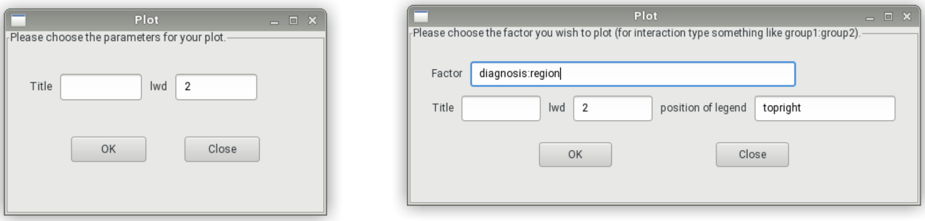

Plot

Please choose the parameters for your plot.

Title

lwd 2

Plot

Please choose the factor you wish to plot (for interaction type something like group1:group2).

Factor diagnosis:region

Title

lwd 2

position of legend topright

Figure 5: Graphical user interfaces for plotting: The left GUI is for the one-way layout (no choice of factors possible), the right one is for a two-way layout with an example for plotting interactions.

Discussion and Outlook

We have explicitly described the usage of the R package **MANOVA.RM** for analyzing various nonparametric multivariate MANOVA and RM designs making use of novel bootstrap- and permutation-approaches. Moreover, the corresponding models and inference procedures that have been derived and theoretically analyzed in previous papers are explained as well. In particular, three different test statistics of Wald-, ANOVA- and modified ANOVA-type are implemented together with appropriate critical values derived from asymptotic considerations, approximations or novel resampling approaches. Here, the latter is recommended in case of small to moderate sample sizes. All methods can be applied *without* assuming usual presumptions such as multivariate normality or specific covariance structures. Moreover, all procedures are particularly constructed to tackle covariance matrix heterogeneity across groups or even covariance singularity (in case of the MATS). In this way **MANOVA.RM** provides a flexible tool box for inferring hypotheses about (i) main and interaction effects in general factorial MANOVA and (ii) between and within subject effects in RM designs with possibly complex factorial structures on both, between and within subject factors.

In addition, we have placed a graphical user interface (GUI) at the users disposal to allow for a simple and intuitive use. It is planned to update the package on a regular basis; respecting the development of new procedures for general RM and MANOVA designs. For example, our working group is currently investigating the implementation of covariates in the above model in theoretical research and the resulting procedure may be incorporated in the future. Other topics include the possible implementation/improvement of subsequent multiple comparisons by the closure principle, as in Burchett et al. (2017), for both MANOVA and RM designs.

Acknowledgments

The work of Sarah Friedrich and Markus Pauly was supported by the German Research Foundation project DFG-PA 2409/3-1.

Bibliography

M. J. Anderson. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1):32–46, 2001. [p]

A. C. Bathke, S. Friedrich, F. Konietschke, M. Pauly, W. Staffen, N. Strobl, and Y. Höller. Testing Mean Differences Among Groups: Multivariate and Repeated Measures Analysis with Minimal Assumptions. *Multivariate Behavioral Research*, 2018. URL `https://doi.org/10.1080/00273171.2018.1446320`. [p]

E. Brunner. Asymptotic and approximate analysis of repeated measures designs under heteroscedasticity. *Mathematical Statistics with Applications in Biometry*, 2001. [p]

E. Brunner, H. Dette, and A. Munk. Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association*, 92(440):1494–1502, 1997. [p]

E. Brunner, S. Domhof, and F. Langer. *Nonparametric Analysis of Longitudinal Data in Factorial Experiments.* John Wiley & Sons, New York, USA, 2002. [p]

W. W. Burchett, A. R. Ellis, S. W. Harrar, and A. C. Bathke. Nonparametric inference for multivariate data: The R package npmv. *Journal of Statistical Software*, 76(4):1–18, 2017. URL `https://doi.org/10.18637/jss.v076.i04`. [p]

W. F. Christensen and A. C. Rencher. A Comparison of Type I Error Rates and Power Levels for Seven Solutions to the Multivariate Behrens-Fisher Problem. *Communications in Statistics - Simulation and Computation*, 26(4):1251–1273, 1997. [p]

E. Chung and J. P. Romano. Exact and asymptotically robust permutation tests. *The Annals of Statistics*, 41(2):484–507, 2013. [p]

G. Cornu, F. Mortier, C. Trottier, and X. Bry. *SCGLR: Supervised Component Generalized Linear Regression*, 2018. URL `https://CRAN.R-project.org/package=SCGLR`. R package version 3.0. [p]

C. S. Davis. *Statistical Methods for the Analysis of Repeated Measurements*. Springer-Verlag, 2002. [p]

B. S. Everitt and T. Hothorn. *HSAUR: A Handbook of Statistical Analyses Using R (1st Edition)*, 2017. URL `https://CRAN.R-project.org/package=HSAUR`. R package version 1.3-9. [p]

L. Finos, with contributions by Florian Klinglmueller, D. Basso, A. Solari, L. Benetazzo, J. Goeman, and M. Rinaldo. *Flip: Multivariate Permutation Tests*, 2018. URL `https://CRAN.R-project.org/package=flip`. R package version 2.5.0. [p]

J. Fox and S. Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, 2nd edition, 2011. URL `http://socserv.socsci.mcmaster.ca/jfox/Books/Companion`. [p]

S. Friedrich and M. Pauly. MATS: Inference for potentially singular and heteroscedastic MANOVA. *Journal of Multivariate Analysis*, 165:166–179, 2018. [p]

S. Friedrich, E. Brunner, and M. Pauly. Permuting longitudinal data in spite of the dependencies. *Journal of Multivariate Analysis*, 153:255–265, 2017a. [p]

S. Friedrich, F. Konietschke, and M. Pauly. Gfd: An r package for the analysis of general factorial designs. *Journal of Statistical Software, Code Snippets*, 79(1):1–18, 2017b. URL `https://doi.org/10.18637/jss.v079.c01`. [p]

S. Friedrich, F. Konietschke, and M. Pauly. *MANOVA.RM: Resampling-Based Analysis of Multivariate Data and Repeated Measures Designs*, 2019. URL `http://github.com/smn74/MANOVA.RM`. R package version 0.4.1. [p]

A. K. Gupta, S. W. Harrar, and Y. Fujikoshi. MANOVA for large hypothesis degrees of freedom under non-normality. *Test*, 17(1):120–137, 2008. [p]

S. W. Harrar and A. C. Bathke. A modified two-factor multivariate analysis of variance: Asymptotics and small sample approximations. *Annals of the Institute of Statistical Mathematics*, 64(1):135–165, 2012. [p]

T. Hothorn, F. Bretz, and P. Westfall. Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3):346–363, 2008a. [p]

T. Hothorn, F. Bretz, and P. Westfall. Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3):346–363, 2008b. [p]

Y. Huang, H. Xu, V. Calian, and J. C. Hsu. To permute or not to permute. *Bioinformatics*, 22(18): 2244–2248, 2006. [p]

R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, 2007. [p]

F. Konietschke, A. C. Bathke, S. W. Harrar, and M. Pauly. Parametric and nonparametric bootstrap methods for general MANOVA. *Journal of Multivariate Analysis*, 140:291–301, 2015. [p]

K. Krishnamoorthy and J. Yu. Modified Nel and Van Der Merwe Test for the Multivariate Behrens–Fisher Problem. *Statistics & Probability Letters*, 66(2):161–169, 2004. [p]

M. Lawrence and D. Temple Lang. RGtk2: A graphical user interface toolkit for R. *Journal of Statistical Software*, 37(8):1–52, 2010. URL `http://www.jstatsoft.org/v37/i08/`. [p]

J. Lemon. Plotrix: a package in the red light district of r. *R-News*, 6(4):8–12, 2006. [p]

P. Livacic-Rojas, G. Vallejo, and P. Fernandez. Analysis of type i error rates of univariate and multivariate procedures in repeated measures designs. *Communications in Statistics - Simulation and Computation*, 39(3):624–640, 2010. [p]

P. Livacic-Rojas, G. Vallejo, P. Fernández, and E. Tuero-Herrero. Power of modified brown-forsythe and mixed-model approaches in split-plot designs. *Methodology*, 13:9–22, 2017. [p]

L. M. Lix and H. Keselman. Multivariate tests of means in independent groups designs: Effects of covariance heterogeneity and nonnormality. *Evaluation & the Health Professions*, 27(1):45–69, 2004. [p]

L. M. Lix and A. M. Lloyd. A comparison of procedures for the analysis of multivariate repeated measurements. *Journal of Modern Applied Statistical Methods*, 6(2):380–398, 2007. [p]

M. McFarquhar, S. McKie, R. Emsley, J. Suckling, R. Elliott, and S. Williams. Multivariate and repeated measures (mrm): A new toolbox for dependent and multimodal group-level neuroimaging data. *NeuroImage*, 132:373–389, 2016. [p]

D. Murdoch and E. D. Chow. *Ellipse: Functions for Drawing Ellipses and Ellipse-Like Confidence Regions*, 2018. URL https://CRAN.R-project.org/package=ellipse. R package version 0.4.1. [p]

D. Nel and C. Van der Merwe. A Solution to the Multivariate Behrens-Fisher Problem. *Communications in Statistics-Theory and Methods*, 15(12):3719–3735, 1986. [p]

K. Noguchi, Y. R. Gel, E. Brunner, and F. Konietschke. nparld: An r software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software*, 50(12):1–23, 2012. URL http://www.jstatsoft.org/v50/i12/. [p]

G. Vallejo, A. Fidalgo, and P. Fernandez. Effects of covariance heterogeneity on three procedures for analyzing multivariate repeated measures designs. *Multivariate Behavioral Research*, 36(1):01–27, 2001. [p]

G. Vallejo Seco, J. A. Gras, and M. Ato García. Comparative robustness of recent methods for analyzing multivariate repeated measures designs. *Educational and Psychological Measurement*, 67(3):410–432, 2007. [p]

S. Xiao and J.-T. Zhang. Modified Tests for Heteroscedastic Two-Way MANOVA. *Journal of Advanced Statistics*, 1(1):1–16, 2016. [p]

H. Yanagihara and K.-H. Yuan. Three Approximate Solutions to the Multivariate Behrens–Fisher Problem. *Communications in Statistics - Simulation and Computation*, 34(4):975–988, 2005. [p]

Y. Yao. An Approximate Degrees of Freedom Solution to the Multivariate Behrens–Fisher Problem. *Biometrika*, 52(1/2):139–147, 1965. [p]

