AUSTRIAN JOURNAL OF STATISTICS
Volume 41 (2012), Number 1, 81–92

Why Does It Always Rain on Me? A Spatio-Temporal Analysis of Precipitation in Austria

Nikolaus Umlauf, Georg Mayr, Jakob Messner, and Achim Zeileis
Universität Innsbruck

Abstract: It is popular belief that the weather is "bad" more frequently on weekends than on other days of the week and this is often perceived to be associated with an increased chance of rain. In fact, the meteorological literature does report some evidence for such human-induced weekly cycles although these findings are not undisputed. To contribute to this discussion, a modern data-driven approach using structured additive regression models is applied to a newly available high-quality data set for Austria. The analysis investigates how an ordered response of rain intensities is influenced by a (potential) weekend effect while adjusting for spatio-temporal structure using spatially varying effects of overall level and seasonality patterns. The underlying data are taken from the HOMSTART project which provides daily precipitation quantities over a period of more than 60 years and a dense net of more than 50 meteorological stations all across Austria.

Zusammenfassung: Gemäß landläufiger Meinung ist das Wetter an Wochenenden häufiger „schlecht" als unter der Woche, was oft mit dem Eindruck einer erhöhten Wahrscheinlichkeit für Regen am Wochenende einhergeht. Tatsächlich liefert die meteorologische Lieferung auch Hinweise auf durch Menschen verursachte wöchentliche Wetterzyklen, auch wenn solche Befunde nicht immer unkontroversiell sind. Als Beitrag zu dieser Diskussion wird ein moderner datengetriebener Ansatz basierend auf strukturierten additiven Regressionsmodellen auf einen neuen qualitativ hochwertigen Datensatz für Österreich angewendet. Die Analyse untersucht, inwieweit eine ordinale Variable von Regenintensitäten von einem (möglichen) Wochenendeffekt abhängt, wobei für die spatio-temporale Struktur durch Schätzung eines räumlich korrelierten Effekts sowie saisonaler Muster adjustiert wird. Die zugrundeliegenden Daten entstammen dem HOMSTART-Projekt, das tägliche Niederschlagsmengen über einen Zeitraum von mehr als 60 Jahren und ein dichtes Gitter von mehr als 50 meteorologischen Stationen über ganz Österreich bereitstellt.

Keywords: Rainfall, Generalized Additive Model, Structured Additive Regression Model, Ordered Probit Model, HOMSTART, **BayesX**, R.

1 Introduction

Many people have the impression that the weather is "bad" more frequently on weekends when they would be able to enjoy outdoor activities much more than during work days. "Bad" weather is often associated with the occurrence of precipitation. Scientific literature also reports that human-induced periodic weekly cycles in climate time series may actually exist, especially due to higher aerosol input into the atmosphere from human activities during the week than on the weekend. For example, Bäumer and Vogel (2007) report evidence for such weekly patterns in data from 12 German meteorological stations. However, such results are not uncontroversial, e.g., the Bäumer and Vogel (2007) results have been challenged by Hendricks Franssen (2008) using data for Swiss stations where there was no evidence for weekly patterns if spatial correlations are taken into account.

Here, we contribute to the discussion by applying a modern flexible regression model for spatio-temporal data to a novel high-quality precipitation data set for Austria. More precisely, we employ data provided by Nemec, Gruber, Chimani, and Auer (2011), consisting of daily precipitation observations over 60 years for a rather dense net of meteorological stations. Moreover, the data are homogenized to adjust for effects, e.g. caused through changes in the data collection process or measurement technology. The statistical model employed assesses the weekend effect while accounting for the inherent temporal and spatial correlations as well as threshold effects in the response by applying a penalized regression approach for an ordered response. It utilizes well-established mixed-model technology to capture the rather complex and possibly nonlinear relationships in the data (e.g., see Lin and Zhang, 1999 and Kneib and Fahrmeir, 2006).

The remainder of this paper is as follows. The next section gives a concise overview of the available data, Section 3 presents the statistical model and briefly discusses the methodological background as well the software used. Estimation results are reported in Section 4 and are further discussed in Section 5. A summary and outlook is given in Section 6.

2 Data

Data are taken from the HOMSTART project

http://www.zamg.ac.at/forschung/klimatologie/klimawandel/homstart/

conducted at the Zentralanstalt für Meteorologie und Geodynamik (ZAMG) and funded by the Austrian Climate Research Programme (ACRP). The project provides daily precipitation time series for 71 Austrian meteorological stations for the period 1948–2009 which are freely available online for research purposes (Nemec, Chimani, Gruber, and Auer, 2011; Nemec, Gruber, et al., 2011).

