 December 2019),

ozone levels then increase and the HMM system starts varying between States 4 and 5. After day 97 (6 December 19) to day 137 (15 January 2020), ozone levels then decrease again and the HMM system remains in State 2, with State 2 (mean of 6803.0 DU, aerosol level = 36.73). From day 137 (15 January 2020) to the end of the time series the mean ozone level then decreases again to the value in State 1 (mean of ozone level = 6563.19 DU, aerosol level = 15.29).

The first observation over the study period of bushfire-induced aerosols occurs on the 68th day (7 November 2019). The first point of change in ozone levels, as demonstrated by HMM, occurred approximately 1 day earlier at 67 days (6 November 2019). The last relatively small emission of bushfire-induced aerosols occurs on the 159th day (6 February 2020), compared with the fifth state change-point, which occurs on day 137 (15 January 2020), 22 days earlier in Region 3 (see Figure 21).

	BinSeg change-points for Ozone (O) and Aerosol (A)					
	Region 1 Ocean		Region 2 Vic		Region 3 NSW	
Date	O	A	O	A	O	A
27 Oct 2019						
05 Nov 2019						66
07 Nov 2019		68				
09 Nov 2019						
13 Nov 2019			74			
17 Nov 2019						
18 Nov 2019	79				79	
19 Nov 2019				80		
21 Nov 2019						
30 Nov 2019						
04 Dec 2019		95				
06 Dec 2019						
07 Dec 2019			98			
12 Dec 2019						
14 Dec 2019			105			
16 Dec 2019				107		
29 Dec 2019	120					
02 Jan 2020					124	
08 Jan 2020		130				
15 Jan 2020						137
18 Jan 2020	140					
19 Jan 2020						
05 Feb 2020						158
08 Feb 2020				161		

	HMM change-points for Ozone (O) and Ozone (A) (see caption)					
	Region 1 Ocean		Region 2 Vic		Region 3 NSW	
Date	O	A	O	A	O	A
27 Oct 2019			57			
05 Nov 2019						66
07 Nov 2019		68				
09 Nov 2019				70		
13 Nov 2019						
17 Nov 2019			78			
18 Nov 2019						
19 Nov 2019	80	80				
21 Nov 2019						82
30 Nov 2019						91
04 Dec 2019						
06 Dec 2019						97
07 Dec 2019						
12 Dec 2019			103			
14 Dec 2019					105	
16 Dec 2019				107		
29 Dec 2019						
02 Jan 2020						
08 Jan 2020						
15 Jan 2020						137
18 Jan 2020		140				
19 Jan 2020	141					
05 Feb 2020						
08 Feb 2020						

Figure 21. Summary of the BinSeg CPs and the HMM derived CPs. In the RHS table, the O columns show the ozone CP (day) based on HMM without covariate adjustment. The A columns show the ozone CP (day) for HMMs with aerosol as a covariate.

4.6. HMM and Binary Segmentation Combined—Comparability Summary

The corresponding results for the optimal HMM models for Region 1 (Pacific Ocean) are given in Appendix B.1 and for Region 2 (Vic) in Appendix B.2. Note that in Figure 21 the O and A columns have a different meaning to the table reporting the Binary Segmentation change-pointsin Section 4.4. Here, on the RHS of Figure 21, the O columns give the day on which ozone states change according to HMM models which do not include aerosol as a

covariate. The A columns report the day on which ozone states change for HMM analyses *with* aerosol added as a covariate, the so-called covariate-adjusted HMM change-point. On the LHS of Figure 21 we report the BinSeg change-points across the three regions. For visual reference of the total sum aerosol and total sum ozone time-series across the three regions, see Figure 9 and Figure 8, respectively.

From the combined HMM and Binary Segmentation change-points (CPs) summarized in Figure 21 we observe the following.

In Region 1 (Pacific Ocean) the exceedance (CP) in Aerosol occurred around 7 November 2019. The associated subsequent decrease in Ozone then occurred around 12 days later on the 19 November 2019. About 40 days later on the 18 January 2020 Ozone levels were still significantly decreasing.

In Region 2 (VIC) the exceedance (CP) in Aerosol occurred around 13–17 November 2019. Then 18 days thereafter around 7–12 December 2019, there was a significant decrease in ozone; about 25 days later, Ozone levels were still significantly decreasing.

In Region 3 (Pacific Ocean) the exceedance (CP) in Aerosol occurred around 5 November 2019, followed 13–16 days later by a decrease in ozone around the 30 November 2019. About 40–45 days later, ozone levels were still significantly decreasing around 15 January 2020.

4.7. Intervention Analysis Results for Region 3 (NSW)

This section reports the results of the intervention analysis for Region 3 (NSW). The corresponding results for the optimal intervention models for Region 1 (Pacific Ocean) and Region 2 (Vic) are given in Appendices A.1 and A.2.

In Region 3, we note the HMM analysis based on five states showed early oscillatory changes in states in ozone levels. This motivated our investigation as to whether an alternative BinSeg routine could detect earlier change-points. This was achieved by using the BinSeg method in the Package Rapture—[77]. These change-points (day 54, 79 and 124) were used for the intervention analysis of Total Sum Ozone for Region 3 (NSW), as shown in Figure 22.

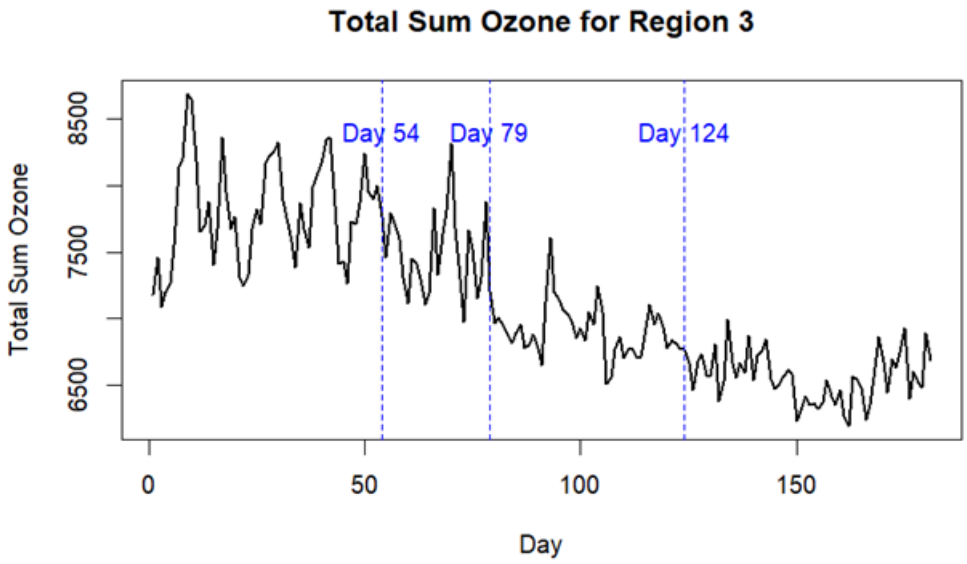

Figure 22. Detected Change Points for Total Sum Ozone (Region 3).

