# Modeling Seasonal Effects on the Lexis Surface

Roland Rau*
Terry Sanford Institute of Public Policy
Duke University
roland.rau@duke.edu

Jutta Gampe
Max Planck Institute
for Demographic Research
gampe@demogr.mpg.de

Paul H.C. Eilers
Faculty of Social and Behavioural
Sciences, Utrecht University
p.h.c.eilers@fss.uu.nl

Brian D. Marx
Dept. of Experimental Statistics
Louisiana State University
bmarx@lsu.edu

February 22, 2007

Paper prepared for

Session 1001: Statistical Modeling Issues in Population Research

of the 2007 PAA Annual Meeting, New York, March 29-31, 2007

*Corresponding Author. Duke University; Terry Sanford Institute of Public Policy; Population, Policy, and Aging Research Center (PPARC); Box 90309; 302 Towerview Road; Durham, NC 27708-0309; USA; Email: roland.rau@duke.edu; Phone: +1-919-613-9349; Fax: +1-919-613-9940

# 1 INTRODUCTION

Trajectories of human mortality are well-known in two dimensions: *Over age*, mortality drops from infant mortality to a minimum during early adolescence and increases exponentially for most of the adult ages before it eventually levels-off (e.g. Robine and Vaupel, 2002; Vaupel et al., 1998, 2007). *Over time* life expectancy rose in most countries for recent history due to reduced death rates, first at infant and younger ages and at advanced ages since the 1950s (e.g. Kannisto et al., 1994; Oeppen and Vaupel, 2002; Tuljapurkar et al., 2000).

Another dimension, however, received less attention, despite being known at least since Hippocrates' times:[1] the *seasonal* fluctuations throughout the year. Due to increased media coverage (e.g. CNN, 2003), the general public often equates seasonal mortality with heat waves; the regular real 'grim reaper', though, is winter in most countries. Previous studies for the elderly population of Denmark have shown that the relative risks of dying are at least 15% higher during winter than during summer (Rau, 2007; Rau and Doblhammer, 2003). The national statistical office for England & Wales estimates a cold-related death toll of more than 32,000 deaths *annually* during the last ten years (National Statistics Online, 2007). As Figure 1 shows in its upper panel, deaths in the United States follow a rather stable sinusoidal pattern over the years with peaks in January and minima in August. When aggregated into twelve months (lower two panels), separately for women and men, one recognizes (a) again the rather smooth curve throughout the year. In addition, (b) differences between the sexes ares rather negligible, a finding which has been reported previously (Eurowinter Group, 1997; Gemmell et al., 1999, 2000; Nakai et al., 1999; Yan, 2000). Hence we

[1] Hippocrates wrote in about 400BC in his influential work "On Airs, Waters, and Places": "Whoever wishes to investigate medicine properly, should proceed thus: in the first place to consider the seasons of the year, and what effects each of them produces for they are not at all alike, but differ much from themselves in regard to their changes."

chose to focus solely on female data in our article.

Two groups of causes of death are mainly affected and are analyzed in our paper: circulatory diseases (cerebrovascular and cardiovascular) and respiratory diseases. According to the biomedical literature, the cold triggers deaths from these causes via various mechanisms such as lower resistance of the immune system against respiratory infections, hypertension, changes in the blood composition or constriction of the blood vessels (e.g. Donaldson et al., 1998a,b; Eurowinter Group, 1997, 2000; Huynen et al., 2001; Johnson and Griffiths, 2003; Keatinge, 1986; Woodhouse et al., 1993).

Paradoxically, colder countries or countries where there are large differences between winter and summer climate appear to suffer less from excess winter mortality than countries with warm or moderate climate. It has been shown for Europe, for example, that the differences between winter and summer mortality are considerably larger in the United Kingdom, Ireland, Portugal, Spain and Italy than in Nordic countries like Sweden, Norway, or Finland (Grut, 1987; Healy, 2003; McKee, 1989). This suggests that social factors such as clothing, housing conditions and exposure to outdoor cold can mediate the physiological impact of cold temperatures on the human body. The spread of central heating is argued to be the main cause for the decline in seasonality in mortality during recent decades (Aylin et al., 2001; Donaldson, 2002; Donaldson and Keatinge, 1997; Keatinge et al., 1989; McDowall, 1981; Sakamoto-Momiyama, 1977, 1978).

A common approach to detect changes in seasonality over age or time is to decompose the time-series of data and to extract then the seasonal component. Examples of such methods are X-11, SABL, STL, BV4, or TRAMO/SEATS (Cleveland et al., 1990, 1981; Speth, 2004; Yaffee, 2000). Simulation studies have shown that these methods work well with simple data structures. However,

they struggle to yield the correct results when faced with a changing trend and a varying seasonal pattern, a framework to be expected in the field of population studies (Rau, 2007). Especially, the popular X-11 method has been subject to criticism (Auerbach and Rutner, 1978; Cleveland and Tiao, 1976; Raveh, 1984). Thus, it can not be ruled out that the recently reported trend reversal of increasing seasonality of deaths in the United States (Feinstein, 2002) is a methodological artifact. Employing the same simulated data, Gampe and Rau (2003, 2004) and Rau (2007) devised a new method which performed better for typical demographic data settings (count data, changes in the trend and in the seasonal component) than the standard models. Whereas that previous approach was restricted to one dimension, changes in seasonality either over time *or* over age, the main contribution of the present paper is to generalize this approach and *estimate seasonality simultaneously over age and time using a modulation model.* The result are smoothed surface maps for overall deaths, for the amplitude *and* for the phase of the seasonal component.

# 2 DATA AND MODEL

## 2.1 Data

In our analysis, we used death counts for the United States for the years 1959 until 1998, which have been originally collected by the National Center for Health Statistics. Data for the years 1959 until 1967 were obtained from the Population, Policy, and Aging Research Center (PPARC) at Duke University. Data for the remaining years are downloadable for member institutions of the “Inter-university Consortium for Political and Social Research” (ICPSR) at the University of

Michigan.[2] For each year, one file is available which lists each death as an individual record with covariates such as age at death, sex, month and year of death, state of residence and state of occurence, cause of death, etc. Combined in all files, 78,286,825 deaths are recorded.[3] Causes of death are coded according to the "International Classification of Diseases" (ICD). During our observation period the 7th, 8th, and 9th revision of the ICD were in use.

Table 1 shows the codes used to extract data and the corresponding counts for female deaths from cardiovascular, cerebrovascular and respiratory diseases. Out of the more than 36 million registered deaths, roughly 40% were due to deaths from cardiovascular diseases whose most prominent sub-cause is ischaemic heart disease. Almost four million women died of stroke or related cerebrovascular diseases ($\approx$11% of all deaths). Although our model can handle zero counts, we cut down the age-range to have some death counts available in each cell (to 44–96 years for respiratory deaths, 40–98 for cerebrovascular deaths, and to 30–100 years for cardiovascular deaths) Despite the drastic reduction on the age-scale, our analysis was still based on 14,461,515 deaths for cardiovascular diseases (98.95% of the number of deaths for all ages), 3,876471 deaths for cerebrovascular diseases (97.62%) and 2,179,242 for respiratory diseases (89.68%). Data were available for the year 1999. We decided, however, not to use them since the adoption of the 10th revision of the ICD in that year would have caused more continuity problems than provided additional insights.[4] Data for the period before 1959 are not available.

