Use of two-point models in “Model choice in time-series studies of air pollution and mortality”

Mieczysław Szyszkowicz¹

Received: 14 January 2019 / Accepted: 27 December 2019 / Published online: 10 January 2020
© The Author(s) 2020

Abstract

In this work, a new technique is proposed to study short-term exposure and adverse health effects. The presented approach uses hierarchical clusters with the following structure: each pair of two sequential days in 1 year is embedded in the year. We have 183 clusters per year with the embedded structure <year:2 days>. Time-series analysis is conducted using a conditional Poisson regression with the constructed clusters as a stratum. Unmeasured confounders such as seasonal and long-term trends are not modelled but are controlled by the structure of the clusters. The proposed technique is illustrated using four freely accessible databases, which contain complex simulated data. These data are available as the compressed R workspace files. Results based on the simulated data were very close to the truth based on the presented methodology. In addition, the case-crossover method with 1-month and 2-week window, and a conditional Poisson regression on 3-day clusters as a stratum, was also applied to the simulated data. Difficulties (high type I error rate) were observed for the case-crossover method in the presence of high concavity in the simulated data. The proposed methods using various forms of a stratum were further applied to the Chicago mortality data. The considered methods have often different qualitative and quantitative estimations.

Keywords Air pollution · Case-crossover · Cluster · Mortality · Poisson

Introduction

Time-series analysis is the most frequent and commonly used method for assessing the association between counts of health outcomes over a considered period of time and exposure to ambient air pollution. The problem of controlling for unmeasured confounders, such as seasonal and long-term trends, in the constructed statistical models, is not fully resolved. Time still remains to be the main concern in modelling Poisson time-series data. There are many criteria to determine an optimal number of knots used in the applied splines for representing the smooth function of time.

The case-crossover (CC) design introduced by Maclure (1991) eliminated time from the model. This fact probably is among the reasons that the CC method is so popular and

attractive. In the CC method, time is controlled by the realized design. The time-stratified approach (Janes et al. 2005) to determine the control periods to an event period separates time into 1-month segments. The standard CC method uses events (individual cases) and assigns the corresponding controls located in one common month. The case and control days are chosen to be the same days of the week (dow). In the CC method, time is still present in relation between case periods and control periods, as the controls are pre- and post-event time. It was especially influential in the original presentation of the CC method (Maclure 1991), where the control period was always before the case period. Such unidirectional (or symmetrical to the event time) strategy to assign controls may result in bias (Janes et al. 2005).

In 2006, Szyszkowicz (2006) proposed considering daily counts of the health events on the hierarchical clusters of the form <year:month:dow>. Such a structure is naturally embedded in the hierarchy of the calendar year. The statistical methodology used was mixed-effect models (Szyszkowicz 2006). In the publication, it was also observed that the structure of the form <year:2 weeks:dow> could also be implemented in the proposed models. It reduced each cluster to the same two weekdays during a 2-week period; and therefore, each cluster has only two data points. Since two points determine a line,

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11869-019-00787-5>) contains supplementary material, which is available to authorized users.

✉ Mieczysław Szyszkowicz
mietek.szyszkowicz@canada.ca

¹ Environmental Health Science and Research Bureau, Health Canada, 101 Tunney's Pasture, Ottawa, ON K1A 0K9, Canada

the slope related to the association between exposure and health counts on these two-point clusters can also be determined. In another publication (Szyszkowicz and Burr 2016), an approach to construct the chained clusters was presented, where each day was used twice in neighbored two point clusters. This technique increases the number of clusters and results in a smaller standard error of the estimated slope.

Further advancement in the statistical methodology proposed the use of a conditional Poisson regression (Armstrong et al. 2014) as a flexible alternative to the standard CC method, with stratum defined as three levels hierarchical cluster <year:month:dow>. As in the previous publications (Szyszkowicz 2006; Szyszkowicz and Burr 2016), daily counts are used here, not individual events. In these approaches, adjustment for day of week (dow) is realized by the design of the study.