In the first step of the intervention analysis, all three change-points were included (denoted by S, P and Q) as different step functions to indicate the intervention points as the baseline. For each intervention point, their first two lags (the first lag is denoted by S.t.1, P.t.1 and Q.t.1 and the second lag is denoted by S.t.2, P.t.2, and Q.t.2) were also included. In addition, the intercept was included, along with one lag of the actual observed time

series (Total Sum Ozone) (denoted by L(Y.t, k = 1) as well as its trend component (denoted by trend (Y.t).

Model involved the addition of a third lag component for aerosol L (Aerosol_Region3, 3) (Table 5); all the following components involving mean ozone level (intercept) and ozone's lagged dependence of 1 and 5 days were found to be highly significant ($p < 0.0001$), as were the negative trends for the first and second intervention points at 54 and 79 days, respectively. Likewise, the sustained negative lag1 effect at the third change-point at 124 days was significant ($p < 0.0001$). Notably, in Region 3 (NSW) aerosol levels lagging at 1 day had a significant impact on the observed ozone levels over time ($p < 0.04$) and the negative impact of lag3 aerosol was significant but only at $p = 0.10$ (Table 5).

Table 5. Model Results.

Variable	Estimate	Std. Error	t Value	Pr(>|t|)
(Intercept)	5462.56	633.00	8.63	<0.0001 ***
L(Y.t, k = 1)	0.50	0.05	8.45	<0.0001 ***
L(Y.t, k = 5)	−0.20	0.05	−3.71	0.000282 ***
S.t	−234.58	59.74	−3.92	0.000126 ***
P.t	−431.34	76.54	−5.63	<0.0001 ***
Q.t.1	−232.57	57.86	−4.02	<0.0001 ***
L(Aerosol_Region3, 1)	2.79	1.30	2.13	0.033857 *
L(Aerosol_Region3, 3)	−2.41	1.26	−1.90	0.05

* $p < 0.05$, *** $p < 0.001$.

It can be seen that the resultant fitted model (Figure 23) was able to capture the behavior of the series quite successfully with small residuals. The results and outputs for the final intervention models for regions 1 and 2 can be found in Appendices A.1 and A.2.

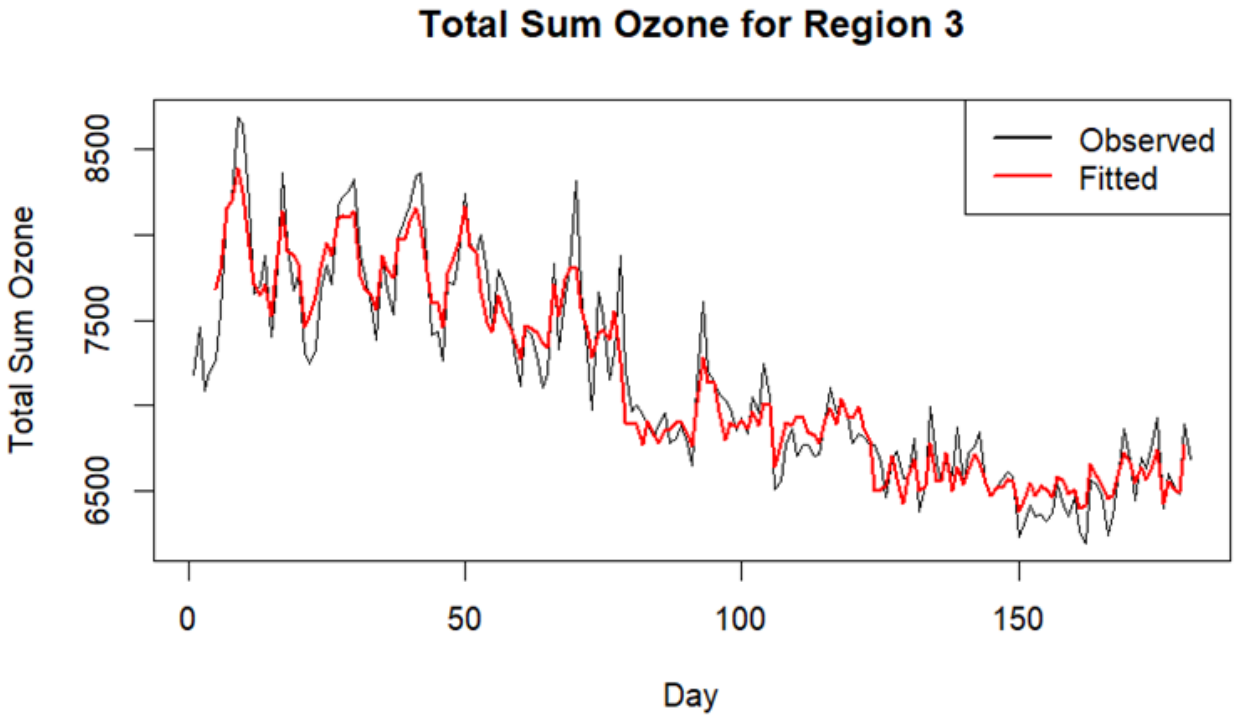

Figure 23. Intervention Analysis Output for Region 3 with Model R3.

5. Conclusions

`RFsp`-based quantile regression predictions, visualized by spatial-temporal maps of the ozone levels, across Australia and the nearby ocean, confirmed our hypothesis that

bushfires result in a localized effect on the ozone layer in the vicinity of the bushfire or downwind of it. Noteworthy also, is that our findings, along with those from [9,10] which discuss how aerosols affect ground ozone levels, in addition to the findings of [15,19,20] on stratospheric ozone depletion, indicate that there is significant evidence that ozone levels are affected by aerosols from bushfires. Random forest spatial-temporal (`RFsp`) analysis of the satellite data from [34] of ozone and aerosols (NASA/NOAA) has shown very accurate ozone predictions from aerosol measurements (achieving an OOB R^2 of 0.996 and MSE of 4.32, with a five-fold cross-validation R^2 of 0.995 and MSE of 5.33), indicating a significant association between bushfire incidence in Australia and ozone depletion in the vicinity of the bushfires.

The use of the total column/sum ozone (TSO) and total sum aerosol (TSA) data for three regions of eastern Australia (Pacific Ocean, Victoria, and NSW) demonstrated significant change-points (or times of change in states) of ozone with respect to aerosol exceedances, as established by binary segmentation and also by Hidden Markov models (HMMs). The HMM approach established that the ozone TSO time series exhibited four or five hidden states. We also observed a significant and quantifiable temporal and lagged effect between times of significant change in aerosol of 10–25 days and change-points in ozone levels in all three regions.

Intervention methods gave similar times of aerosol exceedance to those found by HMM. Intervention analysis increases in aerosol in all three regions had a significant and ongoing impact 1–5 days later in reducing ozone levels ($p < 0.0001$).

For all three regions, there was a lag of 10–25 days between times of aerosol exceedances (or change-points) and subsequent ozone depletion. Specifically, in Region 1 (Pacific Ocean) the exceedance (CP) in Aerosol occurred around 7 November 2019. The associated subsequent decrease in ozone then occurred around 12 days later on 19 November 2019. About 40 days later on 18 January 2020 ozone levels were starting to increase.