Here, we consider the subset of 57 stations for which homogenized precipitation series are provided by the HOMSTART project. The data for the remaining 14 stations is not included in the project as it could not be homogenized, e.g., due to missing appropriate reference stations or other uncertainties in adopting structural changes. The time period covers daily observations from 1948 until the end of 2009 while for some of the stations there are a few gaps in the observed time series. Altogether the data set consists of almost 1,120,000 observations.

Precipitation is measured in millimeters in a standardized reservoir with a resolution of 0.1 mm. No precipitation is indicated in the data with negative data entries (-0.1). When a human observer notices a wetting of the ground but no precipitation could be measured, this trace of precipitation is represented by zero. Because the majority of observations is clustered at -0.1 or 0, the statistical analysis has to appropriately adjust for these threshold or censoring effects in the data. Hence, we adopted an ordered categorical approach based on thresholds. The data are grouped into four rain intensity categories: none (≤ 0), low $(0, 1)$, medium $[1, 5)$ and high (≥ 5). Both -0.1 and 0 are combined into a single category of no *measurable* precipitation. The relative frequencies of daily precipitation sums from 7:00 CET (central European time) of one day to the next in the four categories across all 57 Austrian stations and all 62 years are: 56 % none, 11 % low, 16 % medium and 17 % high.

Besides the rain intensity, the information on the longitude and latitude coordinates of each meteorological station is used to capture a spatially correlated effect of precipitation in Austria.

3 Methods and Software

The space-time structure of the data set, with the repeated categorical observations, requires a flexible regression model that can deal simultaneously with possible nonlinear temporal effects as well as the inherent spatial correlation of meteorological stations. A very general model class supporting these patterns is called structured additive regression (STAR) models (Fahrmeir, Kneib, and Lang, 2004; Brezger and Lang, 2006). E.g., Kneib and Fahrmeir (2006) propose this type of model for explaining the damage state of trees in terms the age of the trees and the longitude-latitude coordinates of the stands, where the modeling problem is similar to the one here.

The STAR model class is based on the framework of (Bayesian) generalized linear models (GLM, see e.g., Fahrmeir, Kneib, and Lang, 2009, and Fahrmeir and Tutz, 2001). In this analysis, we apply a threshold model with cumulative probit link given by

$$\Phi^{-1}\left\{P(\texttt{rain}_{it} \leq r)\right\} = \eta_{it}^{(r)}, \tag{1}$$

with rain intensity categories $r = (\texttt{none}, \texttt{low}, \texttt{medium})$, stations $i = 1, \ldots, 57$ and time $t = 1, \ldots, 22645$. The probabilities for the individual categories can then be obtained by taking differences of the cumulative probabilites; in particular for category `high` the probability is $P(\texttt{rain}_{it} = \texttt{high}) = 1 - P(\texttt{rain}_{it} \leq \texttt{medium})$.

Similar to generalized additive models (GAM), the predictor η in STAR models is relaxed from linearity assumptions, i.e., besides linear modeled terms, the structure of η may additionally include one, two or even higher-dimensional (possibly smooth) functions, e.g., comprising nonlinear effects of continuous covariates, two-dimensional surfaces, spatially correlated effects, varying coefficients, spatially varying effects, random intercepts and slopes, etc. In the rain model (1), the structured additive predictor is represented by

$$\begin{aligned}\eta_{it}^{(r)} = \xi_r - \{&f_{\mathrm{kr}}(\texttt{long}_i, \texttt{lat}_i) + f_{\mathrm{ps}}(t) + \\ &\alpha_{i,1} \cdot \cos(2\pi \cdot t + \phi_{i,1}) + \alpha_{i,2} \cdot \cos(4\pi \cdot t + \phi_{i,2}) + \omega_i \cdot I_{\texttt{weekend}}(t)\},\end{aligned} \tag{2}$$

where ξ_r is the category specific threshold, functions $f_{\mathrm{kr}}(\cdot, \cdot)$ and $f_{\mathrm{ps}}(\cdot)$ are penalized terms, while the remaining parameters are classical parametric terms. Function $f_{\mathrm{kr}}(\cdot, \cdot)$