J.-T. Zhang. Two-way MANOVA with unequal cell sizes and unequal cell covariance matrices. *Technometrics*, 53(4):426–439, 2011. [p]

Øyvind Langsrud and B.-H. Mevik. *Ffmanova: Fifty-Fifty MANOVA*, 2019. URL https://CRAN.R-project.org/package=ffmanova. R package version 1.1.0. [p]

Sarah Friedrich
Department of Medical Statistics, University Medical Center Göttingen
Humboldtallee 32, 37073 Göttingen
Germany
sarah.friedrich@med.uni-goettingen.de

Frank Konietschke
Charité — Universitätsmedizin Berlin
Corporate Member of Freie Universität Berlin
Humboldt-Universität zu Berlin, and
Berlin Institute of Health, Institute of Biometry and Clinical Epidemiology
Charitéplatz 1, 10117 Berlin
Germany
frank.konietschke@charite.de

Markus Pauly
Fakultät Statistik, Technische Universität Dortmund
44221 Dortmund
Germany
markus.pauly@tu-dortmund.de

spGARCH: An R-Package for Spatial and Spatiotemporal ARCH and GARCH models

by Philipp Otto

Abstract In this paper, a general overview on spatial and spatiotemporal ARCH models is provided. In particular, we distinguish between three different spatial ARCH-type models. In addition to the original definition of Otto et al. (2016), we introduce an logarithmic spatial ARCH model in this paper. For this new model, maximum-likelihood estimators for the parameters are proposed. In addition, we consider a new complex-valued definition of the spatial ARCH process. Moreover, spatial GARCH models are briefly discussed. From a practical point of view, the use of the R-package **spGARCH** is demonstrated. To be precise, we show how the proposed spatial ARCH models can be simulated and summarize the variety of spatial models, which can be estimated by the estimation functions provided in the package. Eventually, we apply all procedures to a real-data example.

Introduction

Whereas autoregressive conditional heteroscedasticity (ARCH) models are applied widely in time series analysis, especially in financial econometrics, spatial conditional heteroscedasticity has not been seen as critical issue in spatial econometrics up to now. Although it is well-known that classical least squares estimators are biased for spatially correlated data as well as for spatial data with an inhomogeneous variance across space, there are just a few papers proposing statistical models accounting for spatial conditional heteroscedasticity in terms of the ARCH and GARCH models of Engle (1982) and Bollerslev (1986). The first extensions to spatial models attempted were time series models incorporating spatial effects in temporal lags (see Borovkova and Lopuhaa 2012 and Caporin and Paruolo 2006, for instance). Instantaneous spatial autoregressive dependence in the conditional second moments, i.e., the conditional variance in each spatial location is influenced by the variance nearby, has been introduced by Otto et al. (2016). Further details and derivations can also be found in Otto et al. (2018, 2019). Their models allow for these instantaneous effects but require certain regularity conditions. In this paper, we propose an alternative specification of spatial autoregressive conditional heteroscedasticity based on an exponential definition of the conditional variance. This new model can be seen as the spatial equivalent of the log-GARCH model by Pantula (1986); Geweke (1986); Milhøj (1987). Other recent papers propose a mixture of these two approaches (see Sato and Matsuda 2017, 2018b). Moreover, all these models can be used in spatiotemporal settings (see Otto et al. 2018; Sato and Matsuda 2018a).

In addition to the novel spatial logarithmic ARCH model, this paper demonstrates the use of the R-package **spGARCH**. From this practical point of view, the simulation of several spatial ARCH-type models as well as the estimation of a variety of spatial models with conditional heteroscedasticity are shown. There are several packages implementing geostatistical models, kriging approaches, and other spatial models (cf. Cressie 1993; Cressie and Wikle 2011). One of the most powerful packages used to deal with models of spatial dependence is **spdep**, written by Bivand and Piras (2015). It implements most spatial models in a user-friendly way, such as spatial autoregressive models, spatial lag models, and so forth (see, also, Elhorst 2010 for an overview). These models are typically called spatial econometrics models, although they are not tied to applications in economics. In contrast, the package **gstat** provides functions for geostatistical models, variogram estimation, and various kriging approaches (see Pebesma 2004 for details). For dealing with big geospatial data, the **Stem** package uses an expectation-maximization (EM) algorithm for fitting hierarchical spatiotemporal models (see Cameletti 2015 for details). For a distributed computing environment, the MATLAB software D-STEM from Finazzi and Fasso (2014) also provides powerful tools for dealing with heterogeneous spatial supports, large multivariate data sets, and heterogeneous spatial sampling networks. Additionally, these fitted models are suitable for spatial imputation. Contrary to these EM approaches, Bayesian methods for modeling spatial data are implemented in the R-**INLA** package (see Rue et al. 2009 for technical details of the integrated nested Laplace approximations and Martins et al. 2013 for recently implemented features). Along with this package, the R-**INLA** project provides several functions for diverse spatial models incorporating integrated nested Laplace approximations.

In contrast to the above mentioned software for spatial models, the prevalent R-package for time series GARCH-type models is **rugarch** from Ghalanos (2018). Since **spGARCH** has been developed mainly to deal with spatial data, we aim to provide a package which is user-friendly for researchers and data scientists working in applied spatial science. Thus, the package is coordinated with the objects and ideas of R packages for spatial data rather than packages for dealing with time series.

We structure the paper as follows. In the next Section Spatial ARCH-type models, we discuss all covered spatial and spatiotemporal ARCH-type models. In addition, we introduce a novel logarithmic spatial ARCH model, which has weaker regularity conditions than the other spatial ARCH models. In the subsequent section, parameter estimation based on the maximum-likelihood principle is discussed for both the previously proposed spatial ARCH models as well as the new logarithmic spatial ARCH model. Furthermore, spatial GARCH models are briefly discussed. However, the focus of this paper should be on ARCH-type models. After these theoretical sections, we demonstrate the use of the R-package **spGARCH** in Section Overview of the R-Package spGARCH. Further, we fit a spatial autoregressive model with exogenous regressors and spatial ARCH residuals for a real-world data set. In particular, we analyze prostate cancer incidence rates in southeastern U.S. states. Section Summary and discussion concludes the paper.

Spatial ARCH-type models

Let $\{Y(\boldsymbol{s}) \in \mathbb{R} : \boldsymbol{s} \in D\}$ be a univariate stochastic process having a spatial autoregressive structure in the conditional variance. The process is defined in a multidimensional space D, which is typically a subset of the q-dimensional real numbers $\mathbb{R}^q$, as space is usually finite. For dealing with spatial lattice data, D is subset of the q-dimensional integers $\mathbb{Z}^q$. For both cases, it is important that the subset contains a q-dimensional rectangle of positive volume (cf. Cressie and Wikle 2011). Moreover, this definition is suitable for modeling spatiotemporal data, as one might assume that D is the product set $\mathbb{R}^k \times \mathbb{Z}^l$ with $k + l = d$.

To define spatial models, in particular areal spatial models such as the simultaneous autoregressive (SAR) models, it is convenient to consider a vector of observations $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))'$ at all locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$. For spatial ARCH models, we specify this vector as

$$\boldsymbol{Y} = \operatorname{diag}(\boldsymbol{h})^{1/2}\boldsymbol{\varepsilon}\,, \tag{1}$$

an analogue to the well-known time series ARCH models (cf. Engle 1982; Bollerslev 1986). However, note that the vector $\boldsymbol{h}$ does not necessarily coincide with the conditional variance

$$Var(Y(\boldsymbol{s}_i)|Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_{i-1}))\,,$$

as the variance in any location $\boldsymbol{s}_j$ also depends on $Y(\boldsymbol{s}_i)$ for $j \neq i$ (see Otto et al. 2018 for details). We now distinguish between several spatial ARCH-type models via the definition of $\boldsymbol{h}$.

Spatial ARCH model

First, we define this vector $\boldsymbol{h}$ in such a way as to be analogous to the definition in Otto et al. (2018). For this model, the vector $\boldsymbol{h}_O$ is given by

$$\boldsymbol{h}_O = \alpha\mathbf{1} + \rho\mathbf{W}\operatorname{diag}(\boldsymbol{Y})\boldsymbol{Y}\,, \tag{2}$$

where $\operatorname{diag}(\boldsymbol{a})$ is a diagonal matrix with the entries of $\boldsymbol{a}$ on the diagonal. In order to be consistent with the implementation in the R-package **spGARCH**, we focus on the special case with two parameters α and ρ, whereas Otto et al. (2018) proposed a more general model with a vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)'$ and the first-order spatial lag $\mathbf{W}\operatorname{diag}(\boldsymbol{Y})\boldsymbol{Y}$.

For this definition, there is a one-to-one relation between $\boldsymbol{Y}$ and $\boldsymbol{\varepsilon}$ via the squared observations $\boldsymbol{Y}^{(2)} = (Y(\boldsymbol{s}_1)^2, \ldots, Y(\boldsymbol{s}_n)^2)'$ and squared errors $\boldsymbol{\varepsilon}^{(2)} = (\varepsilon(\boldsymbol{s}_1)^2, \ldots, \varepsilon(\boldsymbol{s}_n)^2)'$ with

$$\boldsymbol{Y}^{(2)} = \alpha\, (\mathbf{I} - \mathbf{A})^{-1} \boldsymbol{\varepsilon}^{(2)}\,, \tag{3}$$

where $\mathbf{W}$ is a predefined spatial weighting matrix and

$$\mathbf{A} = \rho \operatorname{diag}\left(\varepsilon(\boldsymbol{s}_1)^2, \ldots, \varepsilon(\boldsymbol{s}_n)^2\right)\mathbf{W}\,.$$

Thus,

$$\boldsymbol{h}_O = \alpha\mathbf{1} + \rho\alpha\mathbf{W}\,(\mathbf{I} - \mathbf{A})^{-1}\boldsymbol{\varepsilon}^{(2)}\,.$$

It is important to assume that the spatial weighting matrix is a non-stochastic, positive matrix with zeros on the main diagonal to ensure that a location is not influenced by itself (cf. Elhorst 2010; Cressie and Wikle 2011). The vector of random errors is denoted by $\boldsymbol{\varepsilon}$. Due to the complex dependence implied by the weighting matrix $\mathbf{W}$, $\boldsymbol{h}_O$ is not necessarily positive; thus, $\operatorname{diag}(\boldsymbol{h})^{1/2}$ does not necessarily have a solution in the real numbers such that the process in (1) is well-defined. This is only the case if the condition of the following lemma is fulfilled.