In Region 2 (Vic) the exceedance (CP) in Aerosol occurred around 13–17 November 2019. Then, 18 days later around 7–12 December 2019, there was a significant decrease in ozone; 25 days later ozone levels began to increase.

In Region 3 (Pacific Ocean) the exceedance (CP) in Aerosol occurred around 5 November 2019; followed 13–16 days later by a decrease in Ozone around the 30 November 2019. About 40–45 days later, ozone levels begin increasing somewhat around 15 January 2020.

Further work based on HMMs and SSM intervention analyses will incorporate meteorological and climate variables, such as wind and temperature, seasonal effects, and possibly random effects, along with forecasting future ozone levels. More mathematically rigorous Binary Segmentation methods [58], are topics of future work; as are HMM-Gamma [69] models, recently proposed for ozone zone identification models.

Possible overfitting by the `RFsp` model will be addressed by further optimization of the tuning parameters, which govern the number of features randomly chosen for growing each tree from the bootstrapped data. Fixed Rank Kriging [49] and Bayesian Spatial Multivariate Trees (SPAMTrees) [47] are also recent alternatives to test and contrast with `RFsp` in the near future. Further potential approaches to be future work are in regard to Spatio-Temporal Functional Neural Networks following the machine learning (ML) approach of Rao et al. [78].

Author Contributions: Conceptualization, I.H. and P.P.-S.; methodology, I.H. and P.P.-S.; software, P.P.-S. and S.K.; validation, P.P.-S., I.H. and S.K.; formal analysis, P.P.-S., I.H. and S.K.; investigation, P.P.-S. and I.H.; resources, I.H.; data curation, P.P.-S.; writing—original draft preparation, P.P.-S. and I.H.; writing—review and editing, I.H. and S.K.; visualization, P.P.-S. and S.K.; supervision, I.H.; project administration, I.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Daily ozone and aerosol measurements observed by the ozone mapping and profiler suite (OMPS) aboard the Suomi National Polar-orbiting Partnership (Suomi NPP) satellite [34] can be retrieved using the OPeNDAP framework. `OPeNDAP` stands for *Open-source Project for a Network Data Access Protocol*. It provides a way to share data more easily across the world. Documentation for how to use `OPeNDAP` can be accessed via the following link (https://www.opendap.org/support/user-documentation, accessed on 24 June 2022).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Intervention Analysis Results

Appendix A.1. Region 1 (Pacific Ocean) Results

According to the BinSeg segmentation and HMM analysis results in Sections 4.4 and 4.5, respectively, three different change-points were detected for the Total Ozone Column for Region 1, day 59, 79 and 139 (Figure A1).

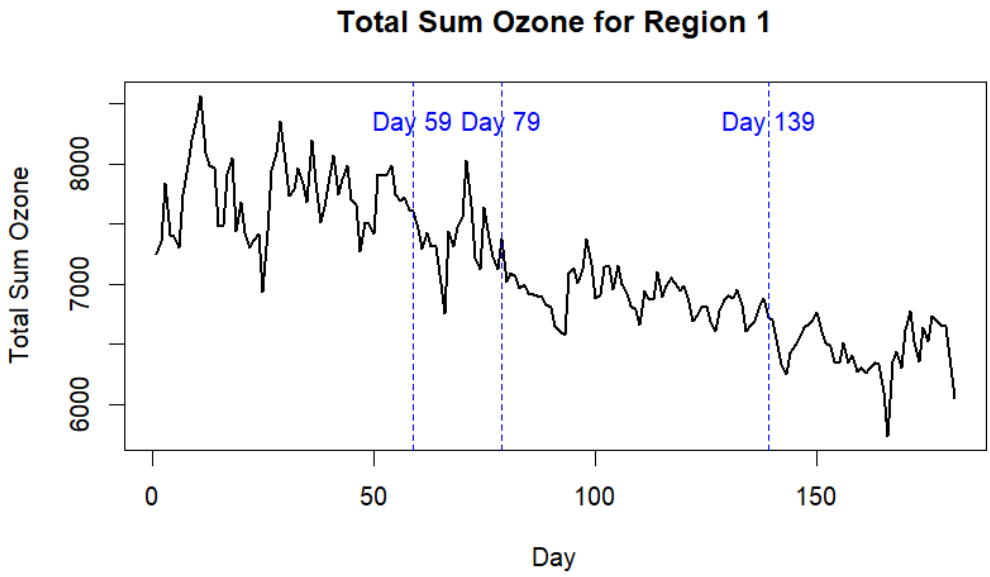

Figure A1. Detected Change Points for Total Sum Ozone (Region 1).

Based on the highest Adjusted R-squared value and lowest MASE and AIC scores (Adjust R^2 = 0.88, AIC = 2356.18, MASE = 0.85), Model R1 is selected as the best model (Table A1). It can be seen that the fitted model (Figure A2) was able to capture the behavior of the series quite successfully with small residuals. In summary, in this model all the following components involving mean ozone level (intercept) and ozone's lagged dependence of 1 and 5 days were highly significant, as were the negative trend lag1 effects at the first and second intervention points at 59 and 79 days, respectively; likewise, the sustained non-lagged negative effect at the third change-point at 139 days. Aerosol levels lagging at 1 day were not shown to be significant.

Table A1. Model Results (Region 1).

Variable	Estimate	Std. Error	t Value	Pr (>|t|)
(Intercept)	4542.37	613.19	7.41	<0.0001 ***
L(Y.t, k = 1)	0.58	0.06	9.80	<0.0001 ***
L(Y.t, k = 5)	−0.17	0.06	−2.97	0.0034 **
S.t.1	−245.98	57.83	−4.25	<0.0001 ***
P.t.1	−281.53	61.86	−4.55	<0.0001 ***
Q.t	−256.53	53.78	−4.77	<0.0001 ***
L(Aerosol_Region1, 1)	0.90	0.64	1.40	0.16

** $p < 0.01$, *** $p < 0.001$.

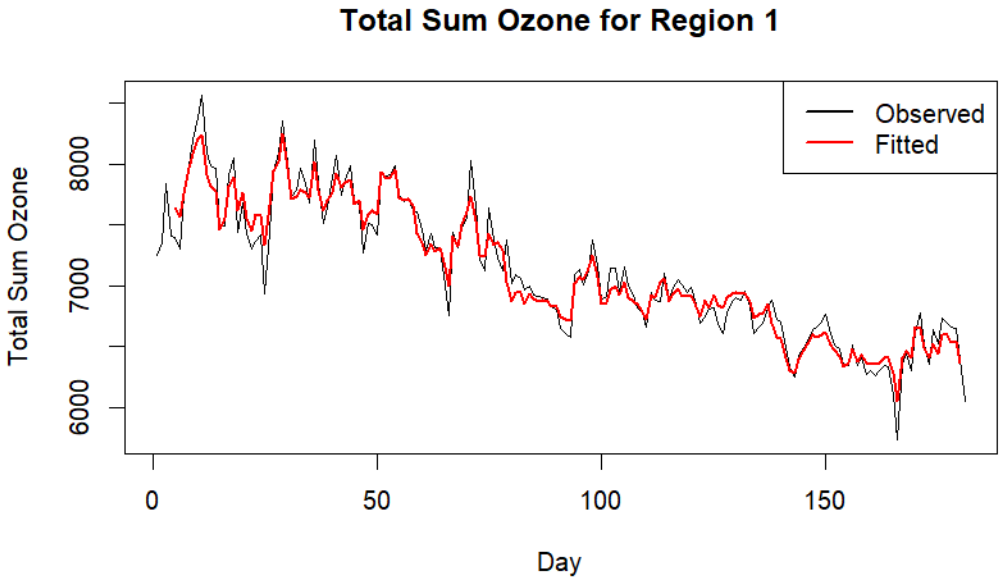

Figure A2. Intervention Analysis Output for Region 1.

