A seasonal Markov chain model for the weather in the central Alps

By PAUL JORDAN[1] and PETER TALKNER[2]*, [1]*Universitätsrechenzentrum, Universität Basel, Klingelbergstr. 70, CH 4056 Basel, Switzerland;* [2]*General Energy Research, Paul Scherrer Institut, CH 5232 Villigen, Switzerland*

ABSTRACT

The dynamics of daily weather types according to Schüepp's classification in the Alpine region is investigated by means of seasonal Markov chain models. A logit model is employed for the transition probabilities of the Markov chain. The parameters follow from a maximum likelihood estimation. A 1st- and a 2nd-order Markov chain model are compared and found to yield very similar results. Model predictions are compared with counted frequencies of seasonally-averaged joint probabilities for the occurrence of weather types at subsequent days, monthly-averaged probabilities of a change of the weather type from one to the next day, and daily averages of the probabilities of occurrence of the different weather types. All these predictions coincide with the observations within the statistical limits. The only large deviations occur in the tails of the lengths distributions of uninterrupted episodes of the two most frequent weather types.

1. Introduction

Since the advent of powerful computers, large numerical models have been employed to better understand the dynamics of the weather. The two main goals one hopes to achieve with these models are improved weather predictions and a better knowledge about the sensitivity of the earth's climate system to anthropogenic influences. Today, the global models together with data from satellite and ground-based observations are used to drive local models with a mesh width of about 50 km. In mesoscalic models with a resolution of about 20 km, coarse topographic influences can be included, which may strongly influence the climate and weather as it is obvious in the case of the Alps. On the other hand, in those times when neither computers nor satellite pictures were available, various classification schemes were introduced in order to categorize the daily weather and to ease the prognosis. Though being no longer in use for these purposes, several meteorological services still determine the weather types for each day. Certain classifications have also been reconstructed from historic data as far back as the necessary informations to perform the classification are available. For example, the classification into Grosswetterlagen (Middle European weather types) by Hess and Brezowsky (1977) exists back to 1881. Klaus (1978) and Klaus and Stein (1978) have analysed these data with respect to possible correlations between the variations of monthly frequencies of the different weather types and temperature fluctuations of the North Atlantic. Spekat et al. (1983) proposed Markovian chains as possible models for the dynamics of three circulation patterns and found good agreement between data and a 1st-order Markovian chain.

Schüepp (1979) adapted the Hess Brezowski classification to the western central part of the Alps and its northern foreland. In contrast to the

* Corresponding author.
e-mail: talkner@psi.ch.

classification of Middle European weather types by Hess and Brezowsky (1977), which is based on information of three consecutive days in order to determine the weather type of the middle day, Schüepp's weather types are determined from information pertaining to a single day. These weather types are known for each day back to 1945. Salvisberg (1996) compared this classification with the one of Hess and Brezowski. Stefanicki et al. (1998) investigated the long-time variability of Schüepp's weather types and found pronounced trends in annual frequencies of some of these types. These trends have been shown to be correlated with trends of independently-determined meteorological parameters (see Stefanicki et al., 1998; Wanner et al., 1997; Wanner et al., 1998). This demonstrates the utility of a weather classification for the detection of climatic changes beyond its primary purpose of a short-term prognosis. In a recent publication, Nicolis et al. (1997) studied the dynamics of Schüepp's weather types by comparing 1st-order Markovian models with different non-Markovian models. However, none of the considered models includes the strong seasonality of the frequencies with which the different types occur. This seasonality has already been noticed by Schüepp (1979) and was further investigated by Baraldi (1994).

In the present paper, the dynamics of Schüepp's weather types is modelled by 1st- and 2nd-order seasonal Markov chains. Since, in a seasonal model, the transition probabilities periodically vary in time, even for the smallest possible number of three weather types there are not enough data to simply estimate the six independent transition probabilities of the 1st-order Markov chain for each day of the year from counted relative frequencies. Therefore we use a generalized linear model which, depending on the choice of parameters yields the transition probabilities of a 1st- or 2nd-order Markov chain and allows for seasonality. The parameters are determined such that the likelihood of the models becomes maximal.

In Section 2, we give a short description of the used data, followed in Section 3 by a presentation of the employed stochastic models. In Section 4, the logit model for the conditional probabilities which specify our stochastic models is reviewed. Using this model we map the transition probabilities which take values between 0 and 1 onto functions with unresticted values. These functions are assumed to depend on the possible causes in a linear way. The likelihood of the model is given in Section 5. The casual reader who is not interested in the details of the statistical analysis may skip Sections 4, 5 and the first half of 6 and continue reading after eq. (6.3). Section 6 contains the model selection and a discussion of the qualitative behavior of the obtained transition probabilities of the 2nd-order Markov chain model. In Section 7, the long-time asymptotic behavior for the resulting two-time joint probabilities of the 1st- and 2nd-order Markov chain model is compared with each other on a daily basis and for monthly averages with direct estimates from the data. Furthermore, the asymptotic single time probabilities resulting from the models are compared with the relative frequencies of the weather types on a daily basis. Moreover, the length statistics of uninterrupted episodes in a single type is investigated. Finally the predictive power of the model is tested on data of the years 1995 till 1997 by means of hit rates and skill scores. The paper closes with a Summary.

2. The data

The data consist of the daily weather types from 1 January 1945 until 31 December 1994 according to Schüepp's coarsest classification scheme (Schüepp, 1979). The classification is made by the Swiss Meteorological Institute on the basis of the ground pressure distribution and the height of the 500 hPa surface both taken at 12 GMT for an area of 440 km diameter covering the western central part of the Alps. There are the advective, the convective and the mixed type. The advective type is characterized by large (horizontal) pressure gradients and pronounced pressure driven winds. Accordingly this type can be further subdivided into subclasses with westerly, northerly, easterly, and southerly winds. The convective type has small pressure gradients and can be subdivided into anticyclonic, indifferent and cyclonic situations. Roughly speaking, those situations that do not fall in either of the two main types, make the mixed type. A further distinction into 40 groups is possible but here we only deal with the three main types. Their relative frequencies of occurrence can be counted from the data. One finds

$$p_1 = 0.5, \qquad p_2 = 0.44, \qquad p_3 = 0.06, \tag{2.1}$$

where the indices 1, 2, and 3 refer to the convective, advective, and mixed type, respectively. The frequencies of occurrence strongly depend on the season of the year (Schüepp, 1979). In summer, the convective and in winter the advective type is more frequent, see Fig. 1.

3. Seasonal Markov chains

In contrast to a time-homogeneous Markov chain, the transition probabilities $p_n(i|j) = \text{prob}(Y_{n+1} = c_i | Y_n = c_j)$ of a *seasonal* Markov chain periodically depend on time n, i.e.,

$$p_{n+N}(i|j) = p_n(i|j), \tag{3.1}$$

where Y_n is a realization of the process at day n, c_i and c_j are states of the process, i.e., in our case 2 of the 3 weather types, and N is the period which here is one year, i.e., $N = 365.25$. Otherwise, the defining property of a seasonal Markov process is the same as for any other Markov process, namely that probabilities that are multiply conditioned in the past only depend on the most recent condition, see, e.g., Cox and Miller (1965).