In this paper, we propose to use clusters of the following form <year:2 days>. Therefore, in 1 year, there are 183 clusters of such type. The clusters are different during each year, as each year (data) has its own specific characteristics. It is the smallest possible time-window in a time-series study of daily counts; just neighbored 2 days are considered a common group. The conditional Poisson regression model was applied using the 2-day (2D) stratum, where day of week was adjusted in the constructed model. The impact of weekdays is modelled, and the time-window considered is narrowed to the smallest one.

Methods

The presented methodology is based on clusters of the form <year:2 days>. For example, the days of a year are grouped as follows: (1—first day of a year, 2—second day of a year), (3, 4), (5, 6), etc. For each pair, we assigned the cluster number: 1 for (1, 2), 2 for (3, 4), 3 for (5, 6), and so on. For example, the days January 1 and January 2 will have cluster number 1. Therefore, the clusters (2000, 1) and (2001, 1) are different, as they belong to different years. This construction results in 183 clusters per year, where the cluster number 183 has 2 days only in leap years. These clusters are used to control time. Time smoothers are not required in the models. The clusters are considered as time units. In addition, days of the week are adjusted for in the final model, which imposes another control for time. In this presentation, conditional Poisson regression model was applied for daily counts (Armstrong et al. 2014), where stratum was defined as the constructed 2-day cluster. This approach is referred to as the 2D method. The model has the following form in the R software (package gnm).

Model2D = gnm (Counts ~ AirPollution + dow + covariates, family = poisson, eliminate = factor(stratum)).

The calculations were carried out with the option family=quasipoisson, which allows for overdispersion in the

used Poisson models. Other covariates can be included in the model, although dow should always be present to control effects of weekdays as other forms of time influence. In addition, for comparison purpose, the clusters of the form <year:3 days> were also used with 122 such structures per year. This approach is referred to as the 3D method.

Very detailed analysis of the impact of time (seasonal, trends, and concavity) was carried out by the authors in the study related to air pollution in the form of particulate matter exposure and mortality (Peng et al. 2006). The authors generated simulated data for four different scenarios. The scenarios reflect the intensity of the used smoothing for time and air pollution with two effects of concavity levels, moderate and high. The results from the simulated data should not demonstrate any effect, i.e. there are no associations with the considered exposures and health outcomes. The true slopes (Beta) are zero. Four scenarios were considered and described in greater detail in (Peng et al. 2006): scenario 1 has more smoothing on time than on air pollution and moderate concavity; scenario 2 similar to scenario 1 but with high concavity; scenario 3 has more smoothing on air pollution than on time and moderate concavity; and scenario 4, similar to scenario 3 but with high levels of concavity.

Data from Minneapolis-St Paul for the years 1987–1994 were used to calculate “true” parameter estimates in order to specify smoothing functions, which were applied to the controlled corresponding simulations. Daily counts and air pollution were simulated 500 times for each considered scenario (Peng et al. 2006). Temperature data remain fixed and is the same in each of the simulations.

Data from the corresponding repository (Data from Peng et al. 2006) representing scenarios 1–4 were used in the form of compressed R workspace files. Each file contains 500 datasets corresponding to each of the four scenarios being considered. Therefore, for each scenario, 500 estimations of the slope are generated. In the current approach, the calendar dates starting from January 1, 1987 and ending on December 31, 1994 (2922 days) were added to the data files. The dates allow us to construct the two-point clusters. In a similar manner, 3-day clusters (3D) were created in order to fit the 3D model. Both variables of daily counts and daily concentrations of air pollution (particulate matter) have simulated values and were used in the analysis. Temperature was represented in the model in the form of natural spline with three degrees of freedom.

For comparison purposes, the same data were assessed using the standard CC method on clusters of the form <year:month:dow>; where each cluster includes 4 or 5 days (referred as CCM, where *M* is for month). The number of days depends on day of the week and length of the month. Also, the CC method was applied on the clusters of the form (year:2 weeks:day of week> (referred as CC2W, where *W* is for week). These clusters are composed of two of the same

weekdays with weekdays separated by 7 days in the calendar, e.g. Monday to Monday of the following week. This differs from the 2D structure, which has two neighboured days in the calendar.