Appendix A.2. Region 2 (Vic) Results

According to the BinSeg segmentation and HMM analysis results in Sections 4.4 and 4.5, respectively, three different change-points were detected for the Total Ozone Column for Region 2, days 49, 74 and 105 (Figure A3). It can be seen that the fitted model (Figure A2) was able to capture the behavior of the series quite successfully with small residuals. In summary, in this model all the following components involving mean ozone level (intercept) and ozone's lagged dependence of 1 day were highly significant. The negative trend lag1 effects at the first and second intervention points at 49 and 74 days, were significant but the negative effect at the third change-point at 105 days was not significant. Aerosol levels lagging by 1 day were also shown to be insignificant.

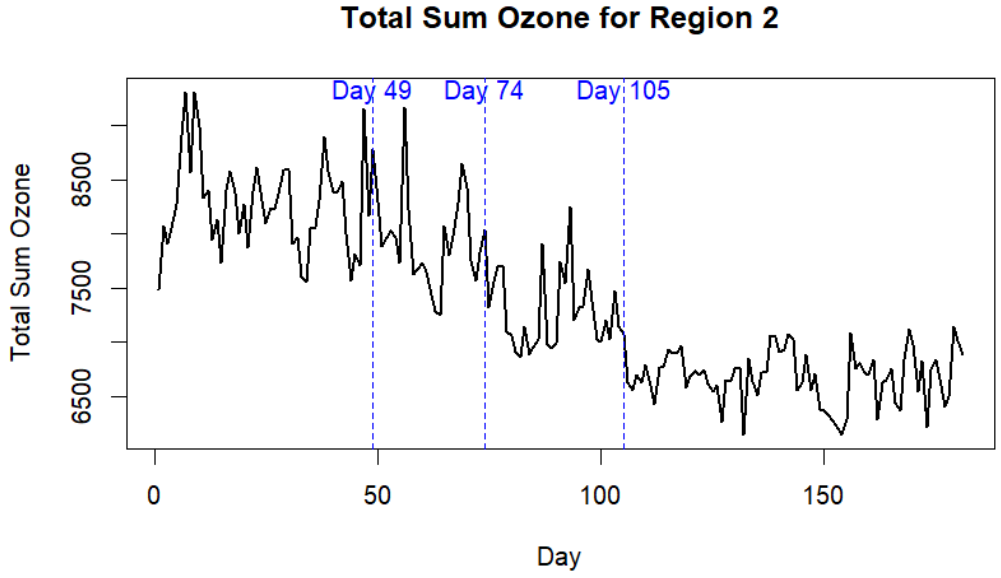

Figure A3. Detected Change Points for Total Sum Ozone (Region 2).

Based on the highest Adjusted R-squared = 0.82, and lowest MASE = 0.83 and AIC = 2588.96, Model R2 was optimal, with the resultant coefficiencies (Table A2) and the fitted model (Figure A4) which was able to capture the general behavior of the series. In the optimal model, the following components involving mean ozone level (intercept) and ozone's lagged dependence of 1 day were highly significant, and the negative trends

for the lag1 effect at the first and second intervention points at 49 and 74 days, respectively, were significant. Aerosol levels or lags were not significant as covariates in this intervention analytic model.

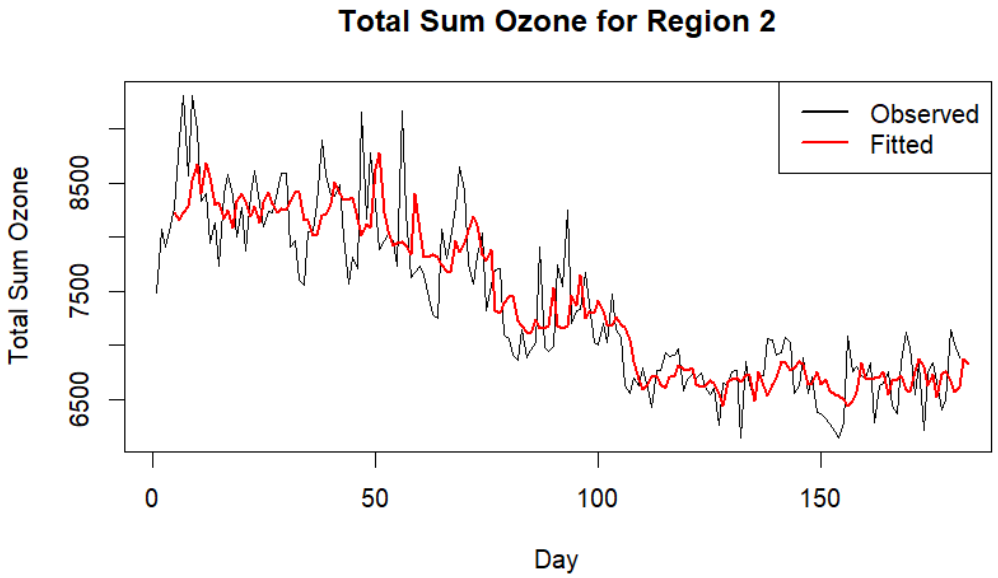

Figure A4. Intervention Analysis Output for Region 2

Table A2. Model Results (Region 2).

Variable	Estimate	Std. Error	t Value	Pr (>|t|)
(Intercept)	5179.19	582.43	8.89	<0.0001 ***
L(Y.t, k = 1)	0.38	0.07	5.37	<0.0001 ***
S.t	539.39	324.21	1.66	0.09
S.t.1	−773.75	327.30	−2.36	0.01 *
P.t.1	−648.94	326.91	−1.98	0.04 *
P.t.2	258.78	330.46	0.78	0.43
Q.t	−184.70	234.72	−0.79	0.43
Q.t.2	−182.28	235.20	−0.78	0.43
Aerosol_Region2	−1.05	1.45	−0.72	0.47
L(Aerosol_Region2, 1)	−0.51	1.73	−0.29	0.77
L(Aerosol_Region2, 2)	0.86	1.44	0.60	0.55

* $p < 0.05$, *** $p < 0.001$.