Lemma 1 (Otto et al. 2018)**.** *Suppose that $\alpha \geq 0$, $\rho \geq 0$ and that $\det(\mathbf{I} - \mathbf{A}^2) \neq 0$. If all elements of the matrix*

$$(\mathbf{I} - \mathbf{A}^2)^{-1} \tag{4}$$

are nonnegative, then all components of $\boldsymbol{Y}^{(2)}$ are nonnegative, i.e., $Y(\boldsymbol{s}_i)^2 \geq 0$ for $i = 1, \ldots, n$. Moreover, $h_O(\boldsymbol{s}_i) \geq 0$ for $i = 1, \ldots, n$.

It is important to note that $\mathbf{A}$ depends on both the weighting matrix and the realizations of the errors. In order to ensure that this condition is fulfilled, Otto et al. (2018) propose to truncate the support of the error distribution on the interval $(-a, a)$ with

$$a = \left\{ \begin{array}{cc} \infty & \exists k > 0 : \rho \mathbf{W}^k = \mathbf{0} \\ 1/\sqrt[4]{\rho^2 ||\mathbf{W}^2||_1} & \rho^2 ||\mathbf{W}^2||_1 > 0 \end{array} \right. ,$$

where $|| \cdot ||_1$ denotes the matrix norm based on the Manhattan norm.

There are two cases in which the support of the errors does not need to be constrained. If $\rho = 0$, the process coincides with a spatial white noise process such that a equals ∞. Moreover, all entries of $\boldsymbol{h}$ are non-negative if $\mathbf{W}$ is similar to a strictly triangular matrix. Then, $\mathbf{W}$ is nilpotent. This case covers the classical time-series ARCH(p) models introduced by Engle (1982) as well as the so-called oriented spARCH processes. For these processes, the spatial dependence has a certain direction, e.g., observations are only influenced by observations in a southward direction or by observations which are closer to an arbitrarily chosen center. The setting also covers recent time-series GARCH models incorporating spatial information (e.g., Borovkova and Lopuhaa 2012; Caporin and Paruolo 2006).

Of course, the truncated support of the errors has an impact on the extent of the spatial dependence on the conditional variances. Obviously, the support need not be constrained regarding $\rho = 0$. However, this support decreases with increasing values of ρ. For instance, if $\rho = 1$, then the parameter a is equal to 0.968 for Rook's contiguity matrices on a two-dimensional lattice. As a measure of the spatial dependence of the variance, one might consider Moran's I for the squared observations (see Moran 1950). Moreover, we observe that the growth rate of I decreases with increasing spatial weights. This trend can be explained by the compact support of the errors. Since there cannot be large variations $\varepsilon(\boldsymbol{s}_i)$ in absolute terms, there also cannot be large spatial clusters of high or low variance. To illustrate this behavior, Figure 1 depicts Moran's I for simulated observations $\boldsymbol{Y}$ and their squares for $\rho \in \{0, 0.05, \ldots, 2\}$. For the Monte Carlo simulation study, we simulate $n = 400$ observation on a two-dimensional lattice $D = \{\boldsymbol{s} = (s_1, s_2)' \in \mathbb{Z}^2 : 0 \leq s_1, s_2 \leq 20\}$. The weighting matrix is a common Rook's contiguity matrix, and the simulation is done for 10^5 replications. Although the exact distribution of Moran's statistic is bounded, the standardized statistic is asymptotically normally distributed for the "majority of spatial structures" (Tiefelsdorf and Boots 1995, see also Cliff and Ord 1981). Thus, the asymptotic 95% confidence intervals are plotted in Figure 1, as well.

Spatial Log-ARCH model

Next, we consider an logarithmic spatial ARCH process (log-spARCH). In this setting, we define the natural logarithm of $\boldsymbol{h}_E = (h_E(\boldsymbol{s}_1), \ldots, h_E(\boldsymbol{s}_n))'$ as

$$\ln \boldsymbol{h}_E = \alpha \mathbf{1} + \rho \mathbf{W} g_b(\boldsymbol{\varepsilon}) \,, \tag{5}$$

with a function $g_b : \mathbb{R}^n \to \mathbb{R}^n$. Like Nelson (1991), we assume that

$$g_b(\boldsymbol{\varepsilon}) = (\ln |\varepsilon(\boldsymbol{s}_1)|^b, \ldots, \ln |\varepsilon(\boldsymbol{s}_n)|^b)'$$

for positive values of b. For this definition, there is a one-to-one relation between $\boldsymbol{Y}$ and $\boldsymbol{\varepsilon}$, as we show in the following theorem.

Theorem 2. *Suppose that $\alpha > 0$, $\rho \geq 0$, and $w_{ij} \geq 0$ for all $i, j = 1, \ldots, n$ and $g_b(\boldsymbol{\varepsilon}) = (\ln |\varepsilon(\boldsymbol{s}_1)|^b, \ldots, \ln |\varepsilon(\boldsymbol{s}_n)|^b)'$. Then there exists one and only one $Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)$ that corresponds to each $\varepsilon(\boldsymbol{s}_1), \ldots, \varepsilon(\boldsymbol{s}_n)$ for $b > 0$.*

At location $\boldsymbol{s}_i$, the value of $h_E(\boldsymbol{s}_i)$ is then given by

$$\ln h_E(\boldsymbol{s}_i) = \alpha + \sum_{v=1}^{n} \rho b w_{iv} \ln |\varepsilon(\boldsymbol{s}_v)| \text{ for } i = 1, \ldots, n \,.$$

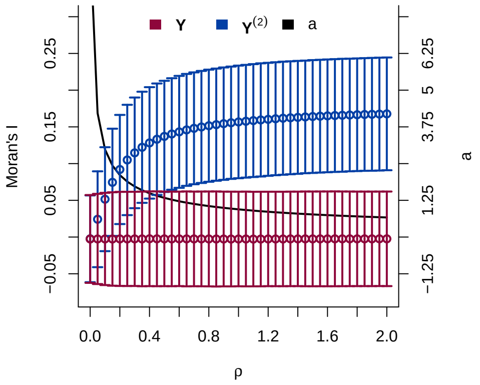


Figure 1: Moran's I of the observations $\boldsymbol{Y}$ and the squared observations $\boldsymbol{Y}^{(2)}$, including the asymptotic 95% confidence intervals of I for $\rho \in \{0, 0.05, \dots, 2\}$. The resulting bound a is plotted as a bold, black line.

For this definition of g_b, one could rewrite $\ln \boldsymbol{h}$ as

$$\ln \boldsymbol{h}_E = \mathbf{S}\left(\alpha \mathbf{1} + \rho b \mathbf{W} \ln |\boldsymbol{Y}|\right) \tag{6}$$

with

$$\mathbf{S} = (s_{ij})_{i,j=1,\dots,n} = \left(\mathbf{I} + \frac{1}{2}\rho b \mathbf{W}\right)^{-1}.$$

In contrast to the spARCH process described in Section Spatial ARCH model, Corollary 1 shows that the entries of $\boldsymbol{h}_E$ are positive for all $\rho \geq 0$ and $\alpha > 0$. Hence, the process is well-defined and there are no further restrictions needed, as in the case for the spARCH model.

Corollary 1. *Assume that the assumptions of Theorem 2 are fulfilled, then $\boldsymbol{h}_E(\boldsymbol{s}_i) \geq 0$ for all $i = 1, \dots, n$.*

For all proofs, we refer to the Appendix.

Complex Spatial ARCH model

Now, we propose a complex-valued spARCH process. In order to obtain a solution of $\text{diag}(\boldsymbol{h})^{1/2}$ in the n-dimensional space of real numbers for the model defined in (2), all elements of the matrix $(\mathbf{I} - \mathbf{A}^2)^{-1}$ must be nonnegative (see Otto et al. 2018). For the complex spARCH process, we relax the assumption that there should be a solution to $\text{diag}(\boldsymbol{h})^{1/2}$ in the real numbers and also consider complex solutions. Thus, the definition of $\boldsymbol{h}$ coincides with $\boldsymbol{h}_O$ of the original model, i.e.,

$$h_C(\boldsymbol{s}_i) = \alpha + \sum_{v=1}^{n} \rho w_{iv} Y(\boldsymbol{s}_v)^2 . \tag{7}$$

Spatiotemporal ARCH model

Finally, we show that spatiotemporal processes are covered directly by these approaches. For spatiotemporal data, the vector $\boldsymbol{s}$ simply includes both the spatial location $\boldsymbol{s}_s$ and the point in time t, i.e., $\boldsymbol{s} = (\boldsymbol{s}_s, t)'$. In addition, it is important to assume that future observations do not influence past observations, i.e., the weights w_{ij} must be zero if $t_j \geq t_i$. However, the dimension of the weighting matrix $\mathbf{W}$ might become very large for this representation. More precisely, the matrix has dimension $NT \times NT$, where N is the total number of spatial locations and T stands for the total number of time points. From a computational perspective, this is not necessarily a drawback since $\mathbf{W}$ is usually sparse and could also have a block diagonal structure. Moreover, it is often reasonable to assume that $h(\boldsymbol{s}_i)$ is only influenced by the neighbors of $\boldsymbol{s}_{s,i}$ at the same point of time and by past observations at the same

Process type	Definition of $\boldsymbol{h}$	Comments
spARCH	$\boldsymbol{h}_O = \alpha\mathbf{1} + \rho\mathbf{W}\left(\mathbf{I}-\mathbf{A}\right)^{-1}\left(\alpha\boldsymbol{\varepsilon}^{(2)}\right)$	$\boldsymbol{\varepsilon}$ is simulated from multivariate normal distribution (MN) truncated on the interval $\left[-1/\sqrt[4]{\|\|\rho^2\mathbf{W}^2\|\|_1}, 1/\sqrt[4]{\|\|\rho^2\mathbf{W}^2\|\|_1}\right]$
spARCH (oriented)	$\boldsymbol{h}_O = \alpha\mathbf{1} + \rho\mathbf{W}\left(\mathbf{I}-\mathbf{A}\right)^{-1}\left(\alpha\boldsymbol{\varepsilon}^{(2)}\right)$	$\boldsymbol{\varepsilon} \sim \mathrm{MN}(\mathbf{0}, \mathbf{I})$, $\mathbf{W}$ must be a strictly triangular weighting matrix
spatial log-ARCH	$\ln \boldsymbol{h}_E = \mathbf{S}\left(\alpha\mathbf{1} + \rho b\mathbf{W}\ln\|\boldsymbol{Y}\|\right)$	$\boldsymbol{\varepsilon} \sim \mathrm{MN}(\mathbf{0}, \mathbf{I})$, but moments of $\boldsymbol{Y}$ differ from the moments of classical spARCH process (cf. Otto et al. 2018)
spARCH (complex)	$\boldsymbol{h}_C = \alpha\mathbf{1} + \rho\mathbf{W}\left(\mathbf{I}-\mathbf{A}\right)^{-1}\left(\alpha\boldsymbol{\varepsilon}^{(2)}\right)$	$\boldsymbol{\varepsilon} \sim \mathrm{MN}(\mathbf{0}, \mathbf{I})$, but complex-valued $\boldsymbol{Y}$

Table 1: Overview of all types of spARCH models implemented in the **spGARCH** package.

location. Then the weighting matrix would have the following structure

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{I} & \mathbf{W}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_T \end{pmatrix}.$$

Indeed, it is plausible to weight the spatial and temporal lags differently by replacing $\rho\mathbf{W}$ by a sum

$$\rho \begin{pmatrix} \mathbf{W}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_T \end{pmatrix} + \phi_1 \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} + \ldots$$

with positive weights ϕ_k for all temporal lags $1 \leq k \leq p$.