models a nonlinearly-correlated spatial effect of the meteorological stations using longitude and latitude coordinates applying kriging (Stein, 1999; Fahrmeir et al., 2009). Function $f_{\text{ps}}(\cdot)$ captures the time trend in t and is also modeled nonlinearly using P-splines (Eilers and Marx, 1996). The terms $\alpha_{i,1} \cdot \cos(2\pi \cdot t + \phi_{i,1})$ and $\alpha_{i,2} \cdot \cos(4\pi \cdot t + \phi_{i,2})$ denote harmonic seasonal terms at annual and half-annual frequencies with station-specific amplitude parameters $\alpha_{i,1}$, $\alpha_{i,2}$ and phases $\phi_{i,1}$, $\phi_{i,2}$, respectively (see e.g., Cryer and Chan, 2008). Note that in a regression setting, the harmonic functions may always be decomposed into linear terms as $\alpha_{i,j} \cdot \cos(2\pi j \cdot t + \phi_{i,j}) = \gamma_{i,j,1} \cdot \cos(2\pi j \cdot t) + \gamma_{i,j,2} \cdot \sin(2\pi j \cdot t)$ such that the amplitude is $\alpha_{i,j} = \sqrt{\gamma_{i,j,1}^2 + \gamma_{i,j,2}^2}$ and the phase is $\phi_{i,j} = \gamma_{i,j,2}/\gamma_{i,j,1}$ for frequency $j = 1, 2$. The term $\omega_i \cdot I_{\texttt{weekend}}(t)$ represents a spatial weekend effect of station i, i.e., $I_{\texttt{weekend}}(\cdot)$ is an indicator function with 1 if t is measured at weekends and 0 otherwise.

Although a STAR predictor may include very flexible functional forms, such as the nonlinear and spatially-correlated functions f_{kr} and f_{ps} in (2), all functions can be represented in a unified way as linear predictors. Here, the predictor (2) may be rewritten as

$$\boldsymbol{\eta}^{(r)} = \mathbf{Z}_{\text{kr}}\boldsymbol{\beta}_{\text{kr}} + \mathbf{Z}_{\text{ps}}\boldsymbol{\beta}_{\text{ps}} + \mathbf{X}\boldsymbol{\gamma}\,,$$

where the design matrix $\mathbf{Z}_k$ depends on the specific functional form chosen in the nonlinear term ($k = \text{ps}$) and the spatial term ($k = \text{kr}$), respectively, and $\boldsymbol{\beta}_k$ are the corresponding unknown regression coefficients to be estimated. Furthermore, $\mathbf{X}\boldsymbol{\gamma}$ corresponds to the parametric part of the STAR model, i.e., including the threshold parameters as well as the parameters of the spatially varying seasonal and weekend effect. Thus, $\mathbf{X}$ contains the threshold dummies, the transformed harmonic regressors (in interaction with station dummies), and a weekend dummy (again in interaction with station dummies), and $\boldsymbol{\gamma}$ collects all corresponding coefficients.

To ensure particular functional forms, prior distributions are assigned to the regression coefficients. The general form of the prior for $\boldsymbol{\beta}_k$ is

$$p(\boldsymbol{\beta}_k|\tau_k^2) \propto \exp\left(-\frac{1}{2\tau_k^2}\boldsymbol{\beta}_k{}'\mathbf{K}_k\boldsymbol{\beta}_k\right)\,,$$

where $\mathbf{K}_k$ is a quadratic penalty matrix that shrinks parameters towards zero or penalizes too abrupt jumps between neighboring parameters. The variance parameter τ_k^2 is equivalent to the inverse smoothing parameter in a frequentist approach and controls the trade off between flexibility and smoothness.

In this analysis, for empirical Bayes inference, τ_k^2 is considered an unknown constant which is determined via restricted maximum likelihood (REML), i.e., models that exhibit penalized terms may be represented as mixed models with i.i.d. random effects (e.g., see Lin and Zhang, 1999, Kamman and Wand, 2003, Wand, 2003, Ruppert, Wand, and Carrol, 2003, Fahrmeir et al., 2004, and for models with categorical responses Kneib and Fahrmeir, 2006). For a more detailed overview of distributions and functional forms that may be modeled using STAR regression see Fahrmeir et al. (2009).

Model fitting is carried out in **R2BayesX** (Umlauf, Lang, Kneib, and Zeileis, 2011), an R interface (R Development Core Team, 2011) to **BayesX** (Brezger, Kneib, and Lang, 2005; Belitz, Brezger, Kneib, and Lang, 2009), which supports estimation of a wide variety of STAR models. All data handling is carried out within R, using the **sp** classes

(Bivand, Pebesma, and Gómez-Rubio, 2008) for managing spatial information and **zoo** (Zeileis and Grothendieck, 2005) for temporal information. For obtaining completely regular series with frequency 365, the observations from February 29 (if any) are omitted prior to constructing the time trend and harmonic seasonal regressors. The preprocessed data is stored on disc prior to calling **BayesX** in order to save memory for fitting the STAR model, and subsequently the results are read back into R (all through **R2BayesX**). Effects with spatial variation are visualized in the following using the **maptools** package (Lewin-Koh and Bivand, 2011), drawing separate points for each meteorological station and using **akima** interpolation (Akima, Gebhardt, Petzoldt, and Mächler, 2009) inbetween. Color palettes based on HCL colors (Zeileis, Hornik, and Murrell, 2009) are employed for coding size and spatialvariation of the effects.