Hence, for a 1st-order Markovian chain the probability $p_n(i) = \text{prob}(Y_n = c_i)$ of having type c_i at day n evolves in time according to

$$p_{n+1}(i) = \sum_j p_n(i|j)p_n(j). \tag{3.2}$$

Starting with arbitrary initial probabilities at a day n_0, the time evolution according to (3.2) leads to asymptotic probabilities $p_n^{\text{as}}(i) = \lim_{k\to\infty} p_{n+kN}$ that are periodic in time:

$$p_{n+N}^{\text{as}}(i) = p_n^{\text{as}}(i). \tag{3.3}$$

In the cases that we are going to study these periodic asymptotic probabilities are independent of the initial probabilities and, as shown below, are approximately reached after a week with an accuracy better than 1%. They correspond to the stationary probabilities in case of a homogeneous Markov chain.

For a seasonal 2nd-order Markov chain, the

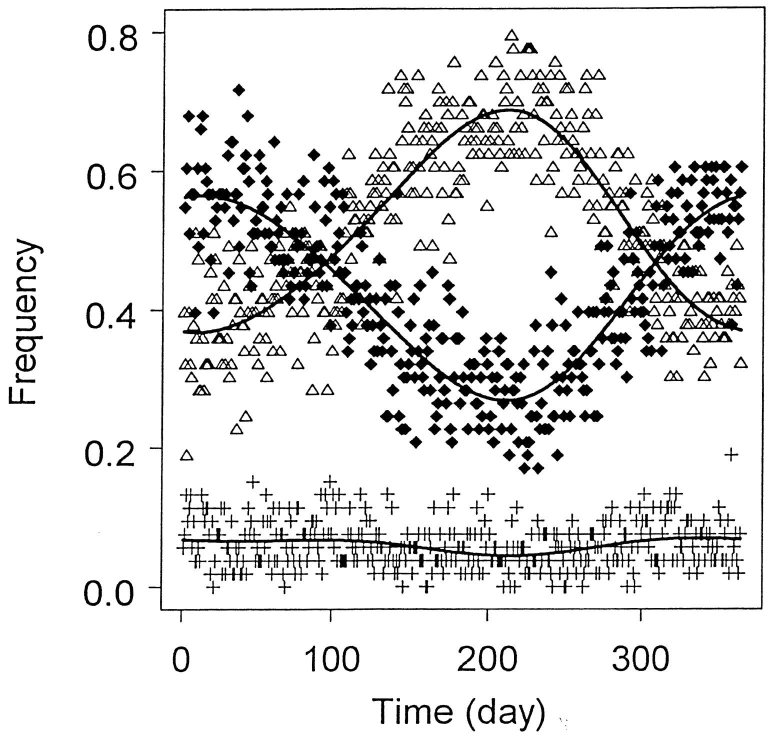


Fig. 1. Daily relative frequencies of the convective (△), advective (◆) and mixed (+) weather types. The curves show the corresponding probabilities following from the 2nd-order Markov chain model.

transition probability $p_{n(i|j,k)} = \mathrm{prob}(Y_{n+1} = c_i | Y_n = c_j, Y_{n-1} = c_k)$ of the type at day $n+1$ depends on both the types at day n and that at the previous day $n-1$. Further it is a periodic function of time n:

$$p_{n+N}(i|j,k) = p_n(i|j,k). \tag{3.4}$$

The joint probability $p_n(i,j) = \mathrm{prob}(Y_n = c_i, Y_{n-1} = c_j)$ of having type c_i on day n and c_j on day $n-1$ evolves in time according to:

$$p_{n+1}(i,j) = \sum_k p_n(i|j,k) p_n(j,k). \tag{3.5}$$

Furthermore, this equation has a periodic solution $p_{n+N}^{\mathrm{as}}(i,j) = p_n^{\mathrm{as}}(i,j)$ that is approached after sufficiently many iterations of equation (3.5).

4. Logit model for a nominal dependent variable

In the framework of ordinary linear models, a vector of observations $\boldsymbol{y}$ having m components $y_1, \dots, y_m$ is assumed to be a realization of a continuous random variable $\boldsymbol{Y}$ with a specific expectation $\boldsymbol{\mu}$.

The systematic part of the model is a specification for the vector $\boldsymbol{\mu}$ in terms of a small number of unknown parameters $\beta_1, \dots, \beta_p$ and other observed quantities $\boldsymbol{x}_i$ (categorical or continuous) by means of which one tries to predict the observations $\boldsymbol{y}$. In the case of ordinary linear models (regression and analysis of variance), this specification takes the form

$$\boldsymbol{\mu} = \sum_{j=1}^{p} \boldsymbol{x}_j \beta_j. \tag{4.1}$$

The generally unknown parameters β_j have to be estimated from the data. If we let i index the observations then the systematic part of the model may be written

$$E(Y_i|\boldsymbol{x}_i) = \mu_i = \sum_{j=1}^{p} x_{ij}\beta_j, \qquad i = 1, \dots, n. \tag{4.2}$$

In matrix notation (where $\boldsymbol{\mu}$ is $n \times 1$, $\boldsymbol{X}$ is $n \times p$ and $\boldsymbol{\beta}$ is $p \times 1$) we may write

$$\boldsymbol{\mu} = \boldsymbol{X\beta}, \tag{4.3}$$

where $\boldsymbol{X}$ is the model or design matrix and $\boldsymbol{\beta}$ is the vector of parameters. The vector $\boldsymbol{X\beta}$ is called the *linear predictor* $\boldsymbol{\eta}$. For classical linear models, the *link* between the systematic and the random component is the identity $\boldsymbol{\mu} = \boldsymbol{\eta}$.

In our case the variable of interest is restricted to a finite set of K distinct categories ($K = 3$ possible values corresponding to our 3 weather-types). Such problems can be adequately treated within the class of Generalized Linear Models (Agresti, 1990). Generalized linear models are an extension of classical linear models: A vector of observations $\boldsymbol{y}$ having n components is assumed to be a realization of a random variable $\boldsymbol{Y}$ with components Y_i which may take categorical values, e.g., a particular weather type at day n, $n = 1, \dots, m$.

Any observation y_n can be coded as a $(K-1)$-tuple whose j-th component is 1 if observation y_i belongs to category c_j and all other components are 0. If y_n belongs to category c_K all $K-1$ components of the tuple are set to 0. Therefore, we are dealing with a generalized Bernoulli trial with a specific expectation $\boldsymbol{\mu}$. The expectation μ_j is then the probability of occurrence of category c_j. This is the same parameterization as for categorical explanatory variables x which is widely used in classical linear models (Analysis Of Variance, ANOVA).

Now, the generalization is the choice of an other link function

$$\eta_j = g(\mu_j). \tag{4.4}$$

For a Bernoulli trial, a link must map the possible values $\mu_i \in (0, 1)$ on the whole real line. An appropriate and theoretically well established choice out of the exponential family is the logit and the generalized logit as link function (Agresti, 1990).