Spatial ARCH Disturbances

Since all conditional and unconditional odd moments of spatial ARCH processes are equal to zero, these ARCH-type models can easily be added to any kind of (spatial) regression model without influencing the mean equation as well as the spatial dependence in the first conditional and unconditional moments. This makes the spatial ARCH models flexible tools for dealing with conditional spatial heteroscedasticity in the residuals of spatial models. For instance, one can consider spatial autoregressive models for $\boldsymbol{Y}$, i.e.,

$$\boldsymbol{Y} = \lambda\mathbf{B}\boldsymbol{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u} \tag{8}$$

with $\boldsymbol{u}$ following either a spatial ARCH model with the original definition $\boldsymbol{h}_O$ or the logarithmic model with $\boldsymbol{h}_E$. Thus,

$$\boldsymbol{u} = \mathrm{diag}(\boldsymbol{h})^{1/2}\boldsymbol{\varepsilon}\,. \tag{9}$$

Further, we call this model the SARspARCH model. For $\lambda = 0$, the model collapses to a simple linear regression model; if, additionally, $\boldsymbol{\beta} = \mathbf{0}$, the model coincides with the previously discussed ARCH models. Thus, these coefficients can be used for testing against nested models.

In contrast to other models for heteroscedastic errors, such as the SARAR or SARMA models, which assume spatial autoregressive or spatial moving average error terms (cf. Kelejian and Prucha 2010; Fingleton 2008; Haining 1978), the SARspARCH model does not affect the spatial autocorrelation of the process, just the spatial heteroscedasticity, because all conditional and unconditional odd moments are equal to zero. Thus, $\lambda\mathbf{B}$ can be interpreted directly as the spatial dependence of the process, while $\rho\mathbf{W}$ describes the spatial dependence in the second conditional moments. Moreover, these two parts can be interpreted separately, as we will demonstrate in the last section via an empirical example.

Generalized Spatial ARCH Models

Additionally, one may include spatially lagged observations of $\boldsymbol{h}$ to construct spatial GARCH-type models. For instance, a spatial GARCH model is given by

$$\begin{aligned} \boldsymbol{h}_G &= \alpha\mathbf{1} + \rho\mathbf{W}\mathrm{diag}(\boldsymbol{Y})\boldsymbol{Y} + \lambda\tilde{\mathbf{W}}\boldsymbol{h}_G \qquad (10)\\ &= (\mathbf{I} - \lambda\tilde{\mathbf{W}})^{-1}\left(\alpha\mathbf{1} + \rho\mathbf{W}\mathrm{diag}(\boldsymbol{Y})\boldsymbol{Y}\right), \qquad (11) \end{aligned}$$

where $\tilde{\mathbf{W}}$ is a second spatial weighting matrix and λ is the corresponding spatial GARCH parameter. Obviously, the spatial GARCH-type models requires that $(\mathbf{I} - \lambda\tilde{\mathbf{W}})$ is non-singular. In a similar manner, $\boldsymbol{h}_{LG}$ can be specified as

$$\ln \boldsymbol{h}_{LG} = \alpha\mathbf{1} + \rho\mathbf{W}g_b(\boldsymbol{\varepsilon}) + \lambda\tilde{\mathbf{W}}\boldsymbol{h}_{LG}\,, \qquad (12)$$

to define a spatial log-GARCH model. For theoretical details of spatial GARCH-type models, we refer to Otto and Schmid (2019) introducing a unified spatial GARCH model covering various spatial ARCH and GARCH models. Moreover, Otto and Schmid (2019) introduce an exponential spatial GARCH model allowing for asymmetry via an alternative definition of g in (5). To be precise, g is given by

$$g(\boldsymbol{\varepsilon}) = (\Theta\varepsilon_1 + \zeta(|\varepsilon_1| - E(|\varepsilon_1|)), \ldots, \Theta\varepsilon_n + \zeta(|\varepsilon_n| - E(|\varepsilon_n|))'$$

for the exponential spatial GARCH model.

Parameter Estimation

The parameters of a spatial ARCH process can be estimated by the maximum-likelihood approach. To obtain the joint density for $\boldsymbol{Y} = k(\boldsymbol{\varepsilon})$, the Jacobian matrix of k^{-1} at the observed values $\boldsymbol{y}$ must be computed (e.g., Bickel and Doksum 2015). If $f_{\boldsymbol{\varepsilon}}$ is the distribution of the error process, then the joint density $f_{\boldsymbol{Y}}$ of $\boldsymbol{Y}$ is given by

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) &= f_{(Y(\boldsymbol{s}_1),\ldots,Y(\boldsymbol{s}_n))}(y_1,\ldots,y_n)\\ &= f_{\boldsymbol{\varepsilon}}\left(\frac{y_1}{\sqrt{h_1}},\ldots,\frac{y_n}{\sqrt{h_n}}\right)|\det\left(\left(\frac{\partial y_j/\sqrt{h_j}}{\partial y_i}\right)_{i,j=1,\ldots,n}\right)|\,. \qquad (13) \end{aligned}$$

If the residuals are additionally independent and identically distributed, the parameter estimates can be obtained from the maximization of the log-likelihood as follows

$$(\hat{\alpha},\hat{\rho}) = \underset{\alpha>0,\rho\geq 0}{\arg\max}\ \ln|\det\left(\left(\frac{\partial y_j/\sqrt{h_j}}{\partial y_i}\right)_{i,j=1,\ldots,n}\right)| + \sum_{i=1}^{n}\ln f_\varepsilon(y_i)\,.$$

The Jacobian matrix, of course, depends on the definition of $\boldsymbol{h}$. For the spARCH process, this Jacobian matrix can be specified as

$$\frac{\partial y_j/\sqrt{h_j}}{\partial y_i} = \begin{cases} 1/\sqrt{h_j} & \text{for} \quad i = j\\ -\frac{y_i y_j}{h_j^{3/2}}\rho w_{ji} & \text{for} \quad i \neq j \end{cases}.$$

In contrast, the Jacobian matrix for the log-spARCH process is slightly different, namely

$$\frac{\partial y_j/\sqrt{h_j}}{\partial y_i} = \begin{cases} 1/\sqrt{h_j} & \text{for} \quad i = j\\ -\frac{b y_j}{2 y_i h_j^{3/2}}\rho s_{ji} w_{ji} & \text{for} \quad i \neq j \end{cases}$$

with

$$h_j = \exp\left(\sum_{v=1}^{n} s_{jv}\left(\alpha + \rho w_{jv}\ln|y_v|\right)\right).$$

From a computational perspective, the computation of the log determinant of this matrix is feasible for large data sets. To be precise, the log-determinant is equal to

$$\ln|\det\left(\mathrm{diag}\left(\frac{h_1}{y_1^2},\ldots,\frac{h_n}{y_n^2}\right) - \rho\mathbf{W}'\right)| + \sum_{i=1}^{n}\ln\frac{y_i^2}{h_i^{3/2}}$$

for the spARCH process. Similarly, it is given by

$$\ln|\det\left(\operatorname{diag}\left(\frac{2h_1}{b},\ldots,\frac{2h_n}{b}\right)-\rho\mathbf{S}'\circ\mathbf{W}'\right)|+\sum_{i=1}^{n}\ln\frac{b}{2h_i^{3/2}}\,.$$

for the log-spARCH process, where $\circ$ stands for the Hadamard product.

In the **spGARCH** package, we implemented the iterative maximization algorithm with inequality constraints proposed by Ye (1988), which is implemented in the R-package **Rsolnp** (see Ghalanos and Theussl 2012). It is important to note that the log determinant of the Jacobian also depends on the parameters in such a way that it needs to be computed in each iteration (see, also, Theorem 13.7.3 of Harville (2008) for the computation of a determinant for the sum of a diagonal matrix and an arbitrary matrix), but $\mathbf{W}$, and therefore $\mathbf{S}\circ\mathbf{W}$, are usually sparse. Thus, the required time for the estimation of the parameters depends mainly on the dimension and sparsity of $\mathbf{W}$.

Certainly, the choice of the weighting matrices are an important design choice of the models, which has to be prespecified. However, the true structure of $\mathbf{W}$ is rarely known in practice. Moreover, all estimated parameters depend on the selection of this matrix. Hence, inference on these parameters and the coefficients itself must be interpreted based on the choice of $\mathbf{W}$. For empirical applications, one might gain insights on the structure of $\mathbf{W}$ by looking at spatial autocorrelation functions or variograms. It is worth noting that the observations are uncorrelated for spatial ARCH models, so one should also look at squared observations. Then, the weighting scheme is typically chosen from a set of candidate schemes by maximizing certain goodness-of-fit criteria, like information criteria or out-of-sample prediction errors. For instance, $\mathbf{W}$ could be chosen as contiguity matrix, i.e., two locations are connected having positive weights, if they share a common border or if their distance is less than a certain threshold. For instance, in studies in spatial econometrics or epidemiology, the spatial domain is often a set of municipalities or counties (e.g., Amin et al. 2014; Buettner 2003). In this case, contiguity matrices are straightforward and if these binary matrices are additionally row-standardized, $\mathbf{W}\operatorname{diag}(\boldsymbol{Y})\boldsymbol{Y}$ can be interpreted as average of the squared neighboring observations. Alternatively, $\mathbf{W}$ can be specified as k-nearest-neighbor matrix, i.e., only the k nearest locations get positive weights, or as inverse-distance matrix, i.e., the weight between two locations is based on the distance between these locations. Further choices of $\mathbf{W}$ are discussed by Otto et al. (2018). Finally, it is worthy to mention that the weights could also depend on exogenous variables or other factors. For instance, they could incorporate economic disparities, e.g., differences in the gross domestic products, poverty rates, household incomes etc., or other covariates, like the wind direction and speed when modeling spatial dependence of air pollutants (cf. Merk and Otto 2019). For spatiotemporal autoregressive processes, there are also some approaches to estimate the entire spatial dependence structure using machine learning methods (e.g., Lam and Souza 2016; Otto and Steinert 2018).