4 Results

Thresholds: The estimated threshold are $\hat{\boldsymbol{\xi}} = (0.17, 0.46, 0.98)$, corresponding to category-specific probabilities of 0.57 (none), 0.11 (low), 0.16 (medium), 0.16 (high) at zero for all other effects. Thus, these thresholds essentially correspond to the mean frequencies of the four categories indicated in Section 2, averaged across space and time.

Spatial effect: In Figure 1, the estimated spatially-correlated effect $\hat{f}_{\text{kr}}(\texttt{long}_i, \texttt{lat}_i)$ is shown. The effect indicates that regions with positive effect (i.e., higher probabilities for higher categories) accumulate in the north-west part of Austria, where the highest estimated effects are located in Vorarlberg and Salzburg. The effects in regions that are south and east of the Alpine mountain range are mostly negative (i.e., with lower rain probabilities). The most pronounced negative value are estimated for regions around Laa an der Thaya in the far north-east.

Trend effect: The estimated nonlinear time trend effect $\hat{f}_{\text{ps}}(t)$ is shown in Figure 4 (left). Although there are clearly some periods with higher and lower precipitation (e.g., the peak between 1960 and 1970) the overall trend seems neither to increase nor to decrease. This can also be seen by the estimated 95% confidence bands, which only cross the zero line at a few points in time. The very high effect at the beginning of the observation period is due to the small number of observations available at this time interval and is most probably an artifact.

Seasonal effect: In Figure 4 (right) the estimated harmonic effect $\hat{\alpha}_{i,1} \cdot \cos(2\pi \cdot t + \hat{\phi}_{i,1}) + \hat{\alpha}_{i,2} \cdot \cos(4\pi \cdot t + \hat{\phi}_{i,2})$ is shown for one year for each meteorological station i. The estimated periodic functions seem to be rather similar, especially in the peak rain season during summertime, which is also indicated by the estimated phases that do not vary too much across stations (results not shown). However, there is some clear spatial variation, especially differences between the regions north and south of the Alps. This can be brought out in two ways: First, the color shading of the curves in Figure 4 (right) illustrates that the southern stations have a clear annual peak while for the northern stations the semiannual pattern is more pronounced. Second, the amplitudes $\hat{\alpha}_{i,1}$ pertaining

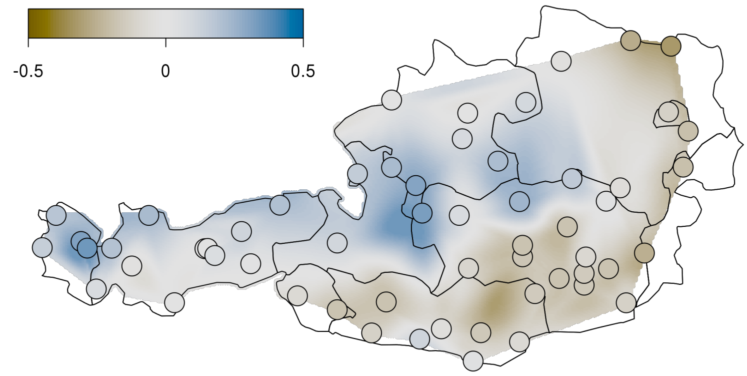

Figure 1: Spatial effect $\hat{f}_{\mathrm{kr}}(\mathtt{long}_i, \mathtt{lat}_i)$. The range of the color scale is 1.0 on the scale of the linear predictor.

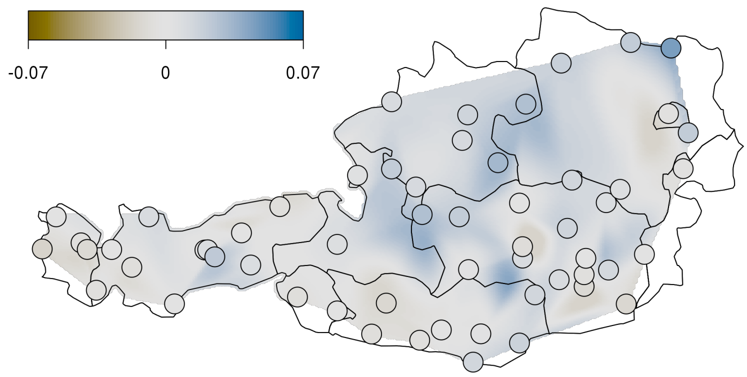

Figure 2: Spatial variation of weekend effect $\hat{\omega}_i$. The range of the color scale is 0.14 on the scale of the linear predictor.