In the case of two categories c_1 and c_2 we have only one probability μ to estimate. With the link function $g(\mu) = \ln(\mu/(1-\mu))$ we have:

$$\ln\left(\frac{\mathrm{prob}(Y = c_1|X)}{\mathrm{prob}(Y = c_2|X)}\right) = \ln\left(\frac{\mathrm{prob}(Y = c_1|X)}{1 - \mathrm{prob}(Y = c_1|X)}\right)$$

$$= \mathrm{logit}\{\mathrm{prob}(Y = c_1|X)\} = X\beta. \tag{4.5}$$

The straightforward generalization in the case of K outcome categories $c_1, \dots, c_K$ is

$$\ln\left(\frac{\mathrm{prob}(Y = c_i|X)}{\mathrm{prob}(Y = c_j|X)}\right) = X\beta_i - X\beta_j. \tag{4.6}$$

This is equivalent to

$$\mathrm{prob}(Y=c_j|X)=\frac{\exp(X\beta_j)}{\sum_{i=1}^{K}\exp(X\beta_i)}, \qquad j=1,\dots,K. \tag{4.7}$$

In order to make the parameters unique we impose the (arbitrarily chosen) constraint

$$\beta_K=0. \tag{4.8}$$

This means that category c_K is the reference group.

5. Parameter estimation

For a multinomial Bernoulli trial the likelihood of the statistical model is defined as

$$L=L(\beta)=\prod_{n=1}^{m}\prod_{j=1}^{K}(\mathrm{prob}(Y_n=c_j))^{Z_n^j}$$
$$=\prod_{n=1}^{m}\prod_{j=1}^{K}\left(\frac{\exp(X\beta_j)}{\sum_{k=1}^{K}\exp(X\beta_k)}\right)^{Z_n^j}, \tag{5.1}$$

where

$$Z_n^j=\begin{cases}1 & \text{if } Y_n=c_j,\\ 0 & \text{if } Y_n\neq c_j.\end{cases} \tag{5.2}$$

The so-called maximum-likelihood estimator (ML) $\hat{\beta}$ of β is the value where $L(\beta)$ takes its maximum. ML-estimators have been the best established estimators in applied and theoretical statistics and their properties are well known (Casella et al., 1990).

6. Model selection

A good statistical model should have parsimonic properties: explaining the most of the variability of the dependent variable with the least number of explanatory variables. As the number of possible explanatory variables and their combinations is practically infinite the following stepwise procedure provides reasonable results: At each step of the process a continuous or categorical variable is added to or removed from the model. As the change of log-likelihood from one step to the next is asymptotically χ^2-distributed one can use this statistic to stop the step procedure.

Table 1 shows the choice of independent variables as specified by their sources for a 1st- and a 2nd-order Markov chain model. These are a constant, the category of today and of yesterday (only needed for the 2nd-order Markov chain) as well as cosine and sine functions with one year and half a year period which should describe the seasonality of the process. Further the pairwise interactions of these variables are taken into account. The corresponding parameters are denoted by Greek letters: α for the constant, β and γ for the the influence of today and yesterday, respectively, and λ for the seasonality. The upper left index denotes the outcome categories $i=1,2$. For the outcome category 3 all parameters are zero. Lower indices at β and γ characterize the categories of today and yesterday and may be 2 or 3. Here the parameters vanish for category 1. The lower indices c and s at λ refer to cosine and sine functions, respectively, with a period one year in case of index 1 and half a year for the index 2. The concatenation of two symbols denotes the parameter for the respective interaction. For the sake of convenience, the variables described by their sources and the corresponding parameters are collected in Table 2.

All calculations were performed by means of the program PR within the BMDP package (Dixon, 1992).

The following example should clarify the construction of the linear predictor η. For convenience we consider a simplified model which only takes into account a constant influence, the category at the previous day, a single periodic function describing the seasonality, and an interaction of the seasonality and the influence of the class at the previous day, see Table 3.

For this toy model, the probability $\mathrm{prob}(Y_n=c_2|Y_{n-1}=c_1)$ of finding c_2 for day n if the category at day $n-1$ was c_1, is determined by the predictor η reading:

$$\eta={}^2\alpha+{}^2\beta_1+({}^2\lambda+{}^2(\lambda\beta_1))\cos\frac{2\pi n}{365.25}. \tag{6.1}$$

Using eq. (4.7), the probability of this event becomes:

$$\mathrm{prob}(Y_n=c_2|Y_{n-1}=c_1)$$
$$=N\exp\left\{{}^2\alpha+{}^2\beta_1+({}^2\lambda+{}^2(\lambda\beta_1))\cos\frac{2\pi n}{365.25}\right\}, \tag{6.2}$$

Table 1. *Parameters of the 1st- and 2nd-order Markov models*

	1st order		2nd order	
Parameter	estimator	std. err.	estimator	std. err.
${}^1\alpha$	2.563	0.0488	2.672	0.0626
${}^2\alpha$	1.753	0.0509	1.722	0.0656
${}^1\beta_2$	−0.816	0.0709	−0.821	0.107
${}^1\beta_3$	−1.734	0.103	−1.828	0.154
${}^2\beta_2$	0.4722	0.0713	0.5281	0.107
${}^2\beta_3$	−0.7018	0.102	−0.6622	0.151
${}^1\gamma_2$			−0.2705	0.108
${}^1\gamma_3$			−0.3644	0.217
${}^2\gamma_2$			0.1197	0.112
${}^2\gamma_3$			−0.1176	0.226
${}^1(\beta_2\gamma_2)$			0.1634	0.154
${}^1(\beta_2\gamma_3)$			0.3402	0.227
${}^1(\beta_3\gamma_2)$			−0.04339	−0.154
${}^1(\beta_3\gamma_3)$			0.2460	0.334
${}^2(\beta_2\gamma_2)$			−0.1107	0.115
${}^2(\beta_2\gamma_3)$			−0.08276	0.223
${}^2(\beta_3\gamma_2)$			−0.2414	0.284
${}^2(\beta_3\gamma_3)$			0.06585	0.330
${}^1\lambda_{c1}$	−0.3344	0.0696	−0.2705	0.0802
${}^1\lambda_{s1}$	−0.1825	0.0683	−0.1546	0.0771
${}^1\lambda_{c2}$			0.01783	0.0462
${}^1\lambda_{s2}$			0.08505	0.0464
${}^2\lambda_{c1}$	0.1244	0.0725	0.1067	0.0831
${}^2\lambda_{s1}$	0.03460	0.0712	0.04701	0.0801
${}^2\lambda_{c2}$			0.01748	0.0464
${}^2\lambda_{s2}$			0.03198	0.0467
${}^1(\lambda_{c1}\beta_2)$	−0.05929	0.100	−0.05623	0.106
${}^1(\lambda_{c1}\beta_3)$	0.00039	0.146	0.01362	0.151
${}^2(\lambda_{c1}\beta_2)$	0.01921	0.101	0.01306	0.106
${}^2(\lambda_{c1}\beta_3)$	−0.08274	0.144	−0.05632	0.149
${}^1(\lambda_{s1}\beta_2)$	0.09207	0.0980	0.06345	0.103
${}^1(\lambda_{s1}\beta_3)$	−0.05381	0.143	−0.03059	0.146
${}^2(\lambda_{s1}\beta_2)$	0.0529	0.0980	0.05257	0.103
${}^2(\lambda_{s1}\beta_3)$	−0.1571	0.141	−0.1075	0.144
${}^1(\lambda_{c1}\gamma_2)$			−0.05055	0.102
${}^1(\lambda_{c1}\gamma_3)$			−0.4102	0.181
${}^2(\lambda_{c1}\gamma_2)$			0.07926	0.103
${}^2(\lambda_{c1}\gamma_3)$			−0.3206	0.180
${}^1(\lambda_{s1}\gamma_2)$			0.06685	0.0996
${}^1(\lambda_{s1}\gamma_3)$			−0.4715	0.174
${}^2(\lambda_{s1}\gamma_2)$			0.04531	0.0999
${}^2(\lambda_{s1}\gamma_3)$			−0.4623	0.173