Overview of the R-Package spGARCH

The R-package **spGARCH** provides several basic functions for the analysis of spatial data showing spatial conditional heteroscedasticity. In particular, the process can be simulated for arbitrarily chosen weighting matrices according to the definitions in Section Spatial ARCH-type models. Moreover, we implement a function for the computation of the maximum-likelihood estimators. To generate a user-friendly output, the object generated by the estimation function can easily be summarized by the generic `summary()` function. We also provide all common generic methods, such as `plot()`, `print()`, `logLik()`, and so forth. To maximize the computational efficiency, the actual version of the package contains compiled C++ code (using the packages **Rcpp** and **RcppEigen**, cf. Eddelbuettel and François 2011; Bates and Eddelbuettel 2013). A brief overview of the package and its main functions is given in Table 2. Further, we focus on the two main aspects of the package, i.e., the simulation (described in detail in Section Simulation of ARCH-type stochastic processes) and estimation (Section Maximum-likelihood estimation) aspects of the spARCH, log-spARCH, and SARspARCH processes.

Simulation of ARCH-type stochastic processes

The simulations of all spatial ARCH-type models are implemented in one function, namely, the `sim.spARCH()` function. The different definitions of the model are specified via the argument `type`. The use of `sim.spARCH()` is very similar to how a basic random number generator is used, meaning that the first argument `n` is the number of generated values and all further arguments specify the parameters of the spARCH process. For instance, one might simulate an oriented spARCH process (meaning $\mathbf{W}$ is triangular) on a $d \times d$ spatial lattice with $\rho = 0.7$ and $\alpha = 1$ using the following lines.

```
R> require("spdep")
R> rho             <- 0.7
```

Function	Description
Main functions	
`sim.spARCH()`	Simulation of spARCH and log-spARCH processes
`sim.spGARCH()`	Simulation of spGARCH, E-spGARCH, and log-spGARCH processes
`qml.spARCH()`	Quasi-maximum-likelihood estimation for spARCH models
`qml.SARspARCH()`	Quasi-maximum-likelihood estimation for SAR models with spARCH residuals
Generic methods	
`summary()`	Summary of an object of '`spARCH`' class generated by `qml.spARCH()` or `qml.SARspARCH()`
`print()`	Printing method for '`spARCH`' class or `summary.spARCH` class
`fitted()`	Extracts the fitted values of an object of '`spARCH`' class
`residuals()`	Extracts the residuals of an object of '`spARCH`' class
`logLik()`	Extracts the log-likelihood of an object of '`spARCH`' class
`extractAIC()`	Extracts the AIC of an object of '`spARCH`' class
`plot()`	Provides several descriptive plots of the residuals of an object of '`spARCH`' class

Table 2: Summary of the main functions of the **spGARCH** package.

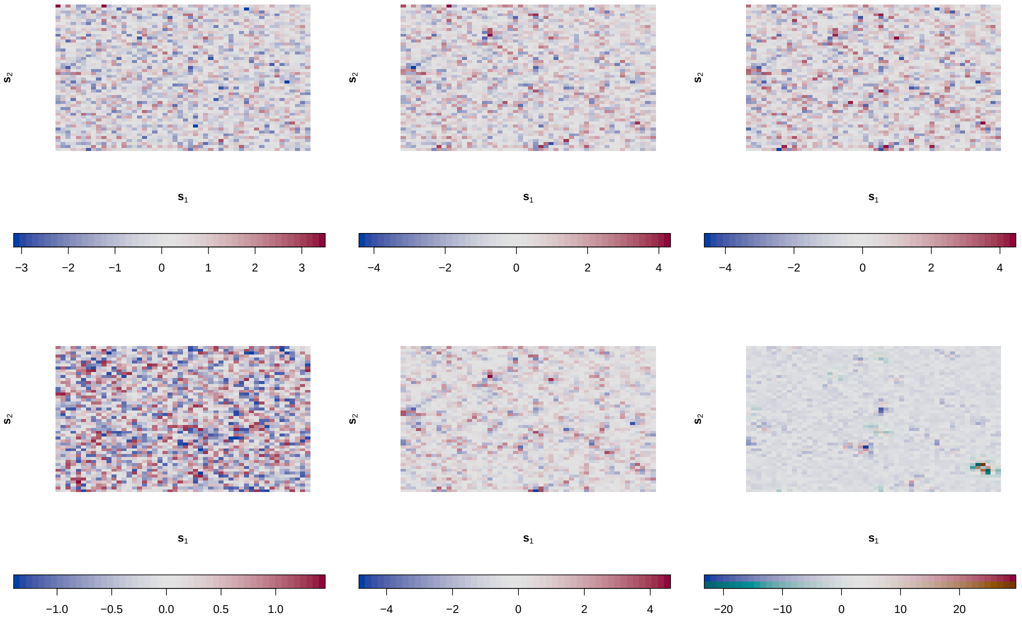
```
R> alpha          <- 1
R> d              <- 50
R> n              <- d^2
R> nblist         <- cell2nb(d, d, type = "queen")
R> W              <- nb2mat(nblist)
R> W[upper.tri(W)] <- 0
R> Y              <- sim.spARCH(n = n, rho = rho, alpha = alpha, W = W,
+                              type = "spARCH", control = list(seed = 5515))
```

To build the spatial weighting matrix, we used `cell2nb()` from the **spdep** package, returning an `nb` object of a $d \times d$ lattice (see Cressie 1993; Bivand and Piras 2015). Further, we converted the `nb` object into a contiguity matrix, as `sim.spARCH()` requires either a matrix (class `matrix`) or a sparse matrix (class `dgCMatrix`) as an argument. Usually, spatial weighting matrices are sparse by construction. Thus, $\mathbf{W}$ is always converted internally to a `dgCMatrix` matrix or rather to a `SparseMatrix` object defined in the eigen library in C++. Via the `control` parameter, a random seed might be passed to the simulation function. If not, a random seed is assigned randomly from a uniform distribution and printed in console in order that one might reproduce the result even without having a random seed specified in advance. We prefer to print a single number in the console rather than returning to the random number generator (RNG) state as an attribute of the returned vector. Thus, a random seed might either be passed as an optional argument to `sim.spARCH()` or set before calling `sim.spARCH()` by `set.seed()`.

There are several types of spatial ARCH processes which can be simulated by `sim.spARCH()`. They are all specified by the argument `type`. If

- `type = "spARCH"`, then the original spARCH process according to the definition in Otto et al. (2018) is simulated.
  - If there exists a permutation such that $\mathbf{W}$ is a strictly triangular matrix, then the function simulates automatically an oriented spARCH process with independent and identically gaussian distributed errors.
  - If there is no such permutation, then the errors are simulated from a truncated normal distribution with $a = 1/\sqrt[4]{\rho^2||\mathbf{W}||_1}$.
- `type = "log-spARCH"`, an log-spARCH process is simulated with an user-specified value of b (default 2) and standard normal random errors.
- `type = "complex-spARCH"`, complex solutions of $\mathrm{diag}(\boldsymbol{h})^{1/2}$ are considered in order to simulate the spARCH process.

Figure 2 illustrates the behavior of different types of spatial ARCH processes. All of them are simulated with the same parameters and random seeds in such a manner that the vector $\boldsymbol{\varepsilon}$ is identical for all types of processes, except for the spARCH process with the truncated normal errors. In the first row, the spatial weighting is achieved via a strictly triangular Queen's contiguity matrix, which



Above left: spatial white noise for comparison; center: oriented spARCH (type = ``gaussian''); right: spatial E-ARCH (type = ``exp'').

Below left: spARCH with truncated normal errors (type = ``gaussian''); center: spatial E-ARCH (type = ``exp''), right: complex spARCH (type = ``complex'').

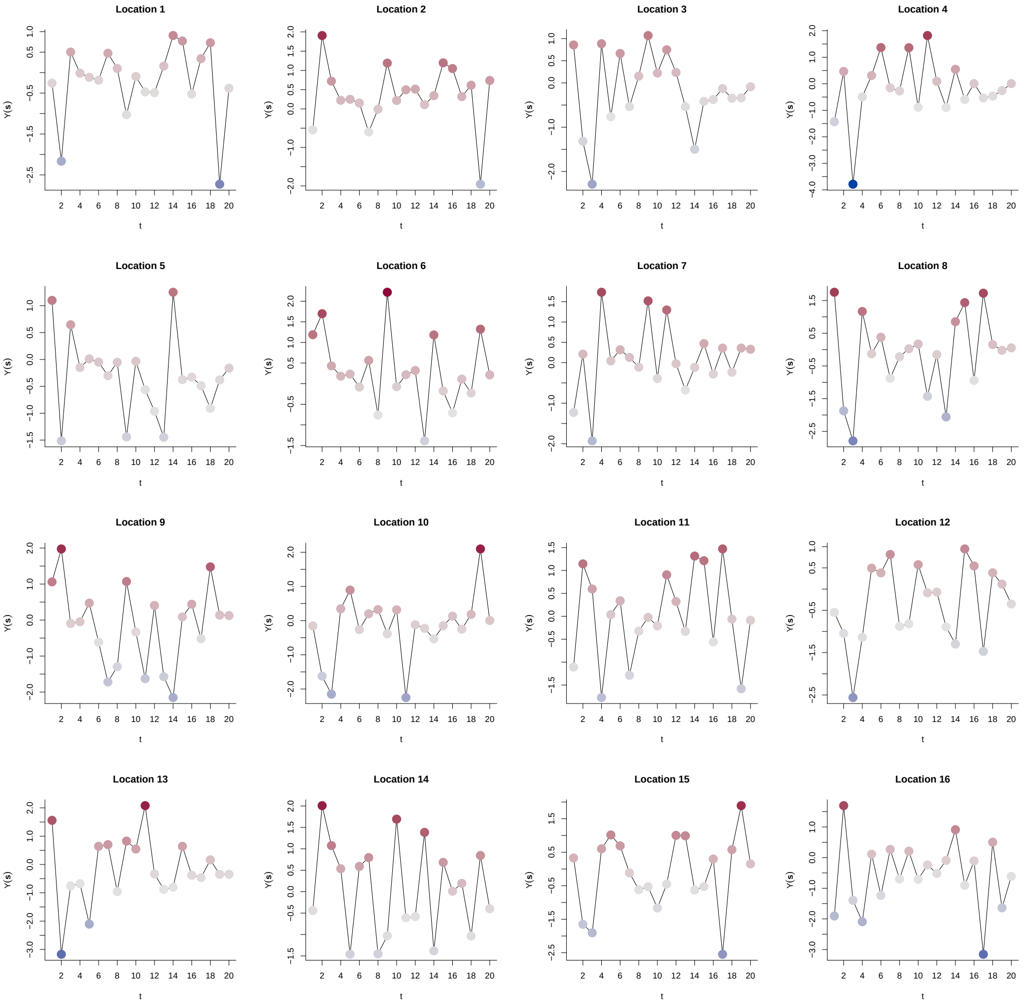
Figure 2: Simulations on a two-dimensional lattice for triangular matrices (above) and non-triangular matrices (below). For all simulations, we set $\rho = 0.7$ and $\alpha = 1$, and $\mathbf{W}$ is chosen to be the Queen's contiguity matrix.