Appendix B. HMM Analysis Results for Region 1 and Region 2

Appendix B.1. HMM Analysis of Ozone for Region 1 with Adjustment by Aerosol as a Covariate

Tables A3–A5, and Figure A5 depict the HMM models when aerosol levels are added as a covariate (the so-called bivariate model).

Table A3. AIC and BIC criteria for HMM model selection with differing number of states for total sum ozone with total sum aerosol as a covariate in Region 1.

Model Name	AIC	BIC
HMM (with Aerosol)—with four States	3912.914	4012.067
HMM (with Aerosol)—with five States	3928.469	4069.203
HMM (with Aerosol)—with three States	3977.753	4041.723
HMM (with Aerosol)—with two States	4162.208	4197.392

Table A4. Transmission matrix of all four states from HMM for total sum ozone with total sum aerosol as a covariate in Region 1.

Transmission Matrix	State 1	State 2	State 3	State 4
State 1	0.903	0.021	0	0.076
State 2	0	1	0	0
State 3	0	0	0.972	0.028
State 4	0.222	0	0.049	0.729

Table A5. Parameter estimates of all four states from HMM for total sum ozone and total sum aerosol Region 1.

	Ozone (DU)		Aerosol	
	Mean	Std. Deviation	Mean	Std. Deviation
State 1	6895.048	152.059	16.482	9.075
State 2	6450.222	208.331	15.354	6.577
State 3	7676.676	342.709	2.645	5.632
State 4	7086.395	294.442	70.048	40.184

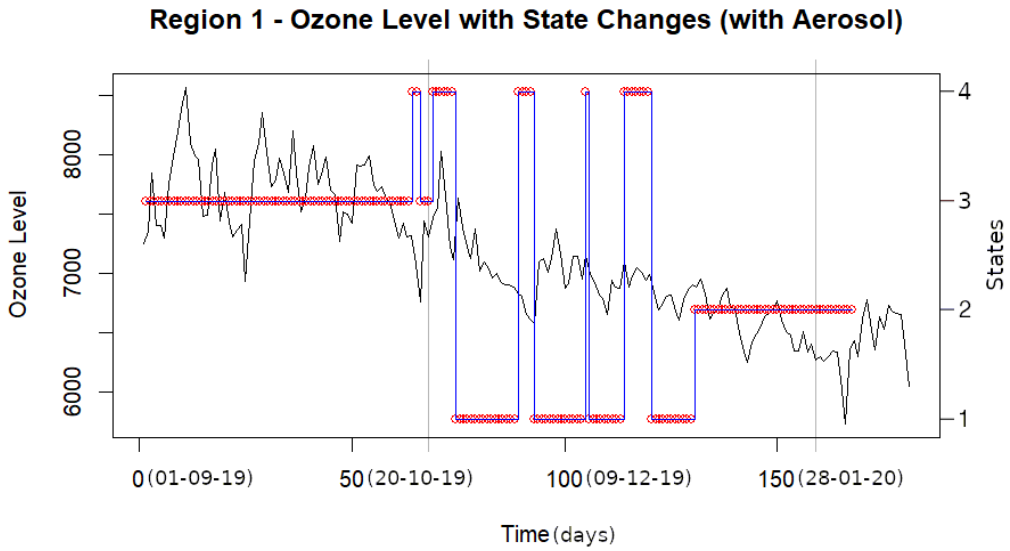

Figure A5. HMM bivariate state changes in total sum ozone time-series with respect to total sum aerosol time-series as a covariate for Region 1.

With reference to Figure A5, the beginning of the ozone series is identified as being in State 3 (highest mean ozone intercept parameter = 7676.67 DU, and lowest mean aerosol = 2.645) until day 68 (7 November 2019), which corresponds to the day of the change-point for the aerosol data (Figure 17). This is also the same day as the first change-point detected using Binary Segmentation.

Between day 68 (7 November 2019) and 80 (19 November 2019), ozone levels vary between States 3 and 4 (Figure A5) with State 4 having the highest aerosol intercept mean of 70.5 (State 4: ozone(DU) = 7086.40).

From day 80 (19 November 2019), ozone levels then alternate between State 1 (mean ozone, aerosol = (6895.04, 16.48)) and State 4 (mean ozone, aerosol) = (7086.40, 70.05)) where they remain until day 140 (18 January 2020), which is 10 days after the last change-point detected for aerosol by Binary Segmentation.

From the transmission probabilities in Table A4 the ozone level in this period primarily remains in State 1 with a probability of 0.903. Note that during this period State 1 can

transition to State 4 with a probability of 0.076, and State 4 can transition to State 1 with a probability of 0.22. After day 141 (19 January 2020) ozone levels were identified as State 2 (mean ozone, aerosol) = (6450.22, 15.35).

Appendix B.2. HMM Analysis of Ozone for Region 2 with Adjustment by Aerosol as a Covariate

Tables A6–A8, and Figure A6 depict the HMM models when aerosol levels are added as a covariate for Region 2.

Table A6. AIC and BIC criteria for HMM model selection with differing number of states for total sum ozone with total sum aerosol as a covariate in Region 2.

Model Name	AIC	BIC
HMM (with Aerosol)—with four States	4003.818	4102.971
HMM (with Aerosol)—with five states	4011.968	4152.702
HMM (with Aerosol)—with three States	4049.551	4113.521
HMM (with Aerosol)—with two States	4146.869	4182.053

Table A7. Transmission matrix of all four states from HMM method for total sum ozone with total sum aerosol as a covariate in Region 2.

Transmission Matrix	State 1	State 2	State 3	State 4
State 1	0.857	0	0.091	0.052
State 2	0	0.742	0	0.258
State 3	0.061	0	0.939	0
State 4	0.018	0.085	0	0.897

Table A8. Parameter estimates of all four states from HMM method for total sum ozone and total sum aerosol Region 2.

	Ozone (DU)		Aerosol	
	Mean	Std. Deviation	Mean	Std. Deviation
State 1	7475.178	340.042	6.038	4.328
State 2	6580.937	232.147	64.789	39.997
State 3	8295.045	409.157	0.04	5.805
State 4	6742.159	240.538	14.767	8.762

Figure A6 displays the state change plot of the optimal HMM bivariate model with four states with minimum AIC and BIC achieved in testing two, three, four and five state HMM models (Table A6) for Region 2. According to the transmission matrix in Table A7 all states tend to remain in their state primarily but main transitions occur between States 1 and 3, and between States 2 and 4.

According to the intercept of the parameters of the HMM model, the intercept of ozone of State 3 has the largest average mean ozone level intercept of 8295.05 DU aligned with the lowest aerosol of 0.04, and State 2 has the smallest average state-specific mean ozone level of 6580.94 DU and highest mean aerosol of 64.79 (Table A8).

From the beginning of the series (see Figure A6) until day 70 (9 November 2019) the system alternates between State 3 and 1, with State 3 having the highest mean ozone intercept parameter of 8295.05 DU and lowest mean aerosol of 0.04, and in contrast State 1 has the second lowest aerosol intercept mean of 6.04 and ozone level intercept mean of 7475.18 DU. Note that day 70 (9 November 2019)) is the same as the change-point for aerosol detected with Binary Segmentation (Figure 18) plus an additional 2 days.