[2]The webpage is: `http://www.icpsr.umich.edu`.

[3]Not all deaths are included in the files; for example in the year 1972, a 50% sample has been used; we multiplied deaths for this year by two.

[4]Reconstructing time-series of causes of death faces many difficulties when countries switch from one revision to another (see, for example, Meslé and Vallin, 1996); the change from ICD-9 to ICD-10 is particularly challenging since the old, purely numerical, system has been replaced by an alphanumeric scheme.

## 2.2 Model

### 2.2.1 Description of the Model

Using a log-link, our modulation model estimates the mean ($\mu_{a,t}$) of death counts, $y_{a,t}$, over age $a$ and time $t$ which are assumed to be independent and follow a Poisson distribution:

$$\log \mu_{a,t} = f_{0,a,t} + f_{1,a,t} \sin\left(\frac{2\pi}{12}t\right) + f_{2,a,t} \cos\left(\frac{2\pi}{12}t\right) \tag{1}$$

The terms to be estimated, $f_{0,a,t}$, $f_{1,a,t}$, and $f_{2,a,t}$ are expected to change smoothly over age and time; they capture the overall trend ($f_{0,a,t}$) and the seasonal fluctuations ($f_{1,a,t}$, $f_{2,a,t}$). Hence, our model follows the classical approach of seasonally adjusting time-series by decomposing the data into the long term trend, the seasonal component (the "signals"), and the irregular component (the "noise") (Shiskin, 1968).[5]

From a technical point of view, Equation 1 represents a combination of two models, both introduced by Hastie and Tibshirani (1987, 1990, 1993): a Generalized Additive Model (GAM) for the trend surface $f_{0,a,t}$, and a Varying Coefficients Model (VCM) for the seasonal components $f_{1,a,t}$ and $f_{2,a,t}$. Both have been chosen because of our implicit assumptions about the data: the trend surface for death counts is influenced by three forces: over age we expect a shape following a Gompertzian density rather closely. Due to improvements in mortality we could expect a dampening of the density over time for certain causes. Simultaneously, we may also observe an increase in death counts over time because of more people attaining higher ages as a result of lower mortality, larger birth cohorts and/or net immigration. A GAM is well-suited for such a situation when we don't want to impose anything apart from smoothness since it provides "a flexible method for identifying nonlinear covariate effects in a variety of modeling situations" (Hastie and Tibshirani,

[5] Long term cycles would be captured in the trend component of our model.

1987, p. 385). It is accomplished by using smooth nonparametric functions instead of the typical linear and parametric functions in a regression context.

For the seasonal components of our model, $f_{1,a,t}$ and $f_{2,a,t}$, we restricted the flexibility. They can change smoothly over age and time, too, but only as smooth functions of our transformed time variable ($\sin\left(\frac{2\pi}{12}\right)t$, $\cos\left(\frac{2\pi}{12}\right)t$), a typical case for VCMs where "the coefficients are allowed to vary as smooth functions of other variables" (Hastie and Tibshirani, 1993, p. 757). The estimated functions $f_{1,a,t}$ and $f_{2,a,t}$ are in themselves not interesting. We can, however, obtain the phase shift $\phi$ of the seasonal component via $\phi = \arctan\left(-\frac{f_{1,a,t}}{f_{2,a,t}}\right)$ and the amplitude $A = \left({f_{1,a,t}}^2 + {f_{2,a,t}}^2\right)^{\frac{1}{2}}$. The amplitude denotes the height of the annual spike in deaths during winter; it is our main indicator to analyze the changing seasonality over age and time. The phase denotes at which point of the year (measured in days) the peak in deaths occurs. Inferring from Figure 1, possible values should be expected between 0 and 90 (January–March).

### 2.2.2 Estimation of the Model

In 2002, Eilers and Marx demonstrated that any combinations of VCMs and GAMS can be estimated in a generalized linear model setting using $P$-Splines. Before we demonstrate the actual estimation procedure in two dimensions, we will briefly explain $P$-Spline smoothing with data $y_t$ along one dimension.

$P$-Splines employ, as shown in Eilers and Marx (1996), cubic $B$-Splines as regression bases. In a pure $B$-Spline setting, regression coefficients $\alpha$ for the smoothing splines $B$ are found by minimizing

$$|y_t - B\alpha|^2 \tag{2}$$

via OLS. The typical challenges of $B$-Spline smoothing such as finding the correct number

and positions of knots are circumvented in the $P$-Spline approach. The standard procedure is to use "too many" $B$-Splines which would normally result in overfitting.[6] Overfitting is prevented by penalizing the roughness of the regression coefficients which is measured by their second differences.

$$|y_t - B\alpha|^2 + \lambda|\Delta_2\alpha|^2 \tag{3}$$

The optimal value of the control parameter $\lambda$ has been determined in the one-dimensional applications of Gampe and Rau (2003, 2004) and Rau (2007) via Akaike's Information Criterion (AIC) using a grid-search for possible $\lambda$s.

In recent years, the framework of Eilers and Marx (1996) has been extended to two or more dimensions with some papers analyzing questions in the field of population studies (e.g. Currie et al., 2004; Eilers et al., 2006; Marx and Eilers, 2005); the current modulation model for seasonality over age and time represents a novel approach, though.

The two-dimensional approach uses the $B$-Spline bases from the one-dimensional case.[7] The compound regression basis is the Kronecker product (i.e. the tensor product of matrices) of the $B$-Spline matrices along age $a$ and time $t$:

$$\boldsymbol{B} = \boldsymbol{B}_a \otimes \boldsymbol{B}_t \tag{4}$$

More detailed explanations of the construction of such bivariate $B$-Splines can be found in Dierckx (1993) or Marx and Eilers (2003). Using the newly built Kronecker products of the $B$-splines, we can re-express our Generalized Linear Model in matrix notation where the linear

[6] "Too many" is, nevertheless, considerably smaller than the number of data $t$; otherwise the model would be overparametrized.

[7] Please note that the $B$-Spline bases for *one*-dimensional smoothing are *two*-dimensional.

predictor $\eta_{a,t}$, i.e. the right hand side of the equation, models the mean, $\mu_{a,t}$, of the death counts $y_{a,t}$ using the canonical log-link for Poisson regression models:

$$\log \mu_{a,t} = \boldsymbol{B}\theta_{0,A,T} + \text{diag}\left(\sin\left(\frac{2\pi}{12}\right)\right)\boldsymbol{B}\theta_{1,A,T} + \text{diag}\left(\cos\left(\frac{2\pi}{12}\right)\right)\boldsymbol{B}\theta_{2,A,T}\,[=\eta_{a,t}] \qquad (5)$$

The corresponding regression coefficients for the functions $f_{0,a,t}$, $f_{1,a,t}$, and $f_{2,a,t}$ from Equation 1 are $\theta_{0,A,T}$, $\theta_{1,A,T}$, and $\theta_{2,A,T}$ — all of dimension $A \times T$ where $A < a$ and $T < t$ to avoid an overparametrized model. In our applications the dimensions were $7 \times 9$.