means that the spatial dependence has its origin in the upper left corner. To the contrary, $\mathbf{W}$ presents a classical Queen's contiguity matrix in the second row. We additionally plot a spatial white noise process for comparison, as we used a rather unconventional two-color scheme. Using this kind of color scheme, one might distinguish between positive and negative observations, such that it is easier to see the spatial volatility clusters. Areas of smaller volatility are characterized by rather evenly gray pixels, whereas clusters of high volatility have rather intense colors. Moreover, the colors fluctuate irregularly between blue and red.

As pointed out in Section Spatial ARCH-type models, spatiotemporal ARCH models are directly covered if time is considered as one dimension of the q-dimensional space D. Thus, a two-dimensional spatiotemporal process $\{Y_t(s) : t = 1, \ldots, T; s \in D_s\}$ with D_s being a $d \times d$ unit grid and $T = 20$ points of time could be simulated as follows.

```
R> d      <- 4
R> T      <- 20
R> D_s    <- 1:(d^2)
R> D_t    <- 1:T
R> n      <- length(D_s) * length(D_t)
R> nblist <- cell2nb(d, d, type = "queen")
R> W_list <- nb2listw(nblist)
R> W_s    <- Matrix(listw2mat(W_list))
R> W      <- W_s
R> for(t in D_t[-1]){
R>   W    <- bdiag(W, W_s)
R> }
R> diag(W[-c(1:length(D_s)), -c((n - length(D_s)+1):n)]) <- 0.2
R> set.seed(1)
R> Y      <- sim.spARCH(n = n, rho = 0.8, alpha = 1, W = W, type = "log-spARCH")
```

The spatial weighting scheme has been chosen as block diagonal matrix like proposed above, i.e., constant matrices $\mathbf{W}_1 = \mathbf{W}_2 = \ldots = \mathbf{W}_T = \mathbf{W}_s$ along the diagonal define the instantaneous



Note that the true spatial orientation is preserved in this representation (4 × 4 grid), i.e., plots appearing close to each other are also located close to each other in space and they are, therefore, more related than more distant locations. An alternative representation as consecutive spatial random fields is shown in Figure 4.

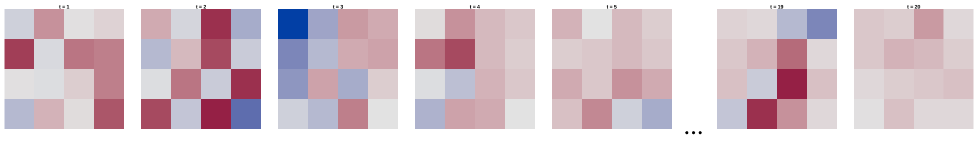
Figure 3: Simulated spatiotemporal log-ARCH process depicted as individual time series. The process has been simulated on a 4×4 spatial unit grid and for 20 points in time. The spatial ARCH parameter ρ equals 0.9 and $\alpha = 1$.

spatial interactions, while constant weights of 0.2 below the diagonal define the extend of the temporal dependence. For instance, a central location would be weighted equally by all eight spatial neighbors with a weight of 0.125 and by the observation of the same location at the previous time point by 0.2. Thus, **W** describes the structure and the extend of the dependence in space and time. Eventually, ρ has been chosen as 0.8 and we simulated the process as logarithmic spatial ARCH process.

The resulting simulation is depicted in Figures 3 and 4. Whereas the simulated values are shown as time series plots placed at their correct spatial locations in Figure 3, 4 depicts the observations as consecutive spatial random fields. Note that the same color coding has been chosen for both representations. On the one hand side, one can observe spatial volatility clusters (e.g. in the pre-last plot in Figure 4, $t = 19$, in which the conditional variance is low in the upper left corner, whereas the conditional variance of the remaining locations is high). On the other hand, temporal volatility cluster can be observed as well. For example, at location 16, the variance is high at the first and last five time points, while it is lower between $t = 6$ and $t = 14$.

For sake of completeness, we briefly demonstrate the simulation of spatial GARCH-type models using the `sim.spGARCH()` function. Like for `sim.spARCH()`, the type of the spatial GARCH model can be chosen by the argument `type`. More precisely, there are the following options

- `type = "spGARCH"` for simulation of spatial GARCH models according to the definition in (10),
- `type = "e-spGARCH"` for simulation of exponential spatial GARCH models according to the



The first five and the last two time points are plotted as spatial random fields, i.e., the simulations are shown in their natural temporal ordering. An alternative representation as time series in their true spatial ordering is shown in Figure 3.

Figure 4: Simulated spatiotemporal log-ARCH process depicted as consecutive spatial random fields. The process has been simulated on a 4 × 4 spatial unit grid and for 20 points in time. The spatial ARCH parameter ρ equals 0.9 and $\alpha = 1$.

definition in (12) with g,

- `type = "log-spGARCH"` for simulation of logarithmic spatial GARCH models according to the definition in (12) with g_b, and
- `type = "complex-spGARCH"` for simulation of a complex-valued spatial GARCH model.

To simulate a spatial GARCH process, two spatial weights matrices need to be specified via the arguments `W1` and `W2`. Moreover, two parameters ρ and λ are passed to the simulation function by the arguments `rho` and `lambda`. For instance, a spatial GARCH model can be simulated on a $d \times d$ spatial unit grid as follows

```
R> require("spdep")
R> rho    <- 0.5
R> lambda <- 0.3
R> alpha  <- 1
R> d      <- 20
R> nblist <- cell2nb(d, d, type = "rook") # Rook's contiguity matrix
R> W_1    <- nb2mat(nblist)
R> W_2    <- W_1
R> Y      <- sim.spGARCH(rho = rho, lambda = lambda, alpha = alpha,
+                        W1 = W_1, W2 = W_2, type = "spGARCH")
```

Similarly, spatial log-GARCH processes and exponential spatial GARCH processes can be simulated by changing the argument `type`. In this case, the parameters b must be provided for the log-GARCH or Θ and ζ for the e-spGARCH, respectively. These parameters can easily passed to `sim.spGARCH()` by the arguments `b`, `theta`, and `zeta`.

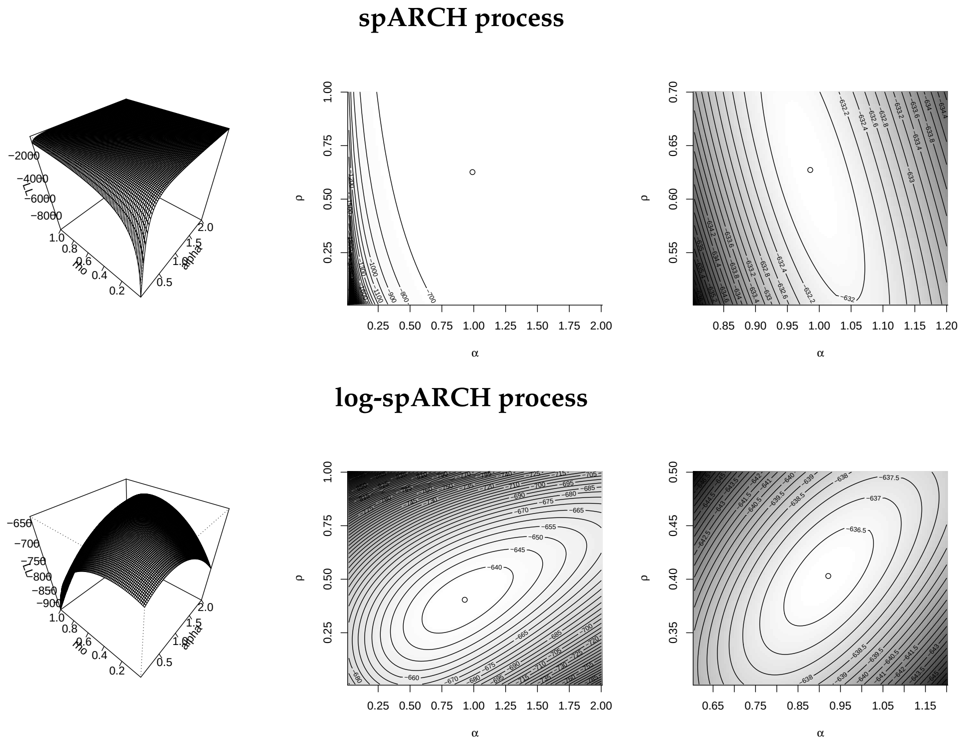
Maximum-likelihood estimation

Other important functions of the package are the `qml.spARCH()` and `qml.SARspARCH()` functions, which implement a quasi-maximum-likelihood estimation algorithm (QML). As for the `sim.spARCH()` function, many spARCH models are covered in the `qml.spARCH()` and `qml.SARspARCH()` function. Thus, the user needs to specify which particular spARCH model is to be fitted via the argument `type`. Moreover, the model for the mean equation is a user-specified `formula`, making the use of the estimation functions similar to the use of the common `lm()` or `glm()` functions.