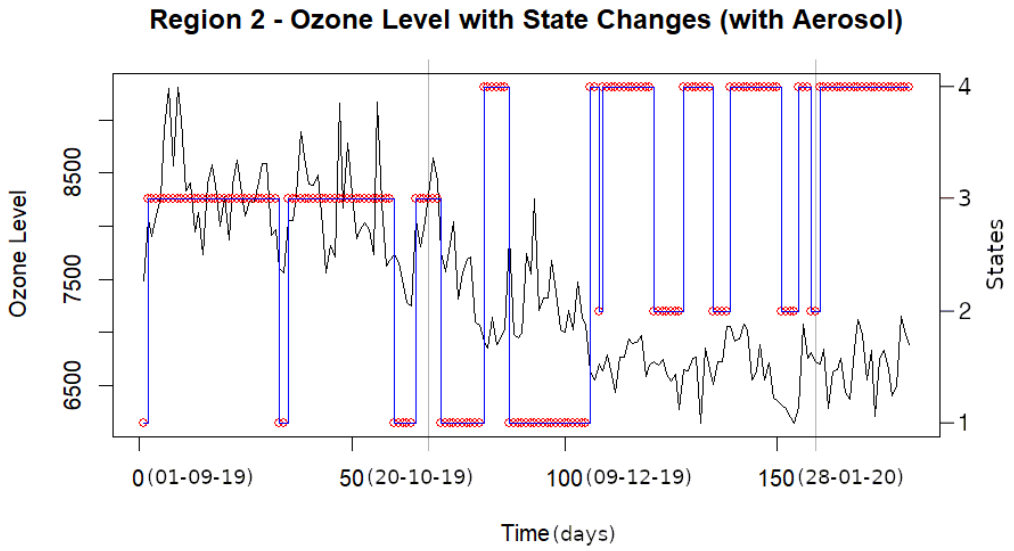

Figure A6. HMM bivariate state changes in total sum ozone time-series with respect to total sum aerosol time-series as a covariate for Region 2.

After day 70 (9 November 2019) up to and including day 107 (16 December 2019), ozone levels then vary between State 1 (mean ozone(DU), aerosol = (7475.18, 6.04)) and State 4 (mean ozone(DU), aerosol = (6742.16, 14.77)).

After day 107 (20 December 2019) to the end of the time series the system is transitioning between State 2 and State 4. Note State 2 (mean ozone(DU), aerosol = (6580.94, 64.79)) has the lowest ozone level and highest aerosol mean intercept, in contrast to State 4 (mean ozone(DU), aerosol = (6742.16, 14.77)).

Notably, the first sign of bushfire-associated aerosols occurs on day 68 (7 November 2019), and the time of state change in ozone levels indicated by HMM occurs approximately 2 days later, at 70 days (9 November 2019) in Region 2.

References

1. Binskin, M.; Bennett, A.; Macintosh, A. *The Royal Commission into National Natural Disaster Arrangements Report*; Final Report; Royal Commission into National Natural Disaster Arrangements: Canberra, Australia, 2020.
2. Australian Institute of Health and Welfare. *Australian Bushfires 2019–20: Exploring the Short-Term Health Impacts*; Cat. No. PHE 276; AIHW: Canberra, Australia, 2020. Available online: https://www.aihw.gov.au/reports/environment-and-health/short-term-health-impacts-2019-20-bushfires/related-material (accessed on 22 June 2022).
3. WWF. *Australian Bushfire Fund Final Report*; WWF: April 2022, 15 pages. Available online: https://www.worldwildlife.org/publications/australia-bushfire-fund-final-report#:~:text=The%20scale%20of%20Australia's%20bushfires,forests%20and%20wildlife%20was%20immense (accessed on 30 May 2022).
4. Lee, M. Australia's Bushfires Could Drive More Than 700 Animal Species to Extinction. Check the Numbers for Yourself. The Conversation. 2020. Available online: https://theconversation.com/australias-bushfires-could-drive-more-than-700-animal-species-to-extinction-check-the-numbers-for-yourself-129773 (accessed on 20 July 2021).
5. Dickman, C.R. Ecological consequences of Australia's "Black Summer" bushfires: Managing for recovery. *Integr. Environ. Assess. Manag.* **2021**, *17*, 1162–1167. [CrossRef] [PubMed]
6. Ward, M.; Tulloch, A.I.; Radford, J.Q.; Williams, B.A.; Reside, A.E.; Macdonald, S.L.; Mayfield, H.J.; Maron, M.; Possingham, H.P.; Vine, S.J.; et al. Impact of 2019–2020 mega-fires on Australian fauna habitat. *Nat. Ecol. Evol.* **2020**, *4*, 1321–1326. [CrossRef] [PubMed]
7. Hyman, I.; Ahyong, S.T.; Köhler, F.; McEvey, S.F.; Milledge, G.A.; Reid, C.A.M.; Rowley, J.J.L. Impacts of the 2019–2020 bushfires on New South Wales biodiversity: A rapid assessment of distribution data for selected invertebrate taxa. *Tech. Rep. Aust. Mus. Online* **2020**, *32*, 1–17. [CrossRef]
8. Shukla, P.R.; Skea, J.; Calvo Buendia, E.; Masson-Delmotte, V.; Portner, H-O.; Roberts, D.C.; Zhai, P.; Slade, R.; Connors, S.; van Diemen, R.; et al. (Eds.) *Climate Change and Land: IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems*; Cambridge Universiyt Press: New York, NY, USA, 2022. [CrossRef]