where N denotes the normalization:

$$N = \left(\exp\left\{ {}^1\alpha + {}^1\beta_1 + ({}^1\lambda + {}^1(\lambda\beta_1)) \cos\frac{2\pi n}{365.25} \right\} + \exp\left\{ {}^2\alpha + {}^2\beta_1 + ({}^2\lambda + {}^2(\lambda\beta_1)) \cos\frac{2\pi n}{365.25} \right\} + 1 \right)^{-1}. \tag{6.3}$$

The transition probabilities for the 1st- and 2nd-order Markov chain models are analogously defined in terms of the parameters given in Table 1, and are shown in Fig. 2. The dependence of the transition probabilities on the previous day is in general considerably weaker than on the present day. The probability $p_n(1|1, j)$ to stay in the convective type c_1 is larger in summer than in winter and always larger than the probability of a change from c_1 to the advective, c_2, or the mixed

Table 2. *Sources and corresponding parameters and their numbers of degrees of freedom (d.f.) of the full model; the abbreviation i.a. stands for interaction*

Parameter	Source	d.f.
${}^{i}\alpha$	constant	1
${}^{i}\beta$	category at day (−1)	2
${}^{i}\gamma$	category at day (−2)	2
${}^{i}(\beta\gamma)$	i.a. category at day (−1) — category at day (−2)	4
${}^{i}\lambda_{c1}$	coeff. of $\cos \frac{2\pi n}{365.25}$	1
${}^{i}\lambda_{s1}$	coeff. of $\sin \frac{2\pi n}{365.25}$	1
${}^{i}\lambda_{c2}$	coeff. of $\cos \frac{4\pi n}{365.25}$	1
${}^{i}\lambda_{s2}$	coeff. of $\sin \frac{4\pi n}{365.25}$	1
${}^{i}(\lambda_{c1}\beta)$	i.a. category at day (−1) — coeff. of $\cos \frac{2\pi n}{365.25}$	2
${}^{i}(\lambda_{s1}\beta)$	i.a. category at day (−1) — coeff. of $\sin \frac{2\pi n}{365.25}$	2
${}^{i}(\lambda_{c1}\gamma)$	i.a. category at day (−2) — coeff. of $\cos \frac{2\pi n}{365.25}$	2
${}^{i}(\lambda_{s1}\gamma)$	i.a. category at day (−2) — coeff. of $\sin \frac{2\pi n}{365.25}$	2

Table 3. *Sources and corresponding parameters and their numbers of degrees of freedom (d.f.) of the toy model*

Parameter	Source	d.f.
${}^{i}\alpha$	constant	1
${}^{i}\beta$	category at day (−1)	2
${}^{i}\lambda$	coeff. of $\cos \frac{2\pi n}{365.25}$	1
${}^{i}(\lambda\beta)$	i.a. category at day (−1) — coeff. of $\cos \frac{2\pi n}{365.25}$	2

type, c_3. The transition probabilities $P(2|1, j)$ are higher in winter than in summer. Those from c_1 to c_3 are rather small all over the year. The transition probabilities $p(2|2, j)$ to stay in c_2 are larger in winter than in summer while transitions from c_2 to c_1 are less frequent in winter than in summer where there even is a short period during which they are more probable than to stay in c_2. Again transitions to c_3 are always rather rare. The seasonal variations of the transition probabilities out of the mixed type c_3 behave in a similar way as those starting in the advective type c_2. There is also a short period in summer in which transitions to the convective type c_1 are more frequent than to the advective type c_2. In winter, the probabilities to stay in c_3 are about equally large as those of going from c_3 to c_1. The probabilities $p(3|3, j)$ to stay in c_3 are approximately twice as large as those to enter c_3 from c_1 or c_2.

If the dependence on the previous day is neglected, a 1st-order Markov chain model results for which the parameters and their standard errors

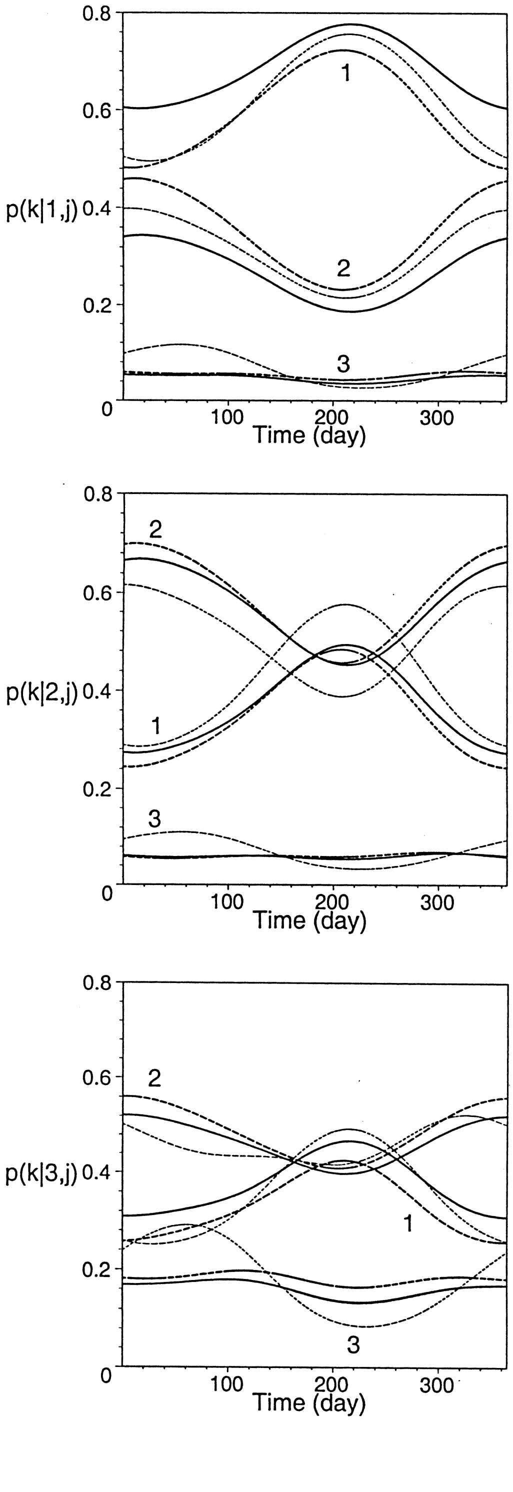


Fig. 2. The seasonal variations of the conditional probabilities $p_n(k|i, j)$. In each panel the type i at the middle day n is specified. The resulting nine curves form groups that are labeled by the type k of the following day $n+1$. Each group consists of three curves specified by the type of the previous day $n-1$ where the solid lines indicate $j=1$, the thin dotted lines $j=2$ and the thick dotted lines $j=3$.

are also given in Table 1. The corresponding transition probabilities are shown in Fig. 3. By definition, they are independent of the type at the penultimate day but otherwise they show the same qualitative seasonal behavior as the transition probabilities of the 2nd-order Markov chain model. A quantitative comparison of the models will be given below.