The log-likelihood function for unknown parameters $\theta$[8] and regressors $\boldsymbol{B}$ generating the Poisson distributed response $y_{a,t}$ is (see McCullagh and Nelder, 1989, p. 211):

$$l\left(\theta, \boldsymbol{B}, y\right) = \sum_{a,t} y_{a,t}\eta_{a,t} - \sum_{a,t} e^{\eta_{a,t}} \qquad (6)$$

The log-likelihood as given in Equation 6 would be appropriate if we were interested in the typical $B$-Spline smoothing. Using the $P$-Spline approach, we penalize the rows and the columns of our $B$-Spline basis for each of the functions we are interested. The penalized log-likelihood function changes therefore to:

$$l^{\star} = l\left(\theta, \boldsymbol{B}, y\right) - \sum_{u=0}^{2} \lambda_{R,u} P_R\left(\theta_u\right) + \lambda_{C,u} P_C\left(\theta_u\right) \qquad (7)$$

where the six $\lambda$s control the roughness of the rows ($R$) and columns ($C$) for the trend ($u = 0$), the sine ($u = 1$), and the cosine term ($u = 2$). The penalty matrices $P_R$ and $P_C$ are Kronecker products with $P_R = I_L \otimes D_R$ and $P_C = D_C \otimes I_K$ where $I$ denotes identity matrices and $D$ the row- or column specific difference matrices as outlined, for example, in Eilers et al. (2006) or Marx and Eilers (2005). The iterative solution to find the maximum of the penalized likelihood (Equation 7) is explained in Eilers et al. (2007).

[8] The combined vector of $\theta_{0,A,T}$, $\theta_{1,A,T}$ and $\theta_{2,A,T}$.

Theoretically, the optimal model is found by conducting a six-dimensional grid-search for possible $\lambda$s on a log-linear scale and choosing the model with the minimum AIC value.[9] In practice, we reduced the dimensions to four by searching across "age-" and "time-" $\lambda$s for the trend surface and by assuming the same $\lambda$ across *age* for the sine and cosine term and the same $\lambda$ across *time* for the sine and cosine term.

### 2.2.3 Possible Extensions of the Model

The model as given in Equation 1 represents an approach tailored to our data setting (count data) and prior knowledge (e.g. frequency of one per year). We would like to point out, however, that our model can accomodate various other situations likewise. For example, adding an offset-term allows modelling of death rates in the presence of exposures, $e_{a,t}$, and occurences (deaths):

$$\log \mu_{a,t} = \log e_{a,t} + f_{0,a,t} + f_{1,a,t} \sin\left(\frac{2\pi}{12}t\right) + f_{2,a,t} \cos\left(\frac{2\pi}{12}t\right) \tag{8}$$

Likewise we could also incorporate higher frequencies, for example, if we expect not only one annual wave in deaths due to cold but an additional peak in mortality possibly caused by summer heat:

$$\begin{aligned} \log \mu_{a,t} =& f_{0,a,t} + \\ &+ f_{1,a,t} \sin\left(\frac{2\pi \times 1}{12}t\right) + f_{2,a,t} \cos\left(\frac{2\pi \times 1}{12}t\right) + \\ &+ f_{3,a,t} \sin\left(\frac{2\pi \times 2}{12}t\right) + f_{4,a,t} \cos\left(\frac{2\pi \times 2}{12}t\right) \end{aligned} \tag{9}$$

[9] A range for one $\lambda$ could be, for example: $10^{-4}, 10^{-3}, 10^{-2}, \ldots 10^{6}, 10^{7}$.

# 3 RESULTS AND DISCUSSION

The results for the three causes of deaths from cardiovascular, cerebrovascular, and respiratory diseases are presented in Figures 2, 3, and 4, respectively. Each figure is divided into four panels. The upper left panels display the overall trend observed in the data (i.e. $\hat{f_{0,a,t}}$ from Equation 1). The seasonality estimates over time and age are depicted in the upper right (amplitude) and the lower left (phase) panel. The amplitude depicts the extent of annual fluctuations and is the main interest in our analysis; the phase shows where mortality peaks throughout the year. The scale of the phase is given in days of the years. Hence a value of 22 would correspond to the 22nd of January. The lower right panel displays the remainder of our seasonal decomposition procedure, the (Pearson) residuals which are scaled raw residuals (McCullagh and Nelder, 1989). Instead of interpreting and discussing each cause of death separately, we investigate each of the four features of our estimation (trend, amplitude, phase, residuals) simultaneously for all three causes of death, starting with the trend, followed by the phase of the seasonal component, the residuals, and finally the amplitude of the seasonal effect.

The single most contributing cause of death for the overall pattern of mortality in general are cardiovascular diseases (cf. Table 2). It should be therefore of little surprise that the overall trend observed in the upper left panel of Figure 2 closely resembles the density of human mortality in general. The peak, which is the modal age at death from cardiovascular diseases, is slowly increasing from about age 80 to about age 90. Although we model death counts—which increased over time for cardiovascular diseases—the relative contribution of this cause of death diminished over time: from 42% during the first five years of the observation period to 35% during 1994–

1998. White (1999) calculated that the age-standardized rate to die from cardiovascular diseases per 100,000 decreased in the United States from 916 in 1940 to 294 in the year 2000,[10] one of the most remarkable developments in public health which led to the expression of the "cardiovascular revolution" (e.g. Meslé and Vallin, 2006). A similar pattern can be observed for cerebrovascular deaths (Figure 2, upper left panel); in contrast to cerebrovascular diseases, the relevance of this group has diminished over time in absolute *and* in relative numbers (see Table 2). Smith (1998) showed that crude death rates (per 100,000) from cardiovascular diseases dropped from 3,500 in 1960 to 44.3 in 2000 for people aged 85–89 in the United States.

The development of respiratory diseases (Figure 4, upper left panel) followed a different trajectory with increasing absolute and relative numbers (Tab. 2): while deaths from this cause constituted only five percent of all deaths during the first five years of the observation period, during the last five years almost 10 percent died of pneumonia and related diseases.[11] Smoking, one of the standard justifications for increasing respiratory mortality, does not appear to be an appropriate explanation in this case: since the 1960s, smoking prevalence of women in the United States decreased steadily (Dettmann, 2006; MacKay and Eriksen, 2002).

The lower left panels in the three Figures 2–4 depict the phase of the seasonal component in days of the year, i.e. the time of year when mortality peaks during the year. On average (across age and calendar time), deaths were most common between the end of January and the middle of March. A slight trend of postponement of the peak can be observed for the two circulatory dis-

[10] The data were standardized using the population from 2000.

[11] Using population estimates from the Human Mortality Database, a quick calculation showed that the crude death rate (per 100,000 females) from respiratory diseases rose from 41 to 80 between 1960 and 1995.

eases. The pattern for respiratory diseases is less stable, reflecting the higher susceptibility of this category to environmental conditions. On average, however, the phase was found to be between beginning and the middle of February (days 35 to 50).