9. Xing, J.; Wang, J.; Mathur, R.; Wang, S.; Sarwar, G.; Pleim, J.; Hogrefe, C.; Zhang, Y.; Jiang, J.; Wong, D.C.; et al. Impacts of aerosol direct effects on tropospheric ozone through changes in atmospheric dynamics and photolysis rates. *Atmos. Chem. Phys.* **2017**, *17*, 9869–9883. [CrossRef]
10. Qu, Y.; Voulgarakis, A.; Wang, T.; Kasoar, M.; Wells, C.; Yuan, C.; Varma, S.; Mansfield, L. A study of the effect of aerosols on surface ozone through meteorology feedbacks over China. *Atmos. Chem. Phys.* **2021**, *21*, 5705–5718. [CrossRef]
11. Li, J.; Pósfai, M.; Hobbs, P.V.; Buseck, P.R. Individual aerosol particles from biomass burning in southern Africa: 2, Compositions and aging of inorganic particles. *J. Geophys. Res. Atmos.* **2003**, *108*, 8484. [CrossRef]
12. Pósfai, M.; Anderson, J.R.; Buseck, P.R.; Sievering, H. Soot and sulfate aerosol particles in the remote marine troposphere. *J. Geophys. Res. Atmos.* **1999**, *104*, 21685–21693. [CrossRef]
13. Ryan, R.G.; Silver, J.D.; Schofield, R. Air quality and health impact of 2019–20 Black Summer megafires and COVID-19 lockdown in Melbourne and Sydney, Australia. *Environ. Pollut.* **2021**, *274*, 116498. [CrossRef]
14. Dickman, C.; McDonald, T. Some personal reflections on the present and future of Australia's fauna in an increasingly fire-prone continent. *Ecol. Manag. Restor.* **2020**, *21*, 86–96. [CrossRef]
15. Yu, P.; Davis, S.M.; Toon, O.B.; Portmann, R.W.; Bardeen, C.G.; Barnes, J.E.; Telg, H.; Maloney, C.; Rosenlof, K.H. Persistent Stratospheric Warming Due to 2019–2020 Australian Wildfire Smoke. *Geophys. Res. Lett.* **2021**, *48*, e2021GL092609. [CrossRef]
16. Peterson, D.A.; Fromm, M.D.; McRae, R.H.; Campbell, J.R.; Hyer, E.J.; Taha, G.; Camacho, C.P.; Kablick, G.P., III; Schmidt, C.C.; DeLand, M.T. Australia's Black Summer pyrocumulonimbus super outbreak reveals potential for increasingly extreme stratospheric smoke events. *NPJ Clim. Atmos. Sci.* **2021**, *4*, 38. [CrossRef]
17. Santee, M.L.; Lambert, A.; Manney, G.L.; Livesey, N.J.; Froidevaux, L.; Neu, J.L.; Schwartz, M.J.; Millán, L.F.; Werner, F.; Read, W.G.; et al. Prolonged and Pervasive Perturbations in the Composition of the Southern Hemisphere Midlatitude Lower Stratosphere From the Australian New Year's Fires. *Geophys. Res. Lett.* **2022**, *49*, e2021GL096270. [CrossRef]
18. Visser, I.; Speekenbrink, M. depmixS4: An R Package for Hidden Markov Models. *J. Stat. Softw.* **2010**, *36*, 1–21. [CrossRef]
19. Solomon, S.; Stone, K.; Yu, P.; Murphy, D.M.; Kinnison, D.; Ravishankara, A.R.; Wang, P. Chlorine activation and enhanced ozone depletion induced by wildfire aerosol. *Nature* **2023**, *615*, 259–264. [CrossRef] [PubMed]
20. Bernath, P.; Boone, C.; Crouse, J. Wildfire smoke destroys stratospheric ozone. *Science* **2022**, *375*, 1292–1295. [CrossRef]
21. Yu, P.; Toon, O.B.; Bardeen, C.G.; Zhu, Y.; Rosenlof, K.H.; Portmann, R.W.; Thornberry, T.D.; Gao, R.S.; Davis, S.M.; Wolf, E.T.; et al. Black carbon lofts wildfire smoke high into the stratosphere to form a persistent plume. *Science* **2019**, *365*, 587–590. [CrossRef] [PubMed]
22. Mersmann, K.; Stein, T.; Carlowicz, M.; Allen, J. Ozone Hole Is Smallest Since 1988. NASA Earth Observatory. Available online: https://earthobservatory.nasa.gov/images/91212/ozone-hole-is-smallest-since-1988 (accessed on 9 April 2022).
23. Lindenmayer, D.B.; Taylor, C. New spatial analyses of Australian wildfires highlight the need for new fire, resource, and conservation policies. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 12481–12485. [CrossRef]
24. Li, M.; Shen, F.; Sun, X. 2019–2020 Australian bushfire air particulate pollution and impact on the South Pacific Ocean. *Sci. Rep.* **2021**, *11*, 12288. [CrossRef]
25. Solomon, S.; Dube, K.; Stone, K.; Yu, P.; Kinnison, D.; Toon, O.B.; Strahan, S.E.; Rosenlof, K.H.; Portmann, R.; Davis, S.; et al. On the stratospheric chemistry of midlatitude wildfire smoke. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2117325119. [CrossRef]
26. McNeill, V.F.; Thornton, J.A. How wildfires deplete ozone in the stratosphere. *Nature* **2023**, *615*, 219–221. [CrossRef]
27. Hengl, T.; Nussbaum, M.; Wright, M.N.; Heuvelink, G.B.; Gräler, B. Random forest as a generic framework for predictive modeling of spatial and spatiotemporal variables. *PeerJ* **2018**, *6*, e5518. [CrossRef] [PubMed]
28. Shumway, R.H.; Stoffer, D.S.; Stoffer, D.S. *Time Series Analysis and Its Applications*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 4.
29. Lopo, A.B.; Spyrides, M.H.C.; Lucio, P.S.; Sigró, J. UV Index Modeling by Autoregressive Distributed Lag (ADL Model). *Atmos. Clim. Sci.* **2014**. [CrossRef]
30. Herman, J.R. Global increase in UV irradiance during the past 30 years (1979–2008) estimated from satellite data. *J. Geophys. Res. Atmos.* **2010**, *115*, D04203. [CrossRef]
31. Herman, J.; Labow, G.; Hsu, N.; Larko, D. Changes in cloud and aerosol cover (1980–2006) from reflectivity time series using SeaWiFS, N7-TOMS, EP-TOMS, SBUV-2, and OMI radiance data. *J. Geophys. Res. Atmos.* **2009**, *114*, D01201. [CrossRef]
32. Hegglin, M.I.; Shepherd, T.G. Large climate-induced changes in ultraviolet index and stratosphere-to-troposphere ozone flux. *Nat. Geosci.* **2009**, *2*, 687–691. [CrossRef]
33. Douglass, A.; Fioletov, V.; Godin-Beekmann, S.; Müller, R.; Stolarski, R.S.; Webb, A.; Arola, A.; Burkholder, J.B.; Burrows, J.P.; Chipperfield, M.P.; et al. Scientific Assessment of Ozone Depletion: 2010, Chapter 2—Stratospheric Ozone and Surface Ultraviolet Radiation. In *Global Ozone Research and Monitoring Project—Report No. 52*; World Meteorological Organization: Geneva, Switzerland, 2010; 516p.
34. Jaross, G. OMPS-NPP L3 NM Ozone (O3) Total Column 1.0 Deg Grid Daily V2. Distributed by Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC). 2017. Digital Science Data. Available online: https://disc.gsfc.nasa.gov/datacollection/OMPS_NPP_NMTO3_L3_DAILY_2.html (accessed on 14 June 2022). [CrossRef]
35. McPeters, R.D.; Bhartia, P.K.; Krueger, A.J.; Herman, J.R. *Earth Probe Total Ozone Mapping Spectrometer (TOMS) Data Product User's Guide*; Goddard Space Flight Center: Greenbelt, MD, USA, 1998.