The results from the simulated data should produce estimated slopes (Beta) equal to zero; therefore, real data was also considered where the true value of the slopes was unknown. The considered database represents mortality in Chicago for the years 1987–2000 (Wand 2003; Wood 2006). The endpoints of interest included all non-accidental death (Death), cardiovascular (CVD) death, respiratory (Resp) death, a combination of cardiovascular and respiratory (CVD&Resp) deaths, and other (Others) deaths defined as mortality not related to cardiovascular and respiratory. This last category was defined as follows: Others = Deaths – CVD&Resp, where CVD&Resp is a summary of both indicated types. These data were analysed using the models presented in this paper, namely 2D, 3D, CCM, and CC2W. Ambient air pollution exposure particulate matters (PM10) and ozone (O3) were considered in the model, as well as temperature which were represented by natural splines of three degrees of freedom.

The main objective of this paper is to present and compare the 2D method and its performance to the CCM method when the models are applied to simulated data (scenarios 1–4) and real Chicago (1987–2000) data. The 2D methodology can be extended to have clusters composed with many days. Following the proposed 3D clusters construction, by analogy one could also form 4D, 5D, and 6D clusters or include even more days, e.g. MD clusters, where the clusters MD are constructed using all days in 1 month.

In this presentation, we used and tested the following methods: CCM, CC2W, 2D, and 3D on four sets of the simulated data and on Chicago mortality data. In the appendix are provided the results for another mortality data generated by various forms of the methods.

Results

The results for the corresponding databases with the simulated data (scenario 1–scenario 4) are summarized in Tables 1 and 2, and illustrated in Figs. 1, 2, and 3. The estimated average of the slopes (Beta) and corresponding standard deviations (SD)

are shown for the CCM method and the 2D (2 days) method in Table 1. The results demonstrate that the proposed 2D method gives an overall average slope less than 0.00001 in all four scenarios, compared with the CC method where the overall average slopes were less than 0.001. Only for a small number of cases (maximum 26), the proposed method estimated 95% confidence intervals (CI) which did not include zero. This is equivalent to the type I error rate being less than 5% (26/500) under the proposed 2D model. In contrast, the standard CC method assessed the slopes incorrectly in the presence of high concurvity (scenarios 2 and 4), i.e. the type I error rate for scenario 2 was 34% (172/500) and for scenario 4, it was 38% (192/500).

Table 2, similar to Table 1, summarizes the results for the methods CC2W and 3D. Comparing these two models, it was observed that the type I error rate (i.e. number of cases that were rejected when the null hypothesis is true) was similar between the two models. For the CC2W model, the type I error rate was less than 6.6% again highest in the scenarios with high concurvity; whereas for the 3D model, the type I error rate was less than 5.6%.

In Fig. 1 there are 4 panels, where each panel represents the results from the four scenarios considered in this analysis, e.g. panels a–d represent scenarios 1–4, respectively. Within each panel, the estimated slopes and corresponding 95% CI based on the CCM and 2D methods for each of the 500 runs are depicted. The slopes were ordered, smallest to largest, according to the values obtained from the standard CCM method. Black symbols were used to represent estimations generated by the CCM method; these appear on the graphs as continuous lines. A horizontal line showing the value of zero is also represented in each of the panels. In scenarios 2 and 4 (panel b and d), the lower limit of the 95% CI is above the zero level for the CCM method for almost 40% of cases. The numerical results from the 2D method are also sorted by the slopes from the CCM method; the represented runs for both methods are similar.

Figure 2 represents the results generated by the CC2W and 3D methods. The estimations are sorted by the values obtained from the CC2W method. The produced estimations from both methods are symmetrical with respect to the line representing zero. Figure 3 displays only the estimated slopes (Beta) based on all 4

Table 1 Estimated values by two methods, CCM and 2D, for the simulations # (1–4): average of Beta's values, their standard deviation (SD), and the number of rejections null hypothesis (RNH)

Simulation #	CCM: Beta (SD)	2D: Beta (SD)	CCM: RNH	2D: RNH
1	1.67E–04 (2.94E–04)	1.47E–05 (3.49E–04)	46	21
2	1.31E–03 (8.17E–04)	6.20E–05 (1.13E–03)	172	23
3	1.68E–04 (2.90E–04)	– 5.00E–06 (3.55E–04)	42	26
4	1.28E–03 (7.65E–04)	4.64E–06 (1.18E–03)	192	26