Based on our model assumption that all signals in the data should be caught either in the trend or in the seasonal component, the residual surface should, hence, contain only noise and no structure. The residuals do not meet this assumption as can be seen in the lower right panels in all three figures (Fig. 2–4): All three causes contain a distinct cohort effect for the birth year 1900. The two circulatory diseases have no other visible features. Vertical lines appear, however, in the panel for respiratory diseases. This could be interpreted as period effects such as minor influenza epidemics which affect all ages.

The upper right panels of Figures 2–4 illustrate the changing development of the seasonal amplitude over age and time. Starting with a simple inspection *by age*, one recognizes a steady ascent with age for all three causes of death. It has to be noted though, that the gradient is considerably steeper for respiratory diseases than for cardiovascular or cerebrovascular diseases which are rather similar in that respect. This general trend of increasing seasonality with age has been recognized at least since 1838 in Quetelet's analysis of Belgian mortality in the 1830s (Quetelet, 1838). Using parish register data, Wrigley et al. (1997) could even show this slope in England for the time period 1580–1837. Of the few studies, which analyzed seasonality to the highest ages (McDowall, 1981; Näyhä, 1980; Rau, 2007; Rau and Doblhammer, 2003; Robine and Vaupel, 2001), it is apparent that the increase of seasonality appears to be a rather universal phenomenon. For example in a previous study (Rau, 2007), the odds-ratio of dying in winter compared to summer was about 1.1

for 65–69 year-old women in Denmark between 1980 and 1998, the excess risk increased for 90+ aged women to 25% and more. Our findings can be consistently interpreted within the standard framework of mortality: rising mortality with age in general is "a deterioration, or an increased inability to withstand destruction" (Gompertz, 1825, p. 517). A constant environment is an implicit assumption in this statement. If, however, adversity of the environment is seasonally changing and the body is with increasing age less able to cope with stress, an increase in the amplitude in deaths and mortality should be not surprising. The development of the seasonal amplitude *over time* is less straightforward: At younger ages (until $\sim 65$), it appears as if the amplitude decreases until the mid 1970s and increases slightly again. This trend is more pronounced for cardiovascular and cerebrovascular diseases than for respiratory diseases. Although the mainstream of seasonality literature reports a decrease in the seasonal amplitude over time in most countries, the reversal of this trend has been reported before for the United States (Feinstein, 2002; Rau, 2007; Seretakis et al., 1997). How can we interpret these findings? According to Seretakis et al. (1997), the decrease and the subsequent slight increase can be traced back, first, to the spread of central heating which helped reducing excess winter mortality and dampened the seasonal amplitude. The increased usage of air conditioning since the mid-1970s may have helped to reduce absolute mortality during the summer, resulting in larger annual fluctuations in mortality — despite a lower overall level in mortality. Especially for respiratory diseases and to a lesser extent for cardio- and cerebrovascular diseases, we recognize a short-lived peak in the early 1970s. We are not sure whether this sudden increase can be traced back to data problems: the data for 1972 contained not the whole population as in all other years but only a 50% random sample. Alternatively, it can be the outcome of smaller influenza epidemics during the winters 1971/72 and 1972/73 (see, for example, Barker, 1986) with a strong impact on respiratory diseases and on a smaller scale the two circulatory diseases in our

analysis (Madjid et al., 2004; Smeeth et al., 2004; Thompson et al., 2003).

At ages above 70, the only visible progress has been made for respiratory diseases. For cardiovascular and cerebrovascular diseases large seasonal amplitudes appear to extend to younger and younger ages. We have no clear explanation for this development. One can raise the conjecture that it is the outcome of a compositional change in the population: as a consequence of progress made in survival in general over time, rather frail individuals reach higher ages than in the past. Those less robust people are then more susceptible to the harsh climatic conditions during the winter than people in previous years who were, on average, less frail.

# 4 SUMMARY AND CONCLUSION

Our analysis investigated how seasonality in deaths in the United States changed over age and time between 1959 and 1998 for deaths of women from cardiovascular, cerebrovascular, and respiratory diseases. The phase, which marks the day of the annual peak in deaths remained relatively constant for the two causes connected with the circulatory system. The timing of the peak was less stable for respiratory diseases which is most likely the outcome of higher sensitivity to quickly changing environmental conditions of this cause of death category. The amplitude was found to be increasing over age; this is what we expected since previous (one-dimensional) studies reported the same development. The most probable explanation is the decreasing ability of the human body to handle environmental stress with increasing ages. While the literature for several countries points at a decreasing amplitude in seasonality *over time*, we found mixed evidence for the three causes we analyzed. For the first two decades the seasonal amplitude appears to decrease for all three causes. Since then, the amplitude started increasing again for ages up to 70 years. This could be explained

by the increased usage of air conditioning. Its widespread use may have allowed summer mortality to decrease faster than mortality in winter (or mortality in general) resulting in a larger seasonal amplitude in deaths.

As far as we know, our modulation model represents a novel approach to analyze seasonality over age and time *simultaneously*. We have shown that the model can be extended easily to incorporate exposures for the estimation of death rates; similarly, the model can include also higher frequencies than annual rhythms. Although we analyzed the *changes in the seasonal component* over age and time, the method can be obviously employed for *seasonal adjustment* in two dimensions — the typical application of one-dimensional methods such as X-11, BV4 or TRAMO/SEATS. In its current form, the model assumes a Poisson distribution for the observed count data. A possible extension could incorporate the Negative Binomial distribution. This distribution relaxes the assumption of equality of mean and variance of the Poisson distribution. It can, thus, account for overdispersion, an effect known among demographers as "unobserved heterogeneity" (Barron, 1992; Lawless, 1987a,b).

# References

Auerbach, R.D. and Rutner, J.L. (1978). The Misspecification of a Nonseasonal Cycle as a Seasonal by the X-11 Seasonal Adjustment Program. *The Review of Economics and Statistics 60*: 601–603.

Aylin, P., Morris, S., Wakefield, J., Grossinho, A., Jarup, L., and Elliott, P. (2001). Temperature,

housing, deprivation and their relationship to excess winter mortality in Great Britain, 1986–1996. *International Journal of Epidemiology 30*: 1100–1108.

Barker, W.H. (1986). Excess pneumonia and influenza associated hospitalization during influenza epidemics in the United States, 1970–78. *American Journal of Public Health 76*: 761–765.

Barron, D.N. (1992). The Analysis of Count Data: Overdispersion and Autocorrelation. *Sociological Methodology 22*: 179–220.

Cleveland, R.B., Cleveland, W.S., McRae, J.E., and Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics 6*: 3–73.

Cleveland, W.S., Devlin, S.J., and Terpenning, I.J. (1981). *The SABL statistical and graphical methods*. Technical report, Computing Information Library, Bell Labs, Murray Hill, NJ.

Cleveland, W. and Tiao, G. (1976). Decomposition of Seasonal Time Series: A Model for the Census X-11 Program. *Journal of the American Statistical Association 71*: 581–587.

CNN (2003). Heat deaths: Chirac pledges action. Available online at: `http://www.cnn.com/2003/WORLD/europe/08/21/france.heatwave/index.html`.

Currie, I.D., Durban, M., and Eilers, P.H. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling 4*: 279–298.