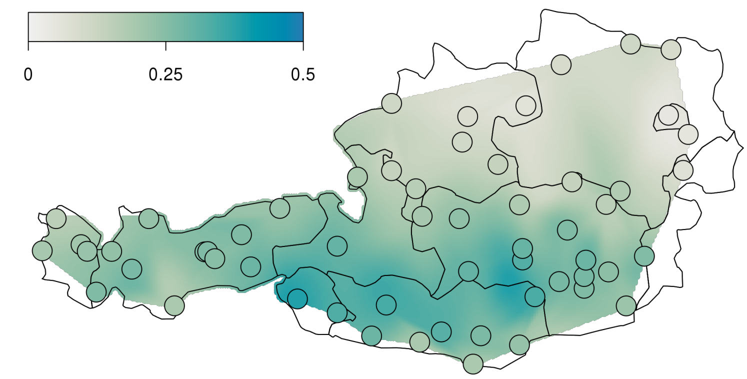

Figure 3: Amplitudes: Spatial variation of estimated amplitudes $\hat{\alpha}_{i,1}$ for annual seasonal changes. The range of the color scale corresponds to 1.0 on the scale of the linear predictor (due to multiplication with the cosine wave).

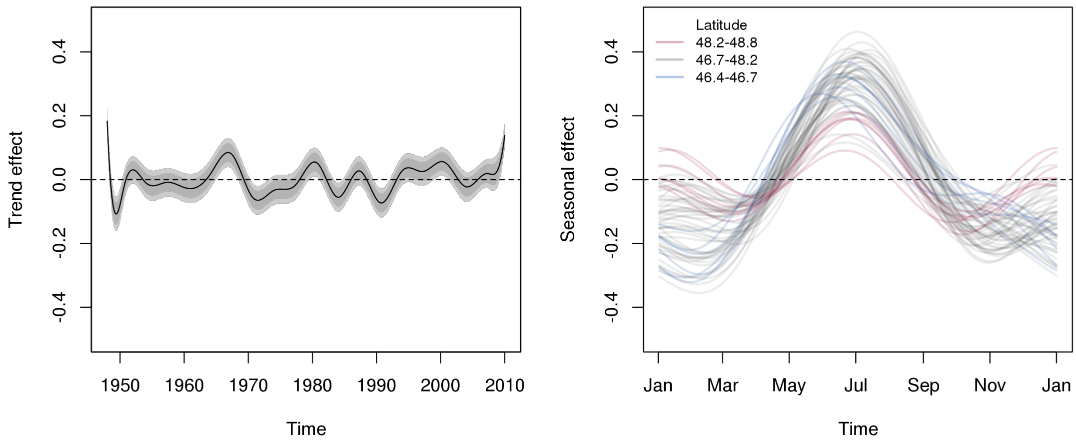

Figure 4: Time trends. Left: Estimated nonlinear time trend across years $\hat{f}_{\mathrm{kr}}(t)$. Right: Estimated seasonal variation within years (harmonic effect of order 2 for each station). To highlight spatial differences in the seasonal patterns, the curves pertaining to the most northern and southern stations are shaded red and blue, respectively.

to the annual frequency in Figure 3 show a similar pattern of low and high amplitudes in the north and south, respectively.

Weekend effect: The spatial variation of the estimated weekend effect $\hat{\omega}_i$ is displayed in Figure 2. Note that while the ranges in all other graphics correspond to changes of ± 0.5 on the latent scale of the linear predictor (i.e., one standard deviation in the probit link), the weekend effect is so small that its legend is almost one order of magnitude smaller (± 0.07). Thus, compared to all other changes, the weekend effect is extremely small. More precisely, the interquartile range of changes in the probability to stay dry (category: none) is -0.6 to 0.1 percentage points, when evaluated at average zero effects for the trend and seasonal terms. The largest change in probability to stay dry is for station Laa an der Thaya (in the north-east) where the probability decreases from 68.7 % during the week to 67.1 % on the weekend. In summary, it can be concluded that there is no relevant weekend effect at all.

Overall effect: To capture the combined effect of all terms, fitted probabilities for all four categories (none, low, medium, high) are computed for the nine stations that are closest to the capitals of the Austrian provinces[1] in Table 1. As the weekend effect is virtually irrelevant, it is excluded from the computations, i.e., set to its reference level zero. Similarly, as there is no systematic upward or downward trend over time, we also set the trend effect to its reference level zero. To capture changes over the year, we employ two time points: first of January and July, respectively. Table 1 brings out several insights that have been discussed separately in the paragraphs above, e.g.: The probability for rain is highest

[1]For Niederösterreich (Lower Austria), the measurements for the capital St. Pölten were excluded from the analysis as the series could not be harmonized in the HOMSTART project (see Section 2). Hence, Zwettl is used as the closest location within Niederösterreich. In Oberösterreich (Upper Austria), the meteorological station Hörsching is very close to the captial Linz.