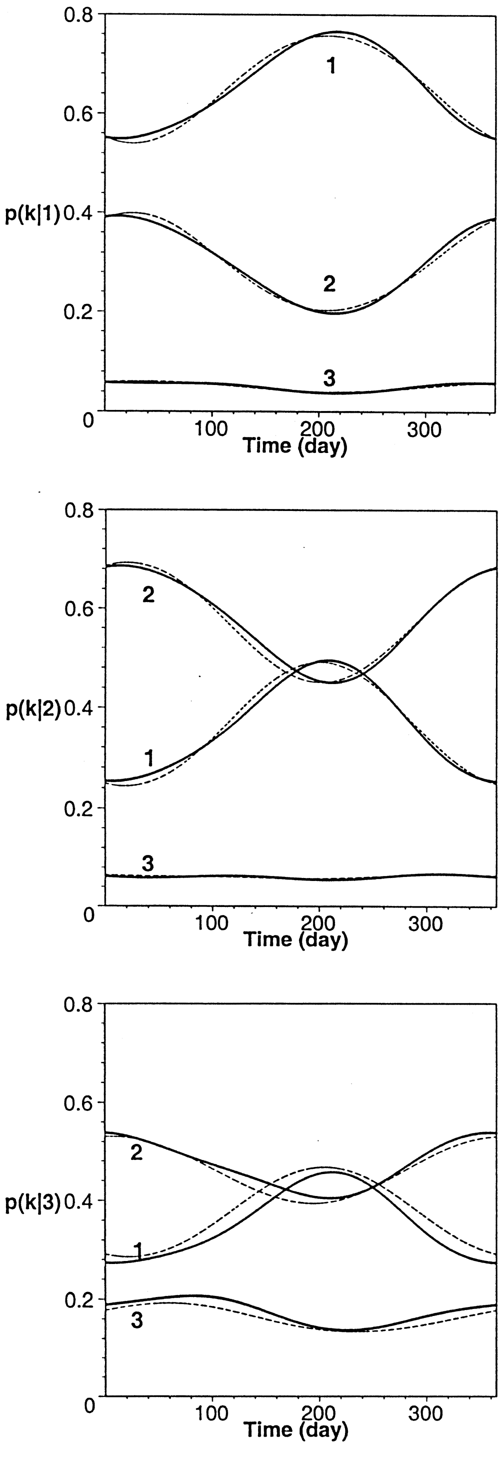
For both models most of the conditional probabilities show strong seasonal variations. For example the transition from c_2 to c_1 is in summer almost twice as frequent as in winter.

7. Comparison of models and data

7.1. Asymptotic probabilities

For the 2nd-order Markov model, the time evolution of the joint probabilities $p_n(j, k)$ is given by eq. (3.5). This is a linear equation which, therefore can be written in vector notation. The particular form of the 9×9-coefficient-matrix depends on the way the pairs of states (j, k) are encoded by a single index. The eigenvalues of this matrix, however, are independent of the particular coding. The total probability is conserved and consequently one eigenvalue is unity. All others have absolute values which are smaller than unity. Fig. 4 shows the 2nd largest eigenvalue as it varies within a year. For larger times, it continues periodically. In our particular case this eigenvalue always is less than one half. This means that on the average after at most five days any initial deviation has died out and an asymptotic state is reached. In contrast to the remaining eigenvalues (not shown here) with smaller modulus it stays real all the time. Since the transition probabilities themselves change with time due to seasonality the asymptotic joint probabilities depend on time in a periodic way with a period of a year, see Fig. 5.

From the available 50 years of data, it is not possible to estimate these joint probabilities from



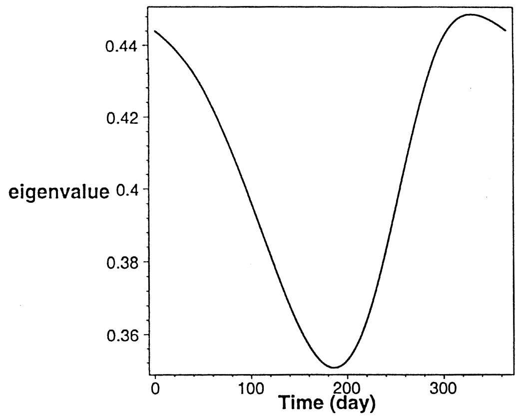


Fig. 4. The eigenvalue of the transition probability of the 2nd-order Markov chain model with largest absolute value different from unity for each day of the year.

relative frequencies on a daily basis. However, for conveniently chosen periods of times as for example for the summer (May until October) or the winter (November until April) months one can reliably estimate relative frequencies for the joint occurrence of pairs of weather types and compare them with the joint probabilities of the respective pairs averaged over the chosen period of time. Table 4 shows a comparison for half year averages. In all cases the agreement is excellent both for the 1st- and 2nd-order Markov chain model.