Dettmann, M. (2006). *Auswirkungen des Rauchverhaltens auf geschlechtsspezifische Unterschiede in der Lebenserwartung*. Master's thesis, Wirtschafts- und Sozialwissenschaftliche Fakultät, Universität Rostock, Rostock, DE.

Dierckx, P. (1993). *Curve and Surface Fitting with Splines*. Monographs on Numerical Analysis. Oxford, UK: Clarendon Press.

Donaldson, G. (2002). Trends in excess winter mortality, by age and sex. Presentation given at the workshop “Seasonality in Mortality”, Duke University, NC, 07–08 March 2002.

Donaldson, G., Ermakov, S., Komarov, Y., and Keatinge, W. (1998a). Cold related mortalities and protection against cold in Yakutsk, eastern Siberia: observation and interview study. *British Medical Journal 317*: 978–982.

Donaldson, G. and Keatinge, W. (1997). Mortality related to cold weather in elderly people in southeast England, 1979-94. *British Medical Journal 315*: 1055–1056.

Donaldson, G., Tchernjavskii, V., Ermakov, S., Bucher, K., and Keatinge, W. (1998b). Winter mortality and cold stress in Yekaterinburg, Russia: interview study. *British Medical Journal 316*: 514–518.

Eilers, P.H.C. and Marx, B.D. (1996). Flexible Smoothing with B-splines and Penalties. *Statistical Science 11*: 89–102.

Eilers, P.H.C. and Marx, B.D. (2002). Generalized linear additive smooth structures. *Journal of Computational and Graphical Statistics 11*: 758–783.

Eilers, P.H., Currie, I.D., and Durbán, M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics and Data Analysis 50*: 62–76.

Eilers, P.H., Gampe, J., Marx, B.D., and Rau, R. (2007). Modulation models for seasonal incidence tables.

Eurowinter Group (1997). Cold exposure and winter mortality from ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe. *Lancet 349*: 1341–1346.

Eurowinter Group (2000). Winter mortality in relation to climate. *International Journal of Circumpolar Health 59*: 154–159.

Feinstein, C.A. (2002). Seasonality of deaths in the U.S. by age and cause. *Demographic Research 6*: 469–486.

Gampe, J. and Rau, R. (2003). Trends in saisonalen Mortalitätsschwankungen — eine Analyse mittels P-splines. Presentation given at the "Deutsche Statistische Woche", Potsdam, Germany, 22 August 2003.

Gampe, J. and Rau, R. (2004). Seasonal Variation in Death Counts: $P$-Spline Smoothing in the Presence of Overdispersion. In A. Bigger, E. Dreassi, C. Lagazio, and M. Marchi (Eds.), *Statistical Modelling. Proceedings of the 19th International Workshop on Statistical Modelling, Florence (Italy), 4–8 July 2004*. Firenze, I: Firenze University Press, pp. 409–409.

Gemmell, I., McLoone, P., Boddy, A., Watt, G., and Dickinson, G. (1999). *Seasonal Variation in Mortality and Morbidity in Scotland 1981–93*. Technical report, University of Glasgow, Public Health Research Unit.

Gemmell, I., McLoone, P., Boddy, F., Dickinson, G.J., and Watt, G. (2000). Seasonal variation in mortality in Scotland. *International Journal of Epidemiology 29*: 274–279.

Gompertz, B. (1825). On the Nature of the Function Expressive of the Law of Human Mortality,

and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London 115*: 513–583.

Grut, M. (1987). Cold-related death in some developed countries. *The Lancet* : 21224 January 1987.

Hastie, T.J. and Tibshirani, R.J. (1987). Generalized Additive Models: Some Applications. *Journal of the American Statistical Association 82*: 371–386.

Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models*. London, UK: Chapman & Hall.

Hastie, T.J. and Tibshirani, R.J. (1993). Varying-Coefficient Models. *Journal of the Royal Statistical Society. Series B (Methodological) 55*: 757–796.

Healy, J. (2003). Excess winter mortality in Europe: a cross country analysis identifying key risk factors. *Journal of Epidemiology and Community Health 57*: 784–789.

Huynen, M.M., Martens, P., Schram, D., Weijenberg, M.P., and Kunst, A.E. (2001). The impact of heat waves and cold spells on mortality rates in the Dutch population. *Environmental Health Perspectives 109*: 463–470.

Johnson, H. and Griffiths, C. (2003). Estimating excess winter mortality in England and Wales. *Health Statistics Quarterly 20*: 9–24.

Kannisto, V., Lauritsen, J., Thatcher, A.R., and Vaupel, J.W. (1994). Reductions in mortality at advanced ages: Several decades of evidence from 27 countries. *Population and Development Review 20*: 793–810.

Keatinge, W. (1986). Seasonal mortality among elderly people with unrestricted home heating. *British Medical Journal 293*: 732–733.

Keatinge, W., Coleshaw, S., and Holmes, J. (1989). Changes in seasonal mortalities with improvement in home heating in England and Wales from 1964 to 1984. *International Journal of Biometeorology 33*: 71–76.

Lawless, J. (1987a). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics 15*: 209–225.

Lawless, J. (1987b). Regression Methods for Poisson Process Data. *Journal of the American Statistical Association 82*: 808–815.

MacKay, J. and Eriksen, M. (2002). *The Tobacco Atlas*. Brighton, UK: World Health Organization.

Madjid, M., Aboshady, I., Awan, I., Litovsky, S., and Casscells, S.W. (2004). Influenza and Cardiovascular Disease: Is There a Causal Relationship? *Texas Heart Institute Journal 31*: 4–13.

Marx, B.D. and Eilers, P.H. (2003). Multidimensional calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems 66*: 159–174.

Marx, B.D. and Eilers, P.H. (2005). Multidimensional Penalized Signal Regression. *Technometrics 47*: 13–22.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. London, UK: Chapman and Hall.

McDowall, M. (1981). Long term trends in seasonal mortality. *Population Trends 26*: 16–19.

McKee, C.M. (1989). Deaths in winter: Can Britain learn from Europe? *European Journal of Epidemiology 5*: 178–82.

Meslé, F. and Vallin, J. (1996). Reconstructing Long-Term Series of Causes of Death. *Historical Methods 29*: 72–87.

Meslé, F. and Vallin, J. (2006). The Health Transition: Trends and Prospects. In G. Caselli, J. Vallin, and G. Wunsch (Eds.), *Demography. Analysis and Synthesis*, volume II, chapter 57. Amsterdam, NL: Elsevier, pp. 247–259.

Nakai, S., Itoh, T., and Morimoto, T. (1999). Deaths from heat-stroke in Japan: 1968–1994. *International Journal of Biometeorology 43*: 124–127.

National Statistics Online (2007). Excess Winter Mortality — Historical Data. Available online at: `http://www.statistics.gov.uk/StatBase/Expodata/Spreadsheets/D7090.xls`.

Näyhä, S. (1980). *Short and medium term variations in mortality in Finland*. Ph.D. thesis, Department of Public Health Service, University of Oulu, Finland.

Oeppen, J. and Vaupel, J.W. (2002). Broken Limits to Life Expectancy. *Science 296*: 1029–1031.