In general, the estimators exhibited good performances for a variety of error distributions in simulation studies, although the likelihood function was derived under the normality assumption. This is not surprising, as the maximum-likelihood estimators have good properties under mild assumptions for the error processes of a variety of similar spatial econometrics models (cf. Lee 2004; Lee and Yu 2012, 2010b,a). Thus, we refer to the approach as the QML approach, and the name of the estimation functions start with `qml` instead of `ml`. In the following paragraphs, we start the simulation of one specific sample, which is then used further to illustrate the log-likelihood functions as well as to demonstrate parameter estimation.

Compared to the log-spARCH processes, the likelihood functions of spARCH models are rather flat around the global maximum. This behavior is illustrated for simulated processes in Figure 5. The observations for the log-spARCH process have been simulated as follows.

```
R> nblist          <- cell2nb(20, 20, type = "queen")
R> W               <- nb2mat(nblist)
```



spARCH process

log-spARCH process

Figure 5: Logarithmic likelihood function.

```
R> y              <- sim.spARCH(n = 20^2, rho = 0.5, alpha = 1, W = W,
+                              type = "log-spARCH", control = list(seed = 5515))
```

To simulate an oriented process, the entries of **W** above the diagonal must be set to zero and the argument `type` must be changed to `"spARCH"`, i.e.,

```
R> W[upper.tri(W)] <- 0
R> y2             <- sim.spARCH(n = 20^2, rho = 0.5, alpha = 1, W = W,
+                                 type = "spARCH",
+                                 control = list(seed = 5515))
```

To estimate the parameters of an intercept-free log-spARCH model without any regressors, the formula passed to the function `qml.spARCH()` should be specified as `y  0`. In addition, a `data.frame` can be passed via the `data` argument to the `qml` functions. Although the likelihood function of a spARCH process is flat, good estimates can be obtained through iterative maximization. Otto et al. (2018) analyze the performance of the estimators in detail. The algorithm implemented in the packages is based on the **Rsolnp** package, allowing for both equality and inequality parameter constraints (cf. Ghalanos and Theussl 2012).

The results of the estimation procedure are returned via an object of the class '`spARCH`', for which we provide additionally several generic functions. First, there is a `summary()` function for the '`spARCH`' object. The summary shows all important estimation results, i.e., the parameter estimates, standard errors, test statistics, and asymptotic p-values, including significance stars. The estimation of the above simulated log-spARCH process would return the following results.

```
R> spARCH_object   <- qml.spARCH(y ~ 0, W = W, type = "log-spARCH")
R> summary(spARCH_object)
 Call:
qml.spARCH(formula = y ~ 0, W = W, type = "log-spARCH")

 Residuals:
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-2.6867629 -0.6197315 -0.0053580 -0.0002615  0.5708346  2.8576621

 Coefficients:
      Estimate Std. Error t value  Pr(>|t|)
alpha 0.919324   0.128544  7.1518 8.564e-13 ***
```

```
rho   0.402998   0.056519  7.1304 1.001e-12 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

 AIC: 543.01, BIC: 539.01 (Log-Likelihood: -269.51)

 Moran's I (residuals): -0.028568, p-value: 0.31795

 Moran's I (squared residuals): 0.035239, p-value: 0.14479
```

The standard errors are estimated as Cramer-Rao bounds from the Hessian matrix of the log-likelihood function. For triangular weighting matrices, the estimators are asymptotically normally distributed (Otto et al. 2018). In addition to the Akaike and Bayesian Schwarz information criteria, the results of Moran's test on the residuals and squared residuals are reported for the spatial autocorrelation of the residuals. However, it is possible to use functions like `AIC()` or `BIC()`, since there is a `logLik()` method for the objects from class 'spARCH'. Additionally, the fitted values and residuals can be extracted by `fitted()` and `residuals()`, respectively.

To analyze the residuals, we provide additionally several descriptive plots via the generic `plot()` function. The first two plots are produced by `moran.plot()` imported from the package **spdep**. They inspect the spatial autocorrelation of the residuals and the squared residuals. In addition, the error distribution is depicted in the third graphic by a normal Q-Q-plot. The output obtained for the above numerical example is given below and in Figure 6.

```
%\begin{CodeInput}
%R> AIC(spARCH_object)
%R> BIC(spARCH_object)
%R> par(mfcol = c(1,3))
%R> plot(spARCH_object)
%\end{CodeInput}
R> AIC(spARCH_object)
[1] 543.0126
R> BIC(spARCH_object)
[1] 550.9956
R> par(mfcol = c(1,3))
R> plot(spARCH_object)
Reproduce the results as follows:
         eps <- residuals(x)
         W <- as.matrix(x$W)
         moran.plot(eps, mat2listw(W), zero.policy = TRUE,
     xlab = "Residuals”, ylab = "Spatially Lagged Residuals”)
Reproduce the results as follows:
         eps <- residuals(x)
         W <- as.matrix(x$W)
         moran.plot(eps, mat2listw(W), zero.policy = TRUE,
     xlab = "Residuals”, ylab = "Spatially Lagged Residuals”)
Reproduce the results as follows:
         eps <- residuals(x)
         std_eps <- (eps - mean(eps))/sd(eps)
         qqnorm(eps, ylab = "Standardized Residuals”)
         qqline(eps)
```

The mean equation can be specified as `formula` for all models, i.e., the spARCH, log-spARCH, and SARspARCH models. Thus, there is a huge variety of possible spatial ARCH models as well as regression models with spARCH residuals which can be fitted by the estimation functions. In addition to linear models of the form `y   a + b`, more sophisticated models can also be fitted, e.g., models with interactions `y   a + b:c`, factor models `y   factor`, polynomial models `y   poly(a,3)`, seasonally or regularly varying models of the form `y   sin(t) + cos(t)` or `y   sin(long) + cos(long) + sin(lat) + cos(lat)`, and so forth. We also implement an `extractAIC()` method for 'spARCH' objects, such that one might also use `step()` for stepwise model selection. Table 3 provides an overview of possible combinations of the arguments `formula` and `type` and shows the resulting models, which can be fitted by the functions `qml.spARCH()` and `qml.SARspARCH()`, respectively.

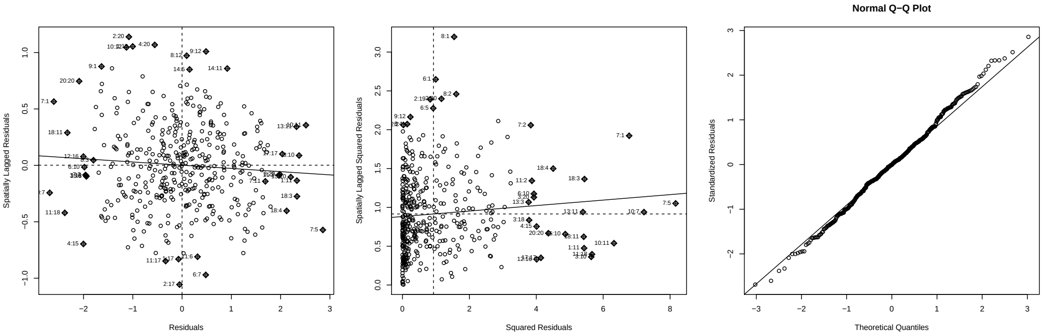


Figure 6: Resulting graphical output of `plot()`.

Function	`formula`	`type`	Resulting model
`qml.spARCH()`	`y 0`	`"spARCH"`	spARCH model (see (1) and (2))
`qml.spARCH()`	`y 1`	`"spARCH"`	spARCH model with an additional intercept for the mean equation
`qml.spARCH()`	`y a + b`	`"spARCH"`	Linear Regression with regressors `a` and `b` and spARCH residuals
`qml.spARCH()`	`y a + b:c`	`"spARCH"`	Linear Regression with more complex expressions and spARCH residuals
`qml.spARCH()`	`y 0`	`"log-spARCH"`	log-spARCH model (see (1) and (5))
`qml.spARCH()`	`y 1`	`"log-spARCH"`	log-spARCH model with an additional intercept for the mean equation
`qml.spARCH()`	`y a + b`	`"log-spARCH"`	Linear Regression with regressors `a` and `b` and log-spARCH residuals
`qml.spARCH()`	`y a + b:c`	`"log-spARCH"`	Linear Regression with more complex expressions and log-spARCH residuals
`qml.SARspARCH()`	`y 0`	`"spARCH"`	SAR model without an intercept, but with spARCH residuals (see (8) and (9))
`qml.SARspARCH()`	`y 1`	`"spARCH"`	SAR model with an intercept and spARCH residuals
`qml.SARspARCH()`	`y a + b`	`"spARCH"`	SAR model with an intercept and the regressors `a` and `b` and spARCH residuals
`qml.SARspARCH()`	`y a + b:c`	`"spARCH"`	SAR model with more complex expressions and spARCH residuals
`qml.SARspARCH()`	`y 0`	`"log-spARCH"`	SAR model without an intercept, but with log-spARCH residuals (see (8) and (9))
`qml.SARspARCH()`	`y 1`	`"log-spARCH"`	SAR model with an intercept and log-spARCH residuals
`qml.SARspARCH()`	`y a + b`	`"log-spARCH"`	SAR model with an intercept and the regressors `a` and `b` plus log-spARCH residuals
`qml.SARspARCH()`	`y a + b:c`	`"log-spARCH"`	SAR model with more complex expressions and log-spARCH residuals

Table 3: Overview of spatial models, which can be fitted by `qml.spARCH()` and `qml.SARspARCH()`.

Real-data example: prostate cancer incidence rates

Below, the focus is on the incidence rates (2008–2012) for prostate cancer provided by the Centers for Disease Control and Prevention (of Health et al. 2015). In particular, we analyze the incidence rates in all counties of several southeastern U.S. states, namely Arkansas, Louisiana, Mississippi, Tennessee, North and South Carolina, Georgia, Alabama, and Florida. This area also covers the counties along the Mississippi River collectively known as "cancer alley" (see Nitzkin 1992; Brent 2010; Berry 2003). All rates are age-adjusted to the 2000 U.S. standard population (cf. of Health et al. 2015). To reproduce the example, the logarithmic incidence rates as well as several covariates are included in the package.