Table 1: Fitted probabilities for all four categories (with amount of rain in mm/day) for nine stations (closest to the capitals of the Austrian provinces) at two seasonal dates (first of January and July, respectively). Trend and weekend effect are set to their reference level zero.

	none: ≤ 0		low: $(0,1)$		medium: $[1,5)$		high: ≥ 5	
Location	Jan 1	Jul 1	Jan 1	Jul 1	Jan 1	Jul 1	Jan 1	Jul 1
Bregenz	51.9	40.3	11.4	11.5	17.3	19.6	19.5	28.6
Innsbruck (University)	61.9	41.6	10.5	11.5	14.4	19.4	13.2	27.4
Salzburg (Airport)	53.8	38.9	11.2	11.5	16.7	19.8	18.2	29.8
Hörsching	55.2	49.3	11.1	11.5	16.4	17.9	17.3	21.4
Klagenfurt	70.8	50.9	9.1	11.4	11.4	17.5	8.7	20.2
Graz (University)	58.2	51.0	10.9	11.4	15.5	17.5	15.4	20.1
Zwettl (Stift)	69.9	48.9	9.3	11.5	11.7	18.0	9.1	21.7
Vienna (Hohe Warte)	57.4	57.9	11.0	10.9	15.8	15.6	15.9	15.6
Eisenstadt	64.4	59.1	10.1	10.8	13.6	15.2	11.9	14.8

in Bregenz and Salzburg. There is only very low seasonal variation in Vienna. Furthermore, these results are complemented with aspects that were not immediately obvious from analyzing the effects separately, e.g.: While the probability for rain in Innsbruck, Klagenfurt, and Zwettl is very low in winter, it is relatively high in summer.

5 Discussion

The wisdom of the crowds holds that the weather is more likely to be "bad" on the weekend than during the week. When this impression is checked objectively where "bad weather" is specified as a day with measurable precipitation, it cannot be substantiated for a longterm (62 years), spatially dense (57 stations) data set from Austria. Cehak (1982) had come to the same conclusion for one Austrian station (Vienna).

The so-called "weekend effect" has been extensively debated in meteorological literature. Effects of a weekly cycle of aerosol concentration from human activities on cloud microphysics were proposed as a possbile explanation. However, while Barmet, Kuster, Muhlbauer, and Lohmann (2009) find a statistically significant Sunday minimum and Wednesday maximum of aerosol concentration at the surface in Switzerland, they could not find a similar weekly cycle for precipitation when using three different methods: the Kruskal-Wallis test, a spectral analysis, and constructing 6 and 8 day weeks. Even in the heavily polluted and thus aerosol-rich region at the border of Germany, the Czech Republic and Poland, no significant signal of a weekly cycle could be found by Stjern (2011), who used the same methods as Barmet et al. (2009) for 30 stations over a 26-year period. A study of 158 stations in western and northern Europe between 1931 and 2005 (Laux and Kunstmann, 2008) using a t-test and stationary block bootstrap resampling also could not find a significant weekly cycle of precipitation. Similarly, Schultz, Mikkonen, Laaksonen, and Richman (2007) noted the absence of a weekly cycle of either the occurrence or the amount of precipitation in the 42-year records of 219 stations in the USA. Bäumer and Vogel (2007) provide the lone supporting evidence of public sentiment. They used 13 stations in Germany from 1991–2005. A careful analysis and Monte Carlo simulations using two Swiss stations by Hendricks Franssen (2008) showed, however, that the apparently significant p-values can be attributable to random effects, and that the failure to include spatial autocorrelation in their analysis might have wrongly led to the significant weekly periodicity in the study of Bäumer and Vogel (2007).

The current study used different statistical methods but still arrives at the same conclusion as the majority of the meteorological literature, adding to the robustness of the result that there is no weekend effect for precipitation.

The spatial part of the rain model in Figure 1 reflects the separation of Austria through the Alps into a wetter northern and a drier southern part. The maxima are in regions where atmospheric flow impinging from westerly to northerly directions first encounters topography which induces strong lifting (e.g. Arlberg region; region south of Salzburg). The low values in the (north)easternmost part of Austria are due to the longer distance from oceanic moisture sources.

Deep convection and more available moisture in the atmosphere cause the precipitation maximum in the warm season for all locations (cf. Figure 4, right). Summer is also the main rainy season throughout the whole Alpine region (cf. Frei and Schär, 1998). Figure 3 concisely depicts the much stronger seasonal differences on the southern side of the Alpine crest than on the northern side, towards which the proximity to the warm Mediterranean Sea and higher solar insolation contribute. The earlier switch on the south side to a positive anomaly in late spring and the later transition back to the negative anomaly in fall are related to frequent south(west)erly flow impinging on the Alps causing strong precipitation. In early fall the Mediterranean is still close to its maximum temperatures and thus an ample source of moisture (cf. Frei and Schär, 1998).