Table 2 Estimated values by two methods, CC2W and 3D, for the simulations # (1–4): average of Beta's values, standard deviation (SD), and the number of rejections null hypothesis (RNH)

Simulation #	CC2W: Beta (SD)	3D: Beta (SD)	CC2W: RNH	3D: RNH
1	7.09E-05 (3.76E-04)	2.27E-05 (3.02E-04)	29	26
2	4.10E-04 (1.11E-03)	2.38E-05 (1.01E-03)	33	28
3	3.43E-05 (3.59E-04)	1.03E-06 (3.13E-04)	24	26
4	4.42E-04 (1.09E-03)	5.31E-05 (9.82E-04)	34	20

methods of analysis (CCM, CC2W, 2D, and 3D) for the simulated data scenarios 1 to 4. Top panel represents the results for the CC methods where each pair of results, generated by CCM and CC2W, is sorted by CCM values. Bottom panel shows the results based on the pairs 2D and 3D, sorted by 2D values. In this figure, only cases which are in one common pair represent the same sample. This figure illustrates the distributions of the estimated slopes. The 2D and 3D methods show symmetrical distribution with respect to the horizontal line at zero. In contrast, the CCM method shows skewed values (above zero), mainly for the scenario 2 and 4.

Figure 4 illustrates the results obtained using the four models (2D, 3D, CCM, and CC2W) applied to the

Chicago mortality database where daily death counts belong to five defined categories (Death, CVD, Resp, CVD&Resp, and Others). The most intriguing result is that the standard CC method (CCM) estimates statistically significant positive slope for all deaths (Deaths) and exposure to particulate matter, but it is not the case when the CC2W, 2D, and 3D methods were applied. The proposed 2D and 3D method estimated positive and statistical significant association for cardiovascular (CVD) death and exposure to ambient ozone.

In addition, the proposed method 2D was realized on the simulation data described and presented in the publication (Burr et al. 2014). The authors created the set of four simulation data (Sim1–Sim4), each with 250 samples for a 10-year period. These data present various

Fig. 1 Estimated slopes (Beta) and 95% CIs obtained by using the standard CC and 2D method. The results are ordered by Beta values based on the CCM method. The results for the CC method are represented by small black symbols (squares: Beta, triangles: lower limit of 95% CI, diamonds: upper limit of 95% CI). The same symbols, but larger and in colour (i.e.

red squares: Beta, blue triangles: lower limit of 95% CI, and green diamonds: upper limit of 95% CI) are used to illustrate the corresponding values generated by the 2D method. The x-axis identifies the simulated samples (1–500), ordered differently for each scenario. **a–d** Results for scenario 1–scenario 4, respectively

Fig. 2 Estimated slopes (Beta) and 95% CIs obtained by using the CC2W and 3D method. The results are ordered by Beta values based on the CC2W method. The results for the CC2W method are represented by small black symbols (squares: Beta, triangles: lower limit of 95% CI, diamonds: upper limit of 95% CI). The same symbols, but larger and in

colour (i.e. red squares: Beta, blue triangles: lower limit of 95% CI, and green diamonds: upper limit of 95% CI) are used to illustrate the corresponding values generated by the 3D method. The x-axis identifies the simulated samples (1–500), ordered differently for each scenario. **a–d** Results for scenario 1–scenario 4, respectively

difficulties for the standard time-series methods. In this case, their true slope is 1.0 (Beta = 1.0). Using the 2D method, the minimum among lower bounds of the 95% confidence intervals were estimated as 0.97, 0.97, 0.95, and 0.94, for Sim1–Sim4, respectively. The corresponding data, codes in R, and the results are presented at <https://github.com/szyszkowicz/Data2D>.

Discussion

The techniques presented here are relatively simple and practical. They can be very easily applied without additional efforts. Existing software allows constructing the corresponding models. The four simulated scenarios that are publically available data sets (Data from Peng et al. 2006) used in this paper allow readers to execute the described calculations. The results also are discussed in publication (Peng et al. 2006). In addition, one real and freely available mortality data (Chicago) can be used to test the methods proposed here. Also the data presented at <https://github.com/szyszkowicz/Data2D>.