Quetelet, A. (1838). *De l'influence des saisons sur la mortalité aux différens ages dans la Belgique*. Bruxelles, B: M. Hayez.

Rau, R. (2007). *Seasonality in Human Mortality. A Demographic Approach*. Demographic Research Monographs. Heidelberg, Germany: Springer.

Rau, R. and Doblhammer, G. (2003). Seasonal mortality in Denmark. The role of sex and age. *Demographic Research 9*: 197–222.

Raveh, A. (1984). Comments on Some Properties of X-11. *The Review of Economics and Statistics 66*: 343–348.

Robine, J.M. and Vaupel, J.W. (2001). Supercentenarians: slower ageing individuals or senile elderly? *Experimental Gerontology 36*: 915–930.

Robine, J.M. and Vaupel, J.W. (2002). Emergence of supercentenarians in low mortality countries. *North American Actuarial Journal 6*: 54–63.

Sakamoto-Momiyama, M. (1977). *Seasonality in Human Mortality. A Medico-Geographical Study*. Tokyo, JP: University of Tokyo Press.

Sakamoto-Momiyama, M. (1978). Changes in the seasonality of human mortality: A medico-geographical study. *Social Science and Medicine 12*: 29–42.

Seretakis, D., Lagiou, P., Lipworth, L., Signorello, L.B., Rothman, K.J., and Trichopoulos, D. (1997). Changing seasonality of mortality from coronary heart disease. *Journal of the American Medical Association 278*: 1012–1014.

Shiskin, J. (1968). Seasonal Adjustment. In D.L. Sills (Ed.), *International Encyclopedia of the Social Sciences*. New York, NY: The Macmillan Company & The Free Press, pp. 80–88.

Smeeth, L., Thomas, S.L., Hall, A.J., Hubbard, R., Farrington, P., and Vallance, P. (2004). Risk of Myocardial Infarction and Stroke after Acute Infection or Vaccination. *New England Journal of Medicine 351*: 2611–2618.

Smith, D.W. (1998). Changing Causes of Death of Elderly People in the United States, 1950–1990. *Gerontology 44*: 331–335.

Speth, H.T. (2004). *Komponentenzerlegung und Saisonbereinigung ökonomischer Zeitreihen mit dem Verfahren BV4.1*. Methodenberichte, Heft 3, 2004, Statistisches Bundesamt, Wiesbaden, D.

Thompson, W.W., Shay, D.K., Weintraub, E., Brammer, L., Cox, N., Anderson, L.J., and Fukuda, K. (2003). Mortality Associated With Influenza and Respiratory Syncytial Virus in the United States. *Journal of the American Medical Association 289*: 179–186.

Tuljapurkar, S., Li, N., and Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. *Nature 405*: 789–792.

Vaupel, J.W., Carey, J.R., Christensen, K., Johnson, T.E., Yashin, A.I., Holm, N.V., Iachine, I.A., Kannisto, V., Khazaeli, A.A., Liedo, P., Longo, V.D., Zeng, Y., Manton, K.G., and Curtsinger, J.W. (1998). Biodemographic trajectories of longevity. *Science 280*: 855–860.

Vaupel, J.W., Kistowski, K.v., and Rau, R. (2007). Mortality. In *The New Palgrave: A Dictionary of Economics*. Macmillan (forthcoming).

White, K.M. (1999). Cardiovascular and Tuberculosis Mortality: The Contrasting Effects of Changes in Two Causes of Death. *Population and Development Review 25*: 289–302.

Woodhouse, P.R., Khaw, K.T., and Plummer, M. (1993). Seasonal variation of blood pressure and its relationship to ambient temperature in an elderly population. *Journal of Hypertension 11*: 1267–1274.

Wrigley, E., Davies, R., Oeppen, J., and Schofield, R. (1997). *English population history from family reconstitution 1580–1837*. Cambridge, UK: Cambridge University Press, Cambridge Studies in Population, Economy and Society in Past Time No. 32 edition.

Yaffee, R. (2000). *Time Series Analysis and Forecasting with Applications of SAS and SPSS*. San Diego, CA: Academic Press.

Yan, Y.Y. (2000). The influence of weather on human mortality in Hong Kong. *Social Science and Medicine 50*: 419–427.

Table 1: ICD Coding and Counts of Female Deaths

| Cause of Death | ICD Revision and Years | | | Death Counts | % |
|---|---|---|---|---|---|
| | ICD-7 | ICD-8 | ICD-9 | | |
| | 1959–67 | 1968–78 | 1979–98 | | |
| All Causes | — | — | — | 36,202,036 | 100.00 |
| Cardiovascular | 400–469 | 390–429, 440–458 | 390–429, 440–459 | 14,615,664 | 40.37 |
| Cerebrovascular | 330–334 | 430–438 | 430–438 | 3,971,043 | 10.97 |
| Respiratory | 241, 470–527 | 460–519 | 460–519 | 2,430,132 | 6.71 |

Table 2: Count and Proportion for Female Deaths from All Causes, Cardiovascular Diseases, Cerebrovascular Diseases, and Respiratory Diseases for the First Five Years (1959–63) and the Last Five Years (1994–98) of the Observation Period

| Cause of Death | Time Period | | | |
|---|---|---|---|---|
| | 1959–63 | | 1994–98 | |
| | Counts | (%) | Counts | (%) |
| Cardiovascular | 1,565,912 | (41.94) | 2,025,439 | (35.21) |
| Cerebrovascular | 515,517 | (13.81) | 482,009 | (8.38) |
| Respiratory | 178,496 | (4.78) | 565,709 | (9.83) |
| All Causes | 3,733,775 | (100.00) | 5,752,471 | (100.00) |