6 Summary and Outlook

In this analysis we apply a modern penalized regression approach based on structured additive regression (STAR) models to a very rich data set of precipitation at 57 meteorological stations across Austria between 1948 and 2009. The model aims to expose whether or not there is a weekend effect while incorporating spatio-temporal patterns, i.e., including spatial correlation, an (inter-annual) time trend, and (intra-annual) seasonal patterns. However, the estimation results cannot support a relevant change of precipitation between weekdays and weekends for any of the locations in the data, whereas considerably large spatial differences in both level and seasonality patterns could be clearly identified. The regions that exhibit the strongest precipitation effects are the northern parts of Austria, and especially the Alpine regions in Vorarlberg and Salzburg. The estimated seasonality also shows a substantial variation between low and high seasonal amplitudes in the north and south, respectively. The estimated time trend remains relatively constant over the observation period.

To enhance future analyses of similar data sets within the framework of STAR models various extensions would be conceivable: One idea would be to raise the order of the harmonic seasonal effect to some large value (20, say) such that the seasonal variation is captured in finer detail. However, to prevent overfitting, a method that controls the trade-off between flexibility and smoothness needs to be adopted, e.g., by using a suitable penalty matrix $\mathbf{K}_k$ on the corresponding regression coefficients or by using a penalized likelihood approach as expounded by Hunsberger, Albert, Follmann, and Suh (2002). A second idea would be to account for spatial correlation not only of the overall level but also of the various regression coefficients. Specifically, spatially-correlated varying effects for the seasonality and the weekend could be employed rather than unrestricted station-specific coefficients. While some building blocks for such models are already available in state-of-the-art GAM and STAR software, such as **BayesX** or **mgcv** (Wood, 2006), further infrastructure is required, especially in combination with ordered categorical models. Hence, it would be desirable to provide further R functionality for model terms utilizing space-time information based on the ideas above.

Computational Details

Our results were obtained using R 2.14.0 with the packages **R2BayesX** 0.1-1/r242, **akima** 1.1-0, **colorspace** 1.1-0, **maptools** 0.8-10, **sp** 0.9-91, **zoo** 1.7-6. **R2BayesX** was used to interface **BayesX** 2.0.1. All software is freely available: R and most packages can be obtained under the General Public License (GPL) or the ACM License (in case of the **akima** package) from the Comprehensive R Archive Network (`http://CRAN.R-project.org/`). **R2BayesX** is under development on the R-Forge system at `http://bayesr.R-Forge.R-project.org/` while **BayesX** can be downloaded at no cost from its web page at `http://www.stat.uni-muenchen.de/bayesx/`.

To replicate our analyses, we provide a zipped data archive containing a `data` subdirectory as well as several R scripts: (a) '`homstart.R`' downloads the HOMSTART data from the ZAMG web page, reads them into R, and combines them with the metainformation about the meteorological stations. (b) '`model.R`' uses the data from (a), sets up the full regressor matrix (using several utility functions from '`functions.R`'), and then calls **R2BayesX** for fitting the STAR model. (c) '`graphics.R`' loads the estimated effects from (b), the shape files for Austria (originally obtained from the public domain Epi Info project page, `http://wwwn.cdc.gov/epiinfo/`), and creates the maps and time series plots from the paper. Note that a sufficient amount of RAM (at least 8 GB) is required for holding the data in memory and fitting the STAR model.

Acknowledgements

We are thankful to Stefan Lang and Thomas Kneib for help and support for **BayesX**, and Ingeborg Auer for the HOMSTART data.

References

Akima, H., Gebhardt, A., Petzoldt, T., and Mächler, M. (2009). **akima**: Interpolation of irregularly spaced data [Computer software manual]. (R package version 0.5-4, `http://CRAN.R-project.org/package=akima`)

Barmet, P., Kuster, T., Muhlbauer, A., and Lohmann, U. (2009). Weekly cycle in particulate matter versus weekly cycle in precipitation over Switzerland. *Journal of Geophysical Research*, *114*(D5), 1–7.

Bäumer, D., and Vogel, B. (2007). An unexpected pattern of distinct weekly periodicities in climatological variables in Germany. *Geophysical Research Letters*, *34*(L03819), 1–4.

Belitz, C., Brezger, A., Kneib, T., and Lang, S. (2009). **BayesX**: Software for Bayesian inference in structured additive regression models [Computer software manual]. (Version 2.0.1, `http://www.stat.uni-muenchen.de/bayesx/`)

Bivand, R. S., Pebesma, E. J., and Gómez-Rubio, V. (2008). *Applied spatial data analysis with R*. New York: Springer-Verlag. (`http://www.asdar-book.org/`)

Brezger, A., Kneib, T., and Lang, S. (2005). **BayesX**: Analyzing Bayesian structured additive regression models. *Journal of Statistical Software*, *14*(11), 1–22. (`http://www.jstatsoft.org/v14/i11/`)

Brezger, A., and Lang, S. (2006). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics & Data Analysis*, *50*, 947–991.