In the situation of the simulated data, the standard CC method demonstrates some difficulties in the presence of high concavity. It is a signal that some real analysis results may be affected by the collinearity for non-linear correlation. The proposed 2D method does not have problems with data sets having high concavity, e.g. scenarios 2 and 4.

Many details, results, and discussion are provided in publication (Peng et al. 2006), which evaluated the problems related to unmeasured confounders. Briefly, for the simulated data, the authors of the presentation (Peng et al. 2006) applied the five df (degree of freedom) selection methods and investigated under which circumstances the used method wrongly reports a statistically significant air pollution association. They found that the bias in the estimates generally decreases with more intensive smoothing. They also reported that the used model selection methods which optimize prediction may not be suitable to create an estimate with small bias. The 2D method approach presented in this paper is one of the possible solutions of these problems in environmental epidemiology. At least it allows to obtain an additional information.

As the results for the real data (Chicago) show we may have various responses generated by the proposed

Fig. 3 Estimated beta values generated by the four considered methods (CCM, CC2W, 2D, and 3D) for the simulated data. The results are grouped into the pairs: (CCM, CC2W) and (2D, 3D). Top panel shows the values for the pairs (CCM, CC2W) sorted using CCM estimations. Bottom panel shows the values for the pairs (2D, 3D) sorted by using 2D estimations. The x-axis identifies the simulated samples (1–500) ordered differently for each scenario. The results are presented for four scenarios # (1–4) identified by (a)–(d), respectively

methods, qualitative (statistically significant or non-significant) and quantitative (estimated slopes and confidence intervals) in nature. For Chicago all mortality (Death), the CCM method indicates associations but 2D method does not. The opposite is true for CVD and exposure to ambient ozone. Such discrepancy is difficult to judge and decide “who is right.” For these data estimated, parametric non-linear concentration-response shapes were reported (Szyszkowicz 2018). The shapes suggest existence of a threshold for analysed exposure. The presence of the threshold may affect associations when we consider 2-day (2D) and whole month (CCM or MD) time-window.

Probably the associations are also related to the properties of the analysed health outcomes. The 2D method realizes many comparisons between two neighbored days with various (lagged) air pollution exposure but

the health events are considered and compared in these 2 days. It may result that for acute health events (such as stroke), the association will be shown for the 2D method (two points) but not for the CC method (4 or 5 points).

The results for Chicago indicate that cardiovascular mortality (CVD) is related to ambient ozone. These associations are estimated by 2D and 3D methods, which use a short distance among the considered data points. In contrast, these methods, 2D and 3D, did not indicate significant associations of total non-accidental mortality (Death) with exposure to ambient particulate matter.

In conclusion, as the proposed methods are easy to implement and realize, it is suggested that during a short-term exposure study, the use of two methods, one with a narrow time-window (e.g. 2D, 3D methods) and another with a wider time-window (e.g. CCM or MD) be fit to the data.

Fig. 4 Estimated slopes and 95% CIs for five causes of deaths and two types of exposure: particulate matter (PM10) and ambient ozone (O3). Four statistical methods (CCM, CC2W, 2D, and 3D) were applied to the Chicago mortality, 1987–2000 data

Appendix

Here are shown the results obtained for Milan mortality data. Ten statistical methods are realized

Data and methods

The approach presented in the paper using 2D (2 days) clusters could be extended to use more days in one cluster. Following the proposed 3D clusters construction, by an analogy one could also form 4D up to 8D clusters, as were used here, or even include more days, and among them also realize MD (month-day, referring to all days in a given month) clusters. These methods were tested using the Milan mortality database spanning the years 1980–1989 in Milan, Italy (Wood 2006). As air pollution exposure total suspended particles (TSP) and sulphur dioxide (SO₂) were used. As in the case of Chicago mortality data in the constructed models, temperature, and also in this case relative humidity, were represented by natural spline of three degrees of freedom

Results

The results for the Milan mortality data are presented on Fig. S1. For exposure to TSP, positive and statistically significant associations were obtained for the methods 4D, 6D, 8D, MD, and CCM. In case of the exposure to sulphur dioxide, the 4D, 8D, MD, and CCM methods estimated a significant impact of the considered exposure on the mortality in Milan. Here we observe various behaviours of the proposed methods. In this considered situation, the models with large clusters (4, 6, 8 days, and 1 month) indicate the statistically significant positive associations. Other presented methods fail to show this association. It however remains difficult to assess which models are correct in their estimations.