As explanatory variables, we included a large set of environmental, climate, behavioral, and health covariates, which might have an influence on incidence rates for prostate cancer. For instance, we consider air pollution, such as $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, O_3, and $CH2O$, as potential environmental hazard factors. Moreover, we account for smoking, drinking, sport activities, and further healthcare-related variables as potential influences on the cancer incidence rates. In total, we account for 34 explanatory variables, which were obtained by inverse-distance-kriging from spatial points processes. Most of the variables are correlated, so we performed a factor analysis on 5 subgroups to identify 10 common factors. The factor loadings are summarized in Table 4. Note that the factor scores are directly included in the dataset `prostate_cancer`. Eventually, the final explanatory factors were chosen by minimizing the Bayesian information criterion using the generic function `step()` as follows.

```
R> data(prostate_cancer)
R> out <- step(qml.SARspARCH(formula, B = B, W = W, type = "spARCH",
+                            data = prostate_cancer), k = log(length(Y)))
```

The `formula` object simply defines a linear model between the logarithmic incidence rates and all factors. Further, matrix **B** describes the predefined spatial dependence structure in the mean equation. For this analysis, **B** has been chosen as a row-standardized contiguity matrix of the direct neighbors. For the spatial dependence in the spatial ARCH term of the residuals, we also included all neighbors up to order 4. Hence, **W** is the row-standardized matrix of the sum of the first-, second-, third-, and fourth-lag neighbors.

By minimizing the BIC criterion, the 2nd and 10th factor has been selected. Whereas the 2nd factor has positive loadings mainly for fine particulate matters, $PM_{2.5}$ and PM_{10}, the 10th describes the tendency for high blood pressure and cholesterol in the county's population. However, note that this analysis is based on aggregated data rather than individual patients; hence, the selected factors cannot be interpreted as carcinogenic factors.

Using the generic `summary()` for the '`spARCH`' class, the estimated model can be summarized as follows.

```
 Call:
qml.SARspARCH(formula = formula, B = B, W = W, type = "spARCH",
    data = prostate_cancer)

 Residuals:
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-0.7492270 -0.1079639 -0.0001509 -0.0005261  0.1121190  0.6404564

 Coefficients:
                Estimate Std. Error t value  Pr(>|t|)
alpha (spARCH)  0.0203839  0.0042674  4.7766 1.783e-06 ***
rho (spARCH)    0.3782104  0.1309656  2.8879  0.003879 **
lambda (SAR)    0.6768133  0.0356765 18.9708 < 2.2e-16 ***
(Intercept)     1.5388985  0.1702222  9.0405 < 2.2e-16 ***
F_2             0.0192857  0.0069917  2.7584  0.005809 **
F_10           -0.0205693  0.0064450 -3.1915  0.001415 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 AIC: -1734.2, BIC: -1746.2 (Log-Likelihood: 873.11)

 Moran's I (residuals): -0.022899, p-value: 0.32023

 Moran's I (squared residuals): 0.021409, p-value: 0.00050052
```

First, we see that the model has a significant spatial autocorrelation in the mean equation since $\hat{\lambda}$

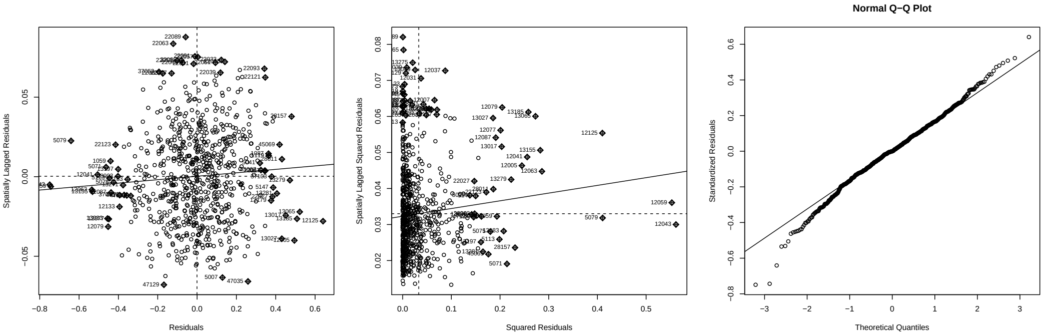


Figure 7: Resulting graphical output of `plot()` for the real-data example.

(`lambda (SAR)`) differs significantly from zero. This implies that there are clusters of higher prostate cancer incidence rates and, vice versa, lower incidence rates. Second, the error process shows conditional, autoregressive heteroscedasticity in space, which is captured by the spARCH component of the model, i.e., $\hat{\rho} = 0.378$ and $\hat{\alpha} = 0.020$. This can be interpreted as differences in the local uncertainty of the model. Hence, there are regions where the model predicts the true incidence rates more accurately, and there are regions with a worse fit. This can also be interpreted as local risks coming from unobserved, hidden factors. Note additionally that it is important to account for spatial conditional heteroscedasticity, as the estimates of spatial autoregressive models are biased if the error variance is not homogeneous across space. Inspecting the residuals, one can see that the spatial autocorrelation has been fully captured by the model, as Moran's I of the residuals is close to zero. In contrast, there is a weak spatial dependence in the squared residuals. To inspect the reason for this dependence graphically, the function `plot()` can be used to produce the plots shown in Figure 7.

After fitting the model, one also may include further regressors or estimate an intercept-only model via `update()`. For illustration, we added the percentage of positive results for a prostate-specific antigen (PSA) test in each county as an additional explanatory variable by

```
R> out2 <- update(out, . ~ . + PSA_test)
```

The PSA test is used for prostate cancer screening, meaning that there should definitely be a positive dependence between the PSA test and the incidence rates. In fact, the estimated parameter is positive, and the AIC is lower compared to the previous model. To be precise, the updated parameters are

```
                      Estimate Std. Error t value  Pr(>|t|)
alpha (spARCH)       0.0199281  0.0043105  4.6231  3.78e-06 ***
rho (spARCH)         0.3902185  0.1280266  3.0479 0.0023041 **
lambda (SAR)         0.6643605  0.0366748 18.1149 < 2.2e-16 ***
(Intercept)          1.1349551  0.2301554  4.9313  8.17e-07 ***
F_2                  0.0198504  0.0069903  2.8397 0.0045159 **
F_10                -0.0224035  0.0065828 -3.4034 0.0006656 ***
PSA_test             0.0095962  0.0042728  2.2459 0.0247125 *
```

Summary and discussion

This paper examines spatial models for autoregressive conditional heteroscedasticity. In contrast to previously proposed spatial GARCH models, these models allow for instantaneous autoregressive dependence in the second conditional moments. Previous approaches only allowed for spatial dependence in the first temporal lag. However, these models are also captured by the spatial ARCH approach, since temporal dependence can be included by appropriate choices of the weighting matrix. In addition to discussing previously proposed models, we introduced a novel spatial logarithmic ARCH model, for which the probability density has been derived and maximum-likelihood estimators discussed.

In addition to this theoretical model, we focus on the computational implementation of all considered spatial ARCH models in the R-package **spGARCH**. In particular, the simulation and estimation has been demonstrated. Regarding maximum-likelihood estimation, a broad range of spatial models are implemented in the package. Furthermore, the spatial weights matrices, as well as the mean model, can easily be specified by the user, providing a flexible and easy-to-use tool for spatial ARCH models. All estimation functions return an object for class '`spARCH`', for which several generic functions are provided, such as `summary()`, `plot()`, and `AIC()`. This setup also allows the use of the R-base functions,

	F. 1	F. 2	F. 3	F. 4	F. 5	F. 6	F. 7	F. 8	F. 9	F. 10
$PM_{2.5}$ concentration	0.69	0.72								
SO_2 concentration	0.33	-0.03								
NO_2 concentration	0.13	-0.12								
CO concentration	0.31	0.05								
PM_{10} concentration	0.07	0.44								
O_3 concentration	1.00	-0.02								
Solar radiation			0.60	0.44						
Precipitation			-0.08	-0.26						
Outdoor temperature			1.00	-0.05						
Temperature differences			0.32	0.94						
Ambient maximal temperature			0.08	-0.39						
CH_2O	-0.23	0.32								
Percentage of current smokers					0.47	-0.85				
Percentage of former smokers					0.92	0.37				
Smoke some days					-0.07	-0.62				
Never smoked					-0.96	0.25				
Aerobic activity							-0.05	0.58		
Exercises							0.41	0.33		
Physical activity index							-0.09	0.99		
Alcohol consumption					0.04	0.62				
Binge drinking					0.07	0.44				
Heavy drinking					0.43	0.02				
High cholesterol									0.00	1.00
Cholesterol checked									0.55	0.00
Overweight (BMI 25.0-29.9)									0.99	0.09
Obese (BMI 30.0 - 99.8)									-0.75	0.01
Blood stool test									0.56	-0.23
Sigmoidoscopy									0.14	-0.16
High blood pressure									0.03	0.79
Flu shot							0.81	-0.13		
Pneumonia vaccination							0.51	-0.26		
Health care coverage							0.58	0.18		
Seatbelt use							-0.58	0.10		

Table 4: Overview of all included regressors and factor loading for the 10 common factors. The regressors were divided into 5 subgroups to allow for distinctions between the factors.

such as `step()` for stepwise model selection or `update()` for updating the results of different mean models. Eventually, the use of these functions are demonstrated by an empirical example, namely county-level incidence rates of prostate cancer.

In the future, the package should be extended for further spatial ARCH-type models. Along this vein, a class for model specifications should be added alongside the actual implementations via arguments for the fitting functions. In that way, the package can be aligned to common time series ARCH packages, such as the **rugarch** package. Furthermore, the package could benefit from robust estimation methods, another focus for future research.

Bibliography

R. Amin, M. Hendryx, M. Shull, and A. Bohnert. A Cluster Analysis of Pediatric Cancer Incidence Rates in Florida: 2000–2010. *Statistics and Public Policy*, 1(1):69–77, 2014. [p]

D. Bates and D. Eddelbuettel. Fast and elegant numerical linear algebra using the RcppEigen package. *Journal of Statistical Software*, 52(5):1–24, 2013. URL http://www.jstatsoft.org/v52/i05/. [p]

G. R. Berry. Organizing against multinational corporate power in cancer alley the activist community as primary stakeholder. *Organization & Environment*, 16(1):3–33, 2003. [p]

P. J. Bickel and K. A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*, volume 117. CRC Press, 2015. [p]

R. Bivand and G. Piras. Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software, Articles*, 63(18):1–36, 2015. ISSN 1548-7660. URL https://doi.org/10.18637/jss.v063.i18. [p]

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3): 307–327, 1986. [p]

S. Borovkova and R. Lopuhaa. Spatial GARCH: A spatial approach to multivariate volatility modeling. *Available at SSRN 2176781*, 2012. [p]

K. Brent. Gender, race, and perceived environmental risk: The "white male" effect in cancer alley, la. *Sociological Spectrum*, 2010. [p]