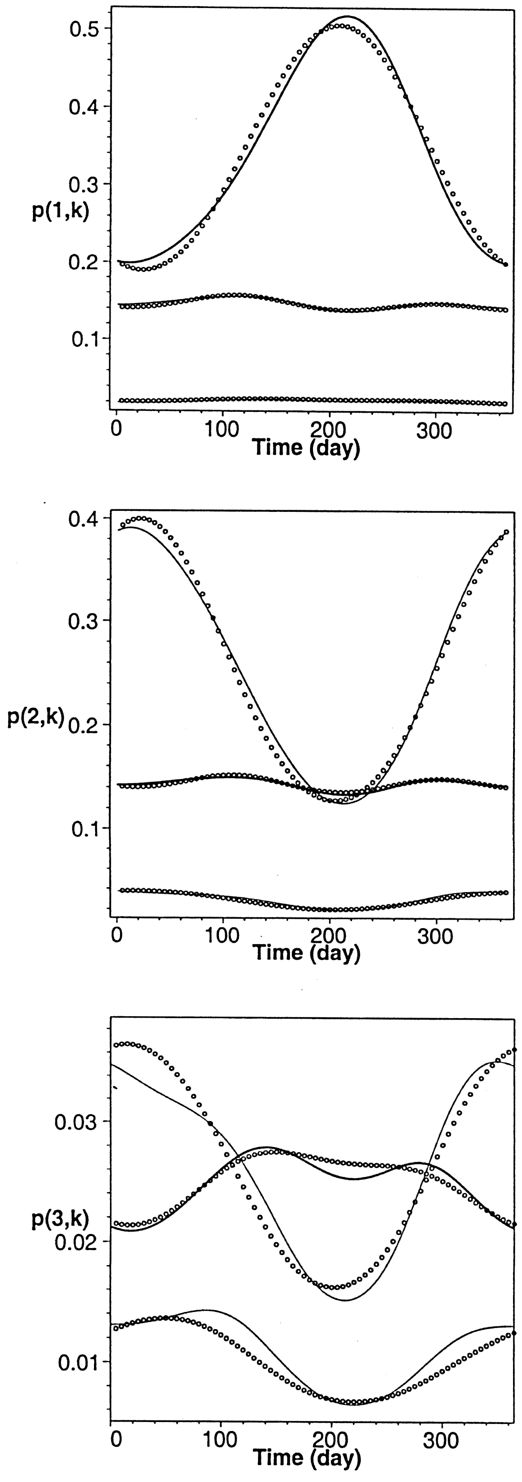
The marginal asymptotic probabilities $p_n^{\mathrm{as}}(i)$ following from the joint probabilities,

$$p_n^{\mathrm{as}}(i) = \sum_j p_n^{\mathrm{as}}(i, j), \tag{7.1}$$

as well as the corresponding relative frequencies obtained from the data are shown in Fig. 1. The agreement again is good.

Within the 1st-order Markov chain model the asymptotic marginal probabilities $p_n^{\mathrm{as}}(i)$ follow from eqs. (3.2) and (3.3). They are shown together with the respective probabilities following from

Fig. 3. The seasonal variations of the conditional probabilities $p_n(k|i)$. The initial day is specified for each panel. The number k next to each curve refers to the respective type. The solid lines show the probabilities of the 2nd-order Markov chain model and the broken lines those of the 1st-order model. The differences between the 1st- and 2nd-order models are rather small for all probabilities $p(k|i)$ with conditions $i = 1, 2$ and somewhat larger if $i = 3$.



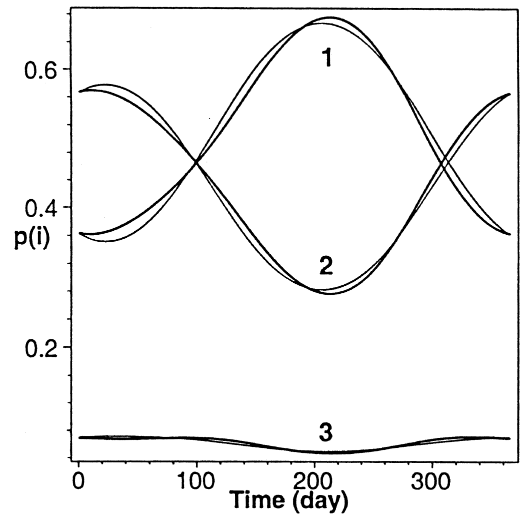


Fig. 6. The asymptotic probabilities $p_n^{as}(i)$ as they result from the 1st- and 2nd-order Markov chain models displayed by thin broken and solid lines, respectively. The corresponding types are indicated by the numbers.

the 2nd-order model in Fig. 6. In Fig. 3, the conditional probabilities

$$p_n^{as}(i|j) = p_{n+1}^{as}(i,j) \Big/ \sum_i p_{n+1}^{as}(i,j) \tag{7.2}$$

are compared with the corresponding transition probabilities of the 1st-order Markov chain model. For all possible pairs of types the 1st- and 2nd-order Markov chain models give very similar results.

7.2. The distribution of lengths of episodes

The time series of weather types contains *episodes* of variable length during which the weather type does not change. The relative frequency of

Fig. 5. The asymptotic joint probabilities $p_n(i,k)$ are shown in three separate panels for $i = 1, 2, 3$. Results of the 2nd-order Markov model are displayed as full lines corresponding to type $k = 1$, thick broken to $k = 2$, and, the lowest, thin broken lines to $k = 3$. The latter is only visible in the panel with $i = 3$; for all i, $k = 3$ has the smallest probability. The small circles represent the results of the 1st-order Markov model where the class k on the previous day coincides with that of the neighboring curve with similar seasonal variation stemming from the 2nd-order Markov model.

Table 4. *Summer and winter averages of the joint probabilities p(i, j) of type j on one day and type i on the consecutive day following from the 1st- and 2nd-order Markov chain models compared with the corresponding observed relative frequencies*

	Summer			Winter		
	Markov model			Markov model		
Probability	1st order	2nd order	observed	1st order	2nd order	observed
$p(1, 1)$	0.443	0.440	0.452	0.240	0.240	0.242
$p(1, 2)$	0.145	0.145	0.144	0.147	0.149	0.147
$p(1, 3)$	0.023	0.022	0.023	0.022	0.020	0.022
$p(2, 1)$	0.142	0.141	0.140	0.145	0.146	0.146
$p(2, 2)$	0.168	0.171	0.164	0.342	0.341	0.340
$p(2, 3)$	0.023	0.024	0.023	0.035	0.035	0.034
$p(3, 1)$	0.027	0.026	0.027	0.023	0.023	0.023
$p(3, 2)$	0.020	0.020	0.019	0.033	0.032	0.033
$p(3, 3)$	0.008	0.009	0.007	0.012	0.013	0.013

the episodes which are characterized by their type and length l are shown in Fig. 7 in a semi-logarithmic plot. They are compared with the probabilities of the corresponding episodes as they result from the 1st- and 2nd-order Markov models. See Section 10 for explicit expression of these probabilities following from the models. The agreement between the models and observations is good for episodes with a length up to 10 days. Longer episodes occur with a systematically higher probability by up to a factor of ten for the most rare ones compared to the predictions of the Markov models. These deviations clearly indicate that there are long-time memory effects which are not contained in the present Markov chain models. Nicolis et al. (1997) proposed stretched exponential distributions of the episode lengths which indeed yield good fits to the distributions of episode lengths in the whole range of observed data. These models, however, do not allow for seasonality which is an inherent property of the data.

Finally, we compare the probabilities p_n^{ch} of a change of the weather type from day n to $n+1$ which are given by the expression:

$$p_n^{\mathrm{ch}} = 1 - p_n(1, 1) - p_n(2, 2) - p_n(3, 3). \tag{7.3}$$

In Fig. 8, the probabilities of change resulting from the 2nd-order Markov chain model with monthly averages obtained from the data. The 1st-order model gives very similar results. The agreement between the models and the data is good. The deviations between data and models can be accounted for statistical effects resulting from the finite size of the data set. According to both data and models, the weather is most stable in summer. It is interesting that the model predicts the highest probability of change in April with its proverbial weather. The direct analysis of the data does not resolve this maximum.