36. Konovalov, I.B.; Beekmann, M.; Kuznetsova, I.N.; Yurova, A.; Zvyagintsev, A.M. Atmospheric impacts of the 2010 Russian wildfires: Integrating modeling and measurements of an extreme air pollution episode in the Moscow region. *Atmos. Chem. Phys.* **2011**, *11*, 10031–10056. [CrossRef]
37. Killick, R.; Eckley, I.A. Changepoint: An R Package for Changepoint Analysis. *J. Stat. Softw.* **2014**, *58*, 1–19. [CrossRef]
38. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
39. Wikle, C.K.; Zammit-Mangion, A.; Cressie, N. *Spatio-Temporal Statistics with R*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2019.
40. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **2006**, *9*, 181–199. [CrossRef]
41. Biau, G.; Scornet, E. A random forest guided tour. *Test* **2016**, *25*, 197–227. [CrossRef]
42. Hooker, G.; Mentch, L. Comments on: A random forest guided tour. *Test* **2016**, *25*, 254–260. [CrossRef]
43. Wong, S.; Flegg, J.A.; Golding, N.; Kandanaarachchi, S. Comparison of new computational methods for geostatistical modeling of malaria. *arXiv* **2023**, arXiv:2305.01907.
44. Liaw, A.; Wiener, M. The R Journal: Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
45. Wright, M.N.; Ziegler, A. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* **2017**, *77*, 1–17. [CrossRef]
46. Murphy, K.P. Machine Learning: A Probabilistic Perspective. In *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012; pp. 515–542.
47. Peruzzi, M.; Dunson, D.B. Spatial Multivariate Trees for Big Data Bayesian Regression. *J. Mach. Learn. Res.* **2022**, *23*, 40.
48. Oliver, M.; Webster, R. A tutorial guide to geostatistics: Computing and modeling variograms and kriging. *Catena* **2014**, *113*, 56–69. [CrossRef]
49. Zammit-Mangion, A.; Cressie, N. FRK: An R Package for Spatial and Spatio-Temporal Prediction with Large Datasets. *J. Stat. Softw.* **2021**, *98*, 1–48. [CrossRef]
50. Sainsbury-Dale, M.; Zammit-Mangion, A.; Cressie, N. Modelling big, heterogeneous, non-Gaussian spatial and spatiotemporal data using FRK. *arXiv* **2022**, arXiv:2110.02507.
51. Meinshausen, N. Quantile Regression Forests. *J. Mach. Learn. Res.* **2006**, *7*, 983–999.
52. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2009.
53. Kämäräinen, M.; Tuovinen, J.P.; Kulmala, M.; Mammarella, I.; Aalto, J.; Vekuri, H.; Lohila, A.; Lintunen, A. Spatiotemporal lagging of predictors improves machine learning estimates of atmosphere–forest CO_2 exchange. *Biogeosciences* **2023**, *20*, 897–909. [CrossRef]
54. Horvath, L. The Maximum Likelihood Method for Testing Changes in the Parameters of Normal Observations. *Ann. Stat.* **1993**, *21*, 671–680. [CrossRef]
55. Gombay, E.; Horvath, L. An application of the maximum likelihood test to the change-point problem. *Stoch. Process. Their Appl.* **1994**, *50*, 161–171. [CrossRef]
56. Edwards, A.W.F.; Cavalli-Sforza, L.L. A method for cluster analysis. *Biometrics* **1965**, *21*, 362–375. [CrossRef]
57. Chen, J.; Gupta, A.K. *Parametric Statistical Change Point Analysis with Applications to Genetics, Medicine, and Finance*, 2nd ed.; Birkhäuser Boston: Boston, MA, USA, 2012. [CrossRef]
58. Fryzlewicz, P. Wild binary segmentation for multiple change-point detection. *Ann. Stat.* **2014**, *42*, 2243–2281. [CrossRef]
59. Baum, L.E.; Petrie, T. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *Ann. Math. Stat.* **1966**, *37*, 1554–1563. [CrossRef]
60. Zucchini, W.; MacDonald, I.L.; Langrock, R. *Hidden Markov Models for Time Series—An Introduction Using R*, 2nd ed.; Chapman and Hall/CRC Press: Boca Raton, FL, USA, 2016; p. 398. [CrossRef]
61. Molenaar, D.; Oberski, D.; Vermunt, J.; Boeck, P.D. Hidden Markov Item Response Theory Models for Responses and Response Times. *Multivar. Behav. Res.* **2016**, *51*, 606–626. [CrossRef] [PubMed]
62. Rabiner, L.; Juang, B. An introduction to hidden Markov models. *IEEE ASSP Mag.* **1986**, *3*, 4–16. [CrossRef]
63. Zhang, M.; Jiang, X.; Fang, Z.; Zeng, Y.; Xu, K. High-order Hidden Markov Model for trend prediction in financial time series. *Phys. A Stat. Mech. Its Appl.* **2019**, *517*, 1–12. [CrossRef]
64. Yip, C.F.; Ng, W.L.; Yau, C.Y. A hidden Markov model for earthquake prediction. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 1415–1434. [CrossRef]
65. Yoon, B.J. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Curr. Genom.* **2009**, *10*, 402–415. [CrossRef]
66. Reynolds, D. *Encyclopedia of Biometrics*; Springer: Boston, MA, USA, 2015; Chapter Gaussian Mixture Models, pp. 827–832. [CrossRef]
67. Alghamdi, R. Hidden Markov Models (HMMs) and Security Applications. *Int. J. Adv. Comput. Sci. Appl.* **2016**, *7*. [CrossRef]
68. Viterbi, A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inf. Theory* **1967**, *13*, 260–269. [CrossRef]
69. Zhang, H.; Zhang, W.; Palazoglu, A.; Sun, W. Prediction of ozone levels using a Hidden Markov Model (HMM) with Gamma distribution. *Atmos. Environ.* **2012**, *62*, 64–73. [CrossRef]

70. Sun, W.; Zhang, H.; Palazoglu, A. Prediction of 8 h-average ozone concentration using a supervised hidden Markov model combined with generalized linear models. *Atmos. Environ.* **2013**, *81*, 199–208. [CrossRef]
71. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 3rd ed.; OTexts: Melbourne, Australia, 2021.
72. McClintock, B.T.; Langrock, R.; Gimenez, O.; Cam, E.; Borchers, D.L.; Glennie, R.; Patterson, T.A. Uncovering ecological state dynamics with hidden Markov models. *Ecol. Lett.* **2020**, *23*, 1878–1903. [CrossRef]
73. Box, G.E.; Tiao, G.C. Intervention analysis with applications to economic and environmental problems. *J. Am. Stat. Assoc.* **1975**, *70*, 70–79. [CrossRef]
74. Chan, K.S.; Cryer, J.D. *Time Series Analysis with Applications in R*; Springer: Berlin/Heidelberg, Germany, 2008.
75. Zeileis, A. dynlm: Dynamic Linear Regression, Package Version 0.3-6. 2019. Available online: https://CRAN.R-project.org/package=dynlm (accessed on 10 October 2024).
76. Demirhan, H. dLagM: An R package for distributed lag models and ARDL bounds testing. *PLoS ONE* **2020**, *15*, e0228812. [CrossRef]
77. Truong, C.; Oudre, L.; Vayatis, N. Selective review of offline change-point detection methods. *Signal Process.* **2020**, *167*, 107299. [CrossRef]
78. Rao, A.R.; Wang, Q.; Wang, H.; Khorasgani, H.; Gupta, C. Spatio-Temporal Functional Neural Networks. In Proceedings of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), Sydney, Australia, 6–9 October 2020; pp. 81–89. [CrossRef]