Cehak, K. (1982). Note on the dependence of precipitation on the day of the week in a medium industrialized city. *Archives for Meteorology, Geophysics, and Bioclimatology Series B*, *30*(3), 247–251.

Cryer, J. D., and Chan, K.-S. (2008). *Time series analysis with applications in R*. Springer-Verlag.

Eilers, P. H. C., and Marx, B. D. (1996). Flexible smoothing using B-splines and penalized likelihood. *Statistical Science*, *11*, 89–121.

Fahrmeir, L., Kneib, T., and Lang, S. (2004). Penalized structured additive regression for space time data: A Bayesian perspective. *Statistica Sinica*, *14*, 731–761.

Fahrmeir, L., Kneib, T., and Lang, S. (2009). *Regression – Modelle, Methoden und Anwendungen* (2nd ed.). Berlin: Springer-Verlag.

Fahrmeir, L., and Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models*. New York: Springer-Verlag.

Frei, C., and Schär, C. (1998). A precipitation climatology of the Alps from high-resolution rain-gauge observations. *International Journal of Climatology*, *18*(8), 873–900.

Hendricks Franssen, H. J. (2008). Comment on "An unexpected pattern of distinct weekly periodicities in climatological variables in Germany" by Dominique Bäumer and Bernhard Vogel. *Geophysical Research Letters*, *35*(L05802), 1–3.

Hunsberger, S., Albert, P. S., Follmann, D. A., and Suh, E. (2002). Parametric and semiparametric approaches to testing for seasonal trend in serial count data. *Biostatistics*, *3*(3), 289–298.

Kamman, E. E., and Wand, M. P. (2003). Geoadditive models. *Applied Statistics*, *52*, 1–18.

Kneib, T., and Fahrmeir, L. (2006). Structured additive regression for multicategorical space-time data: A mixed model approach. *Biometrics*, *62*, 109–118.

Laux, P., and Kunstmann, H. (2008). Detection of regional weekly weather cycles across Europe. *Environmental Research Letters*, *3*(4), 044005.

Lewin-Koh, N. J., and Bivand, R. (2011). **maptools**: Tools for feading and handling spatial objects [Computer software manual]. (R package version 0.8-10, http://CRAN.R-project.org/package=maptools)

Lin, X., and Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society B*, *61*, 381–400.

Nemec, J., Chimani, B., Gruber, C., and Auer, I. (2011). Ein neuer Datensatz homogenisierter Tagesdaten [ÖGM Bulletin]. (2011/1), 19–20. (http://www.meteorologie.at/docs/OEGM_bulletin_2011_1.pdf)

Nemec, J., Gruber, C., Chimani, B., and Auer, I. (2011). *Trends in extreme temperature indices in Austria based on a new homogenised dataset of daily minimum and maximum temperature series.* (submitted to *International Journal of Climatology*)

R Development Core Team. (2011). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. (ISBN 3-900051-07-0, http://www.R-project.org/)

Ruppert, D., Wand, M. P., and Carrol, R. J. (2003). *Semiparametric regression*. New York: Cambridge University Press.

Schultz, D. M., Mikkonen, S., Laaksonen, A., and Richman, M. B. (2007). Weekly precipitation cycles? Lack of evidence from United States surface stations. *Geophysical Research Letters*, *34*(22), 2–5.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory of kriging*. New York: Springer-Verlag.

Stjern, C. W. (2011). Weekly cycles in precipitation in a polluted region of Europe. *Atmospheric Chemistry and Physics Discussions*, *11*(1), 1777–1801.

Umlauf, N., Lang, S., Kneib, T., and Zeileis, A. (2011). Structured additive regression models: An R interface to **BayesX** [Computer software manual]. (R package version 0.1-1/r242, http://bayesr.R-Forge.R-project.org/)

Wand, M. P. (2003). Smoothing and mixed models. *Computational Statistics*, *18*, 223–249.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton: Chapman & Hall/CRC.

Zeileis, A., and Grothendieck, G. (2005). **zoo**: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software*, *14*(6), 1–27. (http://www.jstatsoft.org/v14/i06/)

Zeileis, A., Hornik, K., and Murrell, P. (2009). Escaping RGBland: Selecting colors for statistical graphics. *Computational Statistics & Data Analysis*, *53*(9), 3259–3270.

Authors' addresses:

Nikolaus Umlauf and Achim Zeileis
Department of Statistics
Faculty of Economics and Statistics
University of Innsbruck
Universitätsstr. 15
6020 Innsbruck, Austria

Georg Mayr and Jakob Messner
Institute of Meteorology and Geophysics
Faculty of Geo- and Atmospheric Sciences
Universität Innsbruck
Innrain 52
6020 Innsbruck, Austria