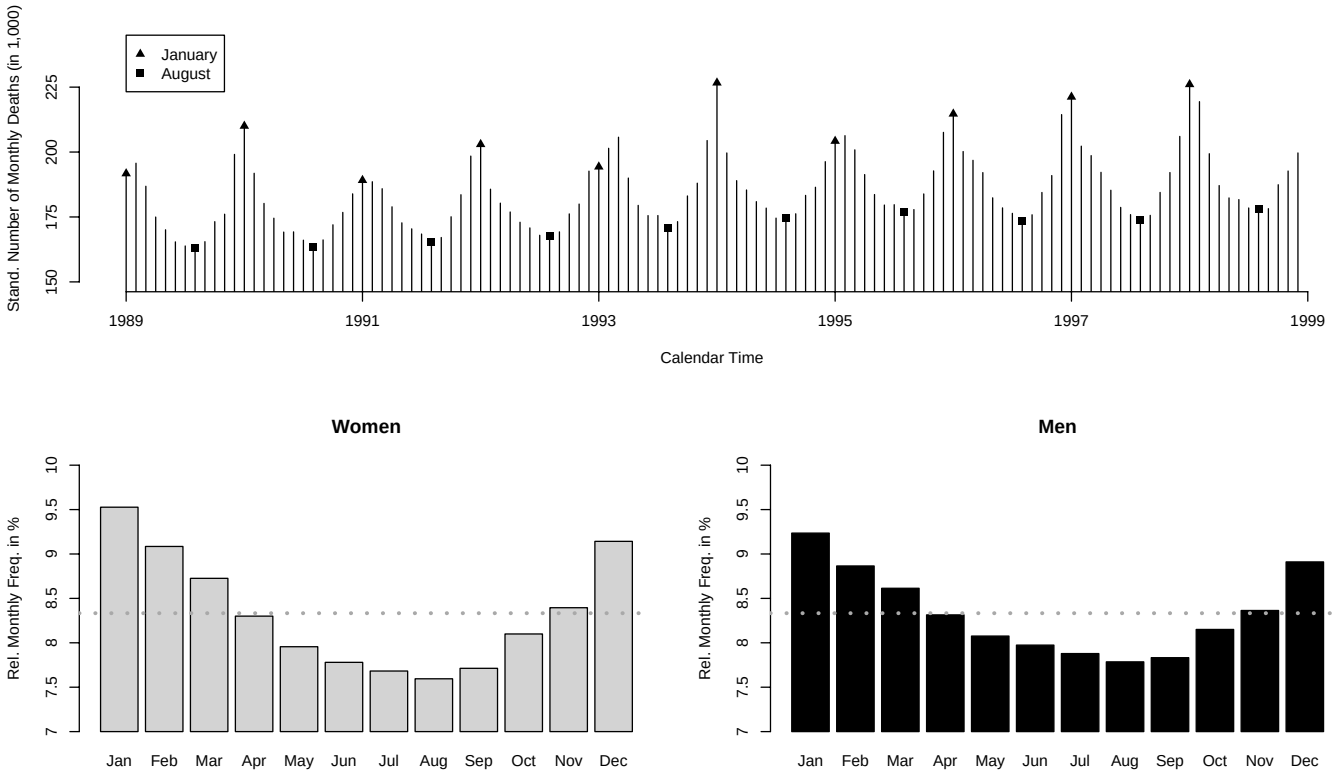


Figure 1: Monthly distribution of deaths in the United States, 1989–98 and its aggregation into twelve months by sex (standardized for length of month)

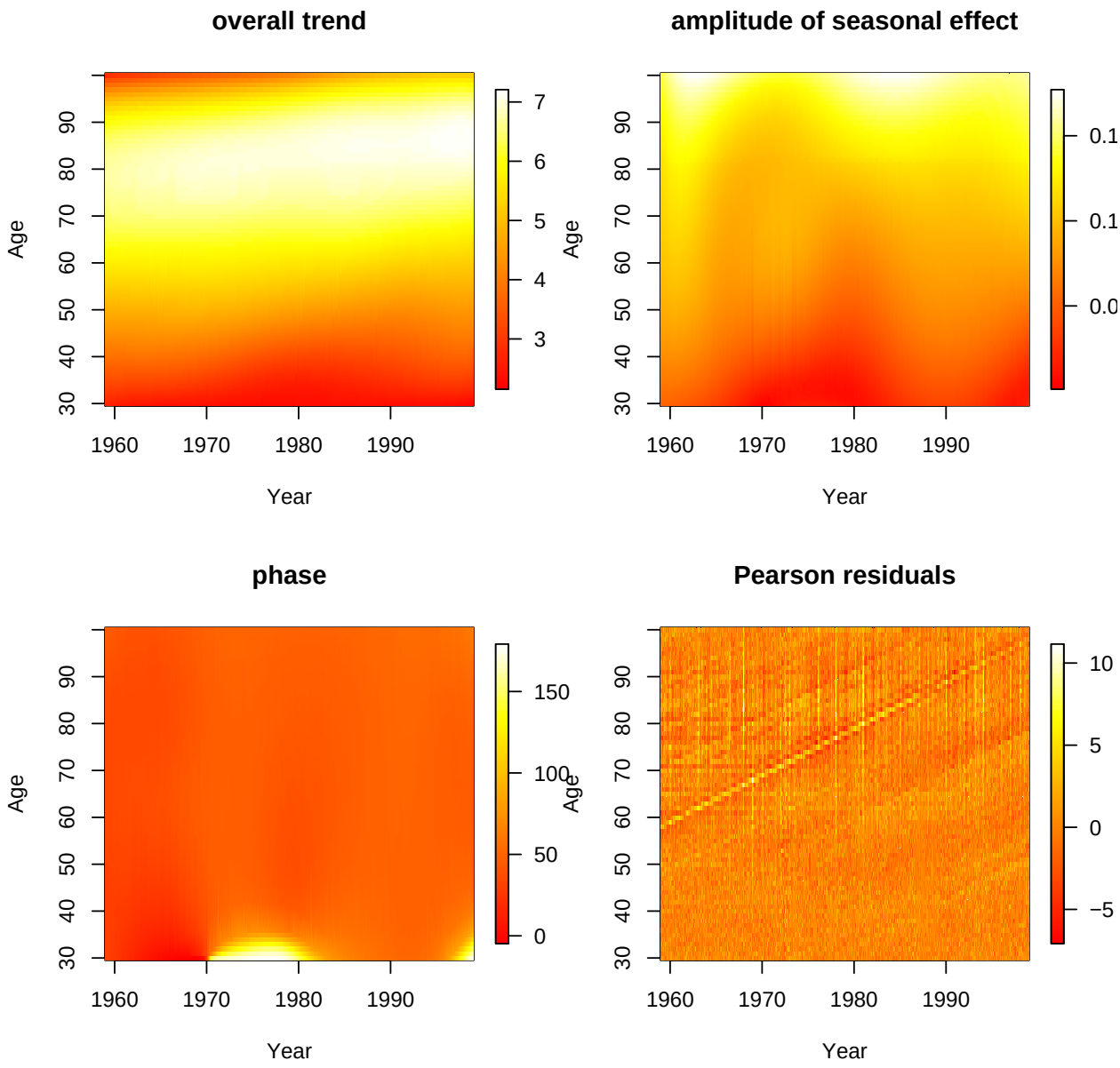


Figure 2: Results for Cardiovascular Mortality, Women, Aged 30–100, 1959–98; Upper Left Panel: Overall Trend, Upper Right: Amplitude of Seasonality, Lower Left: Phase of Seasonality, Lower Right: Pearson Residuals