7.3. *Hit rates and skill scores*

We have exploited the observed weather types of the three years from 1995 till 1997, which were not used for the estimation of the model parameters, in order to evaluate the predictive power of the present model. For this purpose we determined the probabilities $P(i, n+m | j, n; k, n-1) = \mathrm{prob}(Y_{n+m} = c_i | Y_n = c_j,\ Y_{n-1} = c_k)$ of the three possible types at the day $n+m$ conditioned on the types c_j and c_k at m and $m+1$ days earlier, respectively, where $m = 1, \ldots, 20$. For the 2nd-order Markov chain model these probabilities can be expressed by the elementary conditional probabilities $p_n(i | j, k)$ as follows:

$$P(i, n+1 | j, n; k, n-1) = p_n(i | j, k),$$
$$P(i, n+2 | j, n; k, n-1) = \sum_{i_1} p_{n+1}(i | i_1, j) p_n(i_1 | j, k),$$
$$P(i, n+m | j, n; k, n-1)$$
$$= \sum_{i_1, \ldots, i_{m-1}} p_{n+m-1}(i | i_1, i_2) p_{n+m-2}(i_1 | i_2, i_3) \times \ldots$$
$$\times p_{n+1}(i_{m-2} | i_{m-1}, j) p_n(i_{m-1} | j, k)$$
$$\text{for } m > 2. \tag{7.4}$$

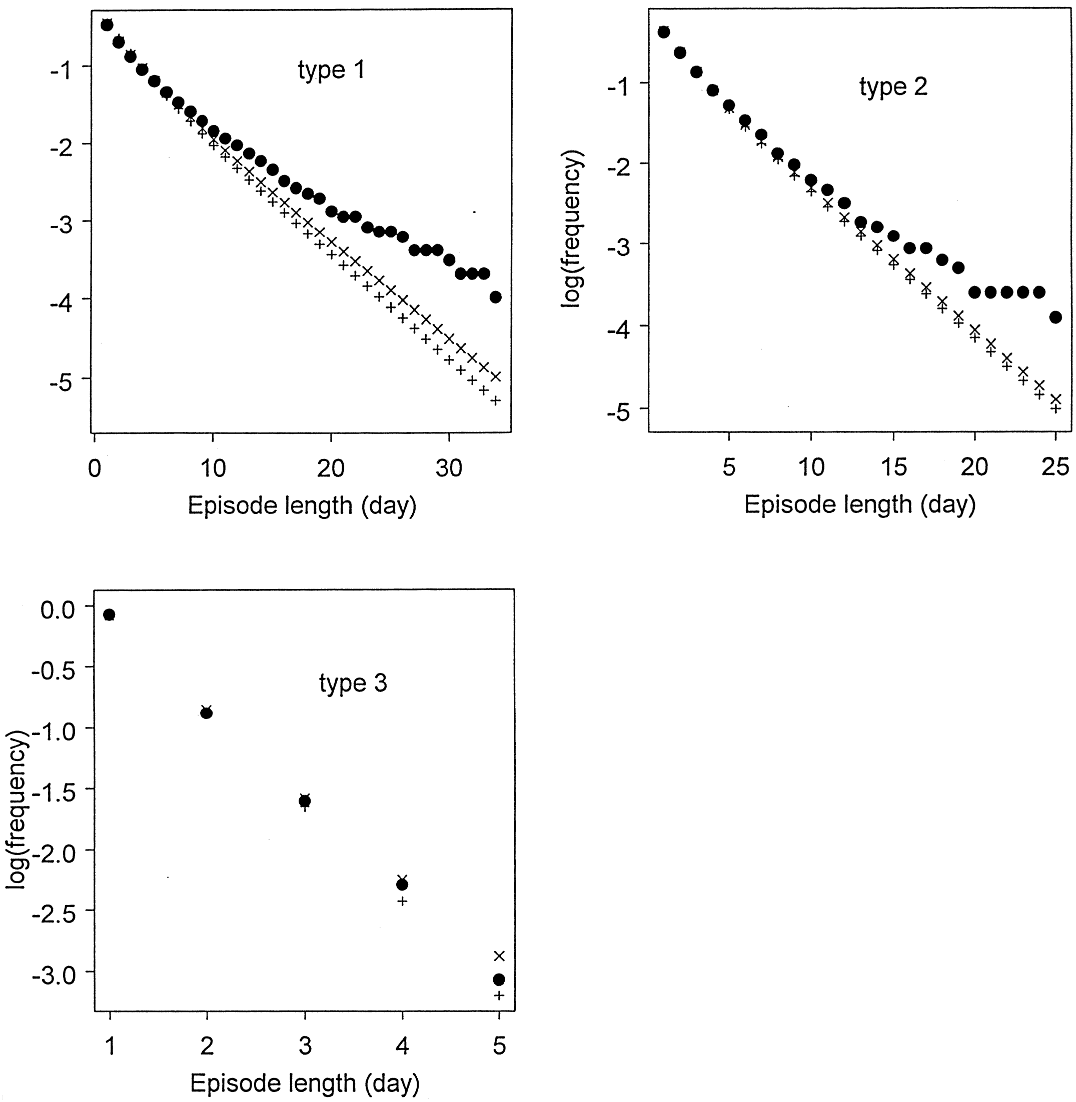


Fig. 7. The logarithm to base ten of the frequency of episodes of lengths l (full circles) for the three weather types are compared with the probabilities of the respective episodes resulting from the 1st-order (+) and the 2nd-order (×) Markov chain model.

For the 1st-order model, the same expressions hold. They, however, are simpler by the fact that then the transition probabilities $p_n(i|j,k)$ are independent of the state k at the earlier time, see Section 3.

Based on the observed weather types at the days n and $n-1$, a prediction for the weather type m days later is obtained by taking the weather type c_i with maximal probability $P(i, n+m|j, n; k, l)$. These predictions have been compared with the observed data by means of a contingency table, i.e., a 3 × 3 matrix. The (i, j)-element of this matrix equals the number of days for which type c_i is predicted and type c_j is observed (Wilks, 1995). Consequently, the diagonal elements indicate the numbers of correctly predicted days for each type. Their sum divided by the total number of observed data yields the hit rate for a prediction m days ahead. Fig. 9

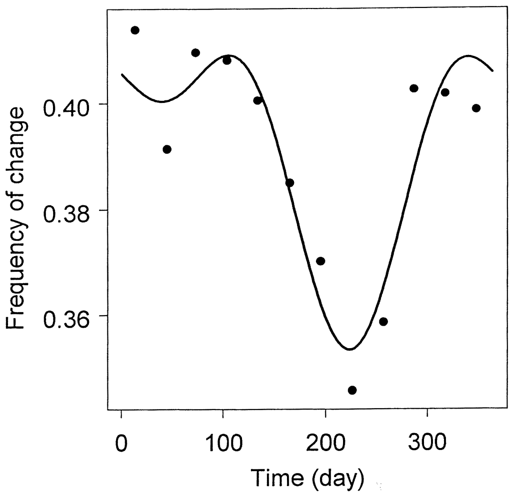


Fig. 8. The probability of a change of the weather type from day n to $n+1$ as a function of the day n is compared with the monthly averaged frequency of such a change.