Discussion

As the results for the real data (Chicago and Milan) show we may have various responses generated by the proposed methods, qualitative (statistically significant or non-significant) and quantitative (estimated slopes have different values) in nature. For Chicago all mortality (Death), the CCM

method indicates associations but 2D method does not. The reverse is true for CVD and exposure to ambient ozone. Such discrepancy is difficult to judge and decide “who is right.” For these data, estimated parametric non-linear concentration-response shapes were reported (Szyszkowicz 2018). The shapes suggest existence of a threshold for analysed exposure. The presence of the threshold may affect associations when we consider 2-day (2D) and whole month (CCM or MD) time-window. In case of the Milan data, we do not have a more specific category of mortality, say CVD. The results suggest that for some health outcomes, information taken from sequential few days (2D, 3D) is different when these days are separated (CCM or CC2W). Of course, we can assign recent air pollution exposure or lagged by a specified number of days.

The associations are probably also related to the properties of the analysed health outcomes. The results for Chicago indicate that CVD is related to ambient ozone levels. These associations are estimated by 2D and 3D methods, which use a short distance among the considered data points. At the same time, both these methods, 2D and 3D, did not describe any relationship between deaths and particulate matter exposure.

It is interesting that the chained clusters 2DC (Szyszkowicz and Burr 2016), where 2 days are paired as (1, 2), (2, 3), and (3, 4) also do not suggest any association for these data. Estimated slopes and their 95% confidence intervals are as follows: $9.19\text{e-}05$ (95% CI, $-8.67\text{e-}05$, $2.70\text{e-}04$) for TSP and $-1.01\text{e-}05$ (95% CI, $-1.55\text{e-}04$, $1.39\text{e-}04$) for SO₂.

In conclusion, as the proposed methods are easy to implement, we recommend using two methods, one with a narrow time-window (e.g. 2D, 3D) and another with a wider time-window (e.g. CCM or MD) when assessing the impact of air pollution on health and related outcomes

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article

are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Armstrong BG, Gasparrini A, Tobias A (2014) Conditional Poisson models: a flexible alternative to conditional logistic case cross-over analysis. *BMC Med Res Methodol* 14:122. <https://doi.org/10.1186/1471-2288-14-122>
- Burr WS, Takahara G, Shin HH (2014) Bias correction in estimation of public health risk attributable to short-term air pollution exposure. *Environmetrics*. <https://doi.org/10.1002/env.2337>
- Data associated with (Peng et al., 2006): <http://www.biostat.jhsph.edu/~rpeng/RR/tsmodelchoice/h> Accessed 6 Nov 2018
- Janes H, Sheppard L, Lumley T (2005) Case-crossover analyses of air pollution exposure data. Referent selection strategies and their implications for bias. *Epidemiology* 16(6):717–726
- Maclure M (1991) The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol* 133(2):144–153
- Peng RD, Dominici F, Louis TA (2006) Model choice in time series studies of air pollution and mortality. *J R Stat Soc Ser A* 169:179–203
- Szyszkowicz M (2006) Use of generalized linear mixed models to examine the association between air pollution and health outcomes. *Int J Occup Med Environ Health* 19(4):224–227
- Szyszkowicz M (2018) Concentration–response functions for short-term exposure and air pollution health effects. *Environ Epidemiol* 2(2). <https://doi.org/10.1097/EE9.0000000000000011>
- Szyszkowicz M, Burr WS (2016) The use of chained two-point clusters for the examination of associations of air pollution with health conditions. *Int J Occup Med Environ Health* 29(4):613–622. <https://doi.org/10.13075/ijomch.1896.00379>
- Wand M (2003) The package “SemiPar” in the R library. In the package milan.mortality data frame. Functions for semiparametric regression analysis, to complement the book. In: Ruppert D, Wand MP, Carroll RJ (eds) *Semiparametric regression*. Cambridge University Press, Cambridge
- Wood S (2006) The gamair package. In: Wood S (ed) *Generalized additive models: an introduction with R*. CRC, Boca Raton

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.