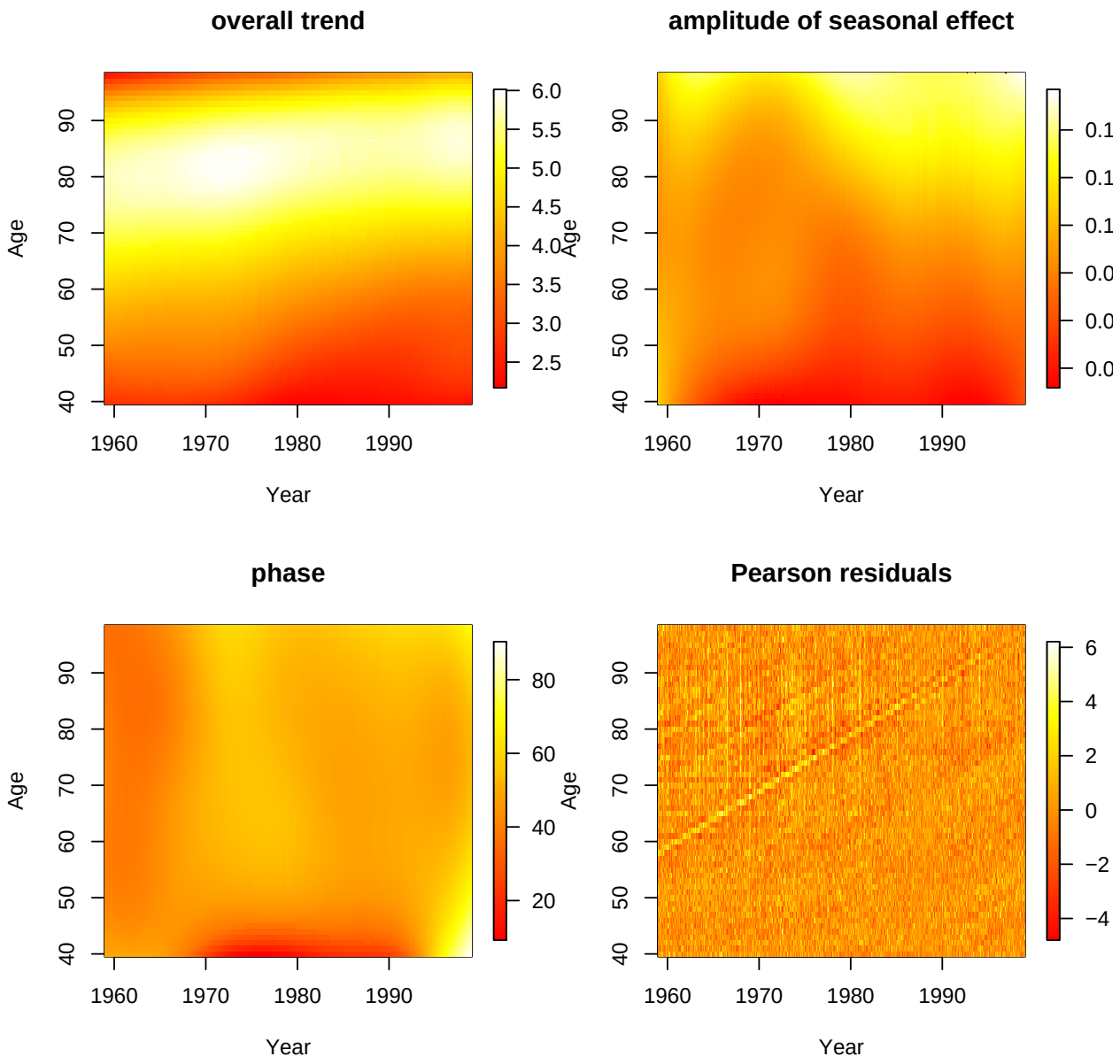


Figure 3: Results for Cerebrovascular Mortality, Women, Aged 40–98, 1959–98; Upper Left Panel: Overall Trend, Upper Right: Amplitude of Seasonality, Lower Left: Phase of Seasonality, Lower Right: Pearson Residuals

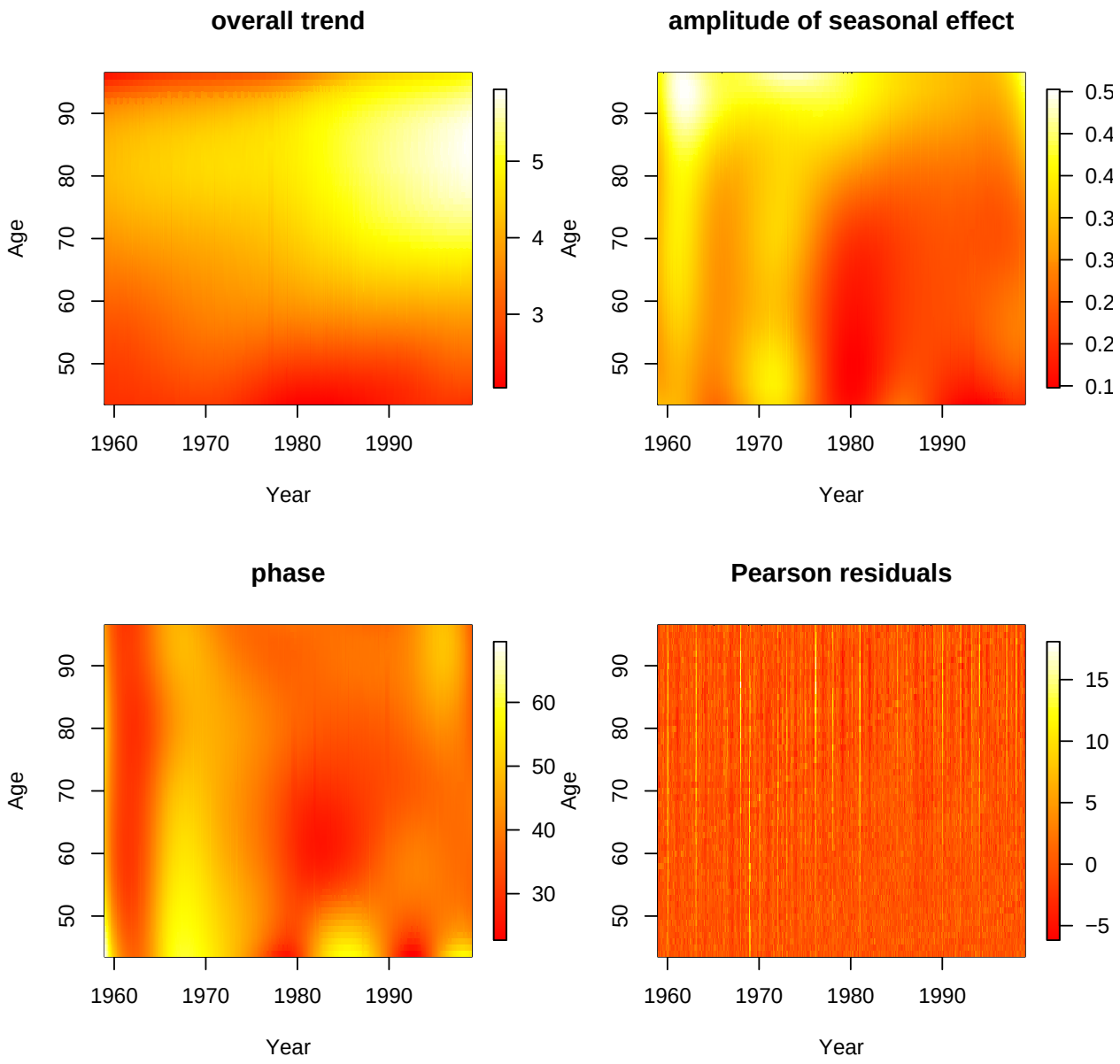


Figure 4: Results for Respiratory Mortality, Women, Aged 44–96, 1959–98; Upper Left Panel: Overall Trend, Upper Right: Amplitude of Seasonality, Lower Left: Phase of Seasonality, Lower Right: Pearson Residuals