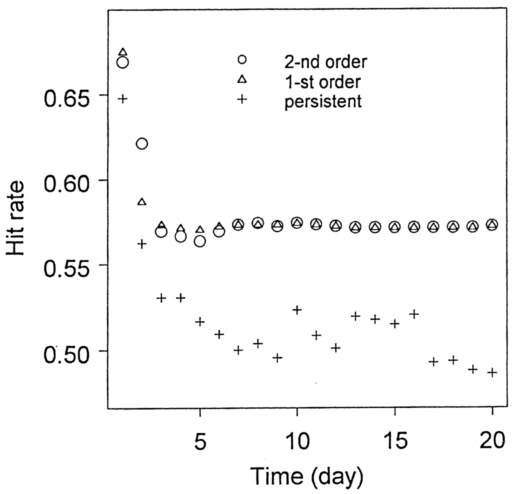


Fig. 9. Hit rates of predictions up to 20 days ahead, based on 1st- and 2nd-order Markov chain models and on a persistence assumption.

shows the hit rates as functions of the prediction time for the 1st- and 2nd-order model and a prediction based on persistence. There are only minor differences between the hit rates of the 1st- and 2nd-order model while the persistence assumption yields always less good results in particular for larger prediction times. After a few days the hit rate based on persistence approaches a value of roughly 0.5 which is close to the value

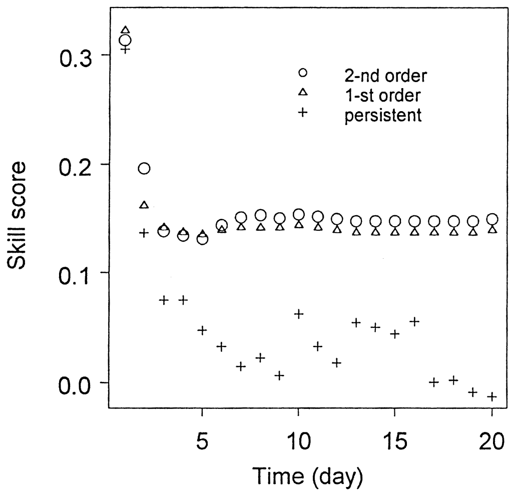


Fig. 10. Heidke skill scores of predictions of the same models as in Fig. 9.

of 0.45 that one would expect for three independent events occurring with the probabilities of the weather types given in eq. (2.1).

Other measures of the predictive power of a model are skill scores which also can be determined from the above described contingency tables (Wilks, 1995). Fig. 10 shows the Heidke skill score (Wilks, 1995) which is given by the difference of the actual hit rate and the hit rate one would observe if observations and predictions were independent of each other. This difference is normalized by the analogously defined difference for an ideal prediction with unit hit rate. The skill scores are at most unity and are zero for an independent prediction. Negative values may result from a systematic anticorrelation between predictions and observations. The skill scores are shown for predictions up to 20 days. Again there are only minor differences between the 1st- and 2nd-order Markov chain models which perform always better than predictions based on the persistence assumption. We have also calculated the Kuipers skill score which differs by its normalization but yields almost identical results as the Heidke skill score for the present data. We finally determined the hit rates and skill scores from data that were simulated by means of the 2nd-order Markov model. The resulting hit rates and skill scores are very similar to those based on the observed data. This further supports the validity of our model.

8. Summary

We have proposed seasonal Markov chains to describe the dynamics of Schüepp's weather types. The transition probabilities were determined by means of linear logit models and the parameters followed from maximizing the likelihood of this model. In this way it is possible to use all data for the estimate of the model paramaters and still to resolve the dynamics on a daily basis. The resulting 1st- and 2nd-order Markov chains lead to very similar results, e.g., for joint and conditional probabilities $p_n(i, j)$ and $p_n(i|j)$, respectively, for the statistics of episode lengths and also for prediction performance measures as hit rates and skill scores. From this point of view the 1st-order Markov chain is the model of choice; it is much simpler than the 2nd-order model and still yields almost the same results. As can be seen from Table 1 the 1st-order model has 9 parameters which are significantly different from zero. For the determination of these parameters more than 18,000 data have been used. Once this class of models is fixed, one can estimate the first few Fourier coefficients of the transition probabilities directly, e.g., by means of a least square fit. In this way one obtains almost the same transition probabilities as with the logit model in a much simpler way.

The predictive power of the model is not very high, but this is caused by the high randomness of the data and cannot be seen as a weekness of the model. The very stochastic nature of the weather type dynamics also is evident from the large entropy values of short weather type sequencies which have been estimated by Nicolis et al. (1997) in a nonparametric way.

Concerning the long-time dynamics both Markov models fail as they predict a significantly smaller frequency of long episodes of the types c_1 and c_2. In view of the present results and those of Nicolis et al. (1997), it would be interesting to combine both aspects and to investigate a seasonal semi-Markov process with stretched exponential waiting times.

Another promising approach to better understand the long-time dynamics can be based on the present seasonal Markov models. As a periodic matrix the transition probability possesses a Floquet representation which is analogous to the spectral representation of the transition probability of a homogeneous process. Since the period is long compared to the typical relaxation time, the semiadiabatic method of Talkner (1999) can be used to determine the Floquet states and multipliers. The components of the Floquet states can be used to numerically decorate the states. The resulting numerical time series can then be further analysed.

It would also be very interesting to see whether climatological models reproduce correctly the short and long-time behavior of weather types as they have been observed. This would require to determine daily weather types from model data by means of the same prescription used for real weather data, and, in a second step to perform the corresponding analysis of the obtained weather types along the lines explained in the present paper.

9. Acknowledgements

We would like to thank G. Stefanicki and R. Weber for many stimulating discussions. The Schweizerische Meteorologische Anstalt in Zürich has made us available the weather type data.

10. Appendix

Distribution of episode lengths

The average distribution of episode length $p^{\mathrm{epi}}(i, l)$ follows upon the time average over the starting day n from the probability $p_n^{\mathrm{epi}}(i, l) = \mathrm{prob}(Y_{n+1} = \ldots Y_{n+l} = c_i,\ Y_{n=l+1} \neq c_i | Y_n = c_i)$ to stay for l days in class c_i without interruption:

$$p^{\mathrm{epi}}(i, l) = \frac{1}{365.25} \sum_{n=1}^{365.25} p_n^{\mathrm{epi}}(i, l). \tag{A1}$$

From the independence of the conditional probabilities on the weather types beyond the previous or beyond the penultimate day for the 1st- and 2nd-order Markov chain models, respectively, one finds

$$p_n^{\mathrm{epi},1}(i, 1) = 1 - p_n(1|1),$$
$$p_n^{\mathrm{epi},1}(i, l) = (1 - p_{n+l}(i|i)) \prod_{k=1}^{l-1} p_{n+k}(i|i), \quad \text{for } l > 1, \tag{A2}$$

for the 1st-order model, where $p_n(i|j)$ is the transition probability of the 1st-order model, and

$$p_n^{\text{epi},2}(i,1) = 1 - \frac{p_n^{\text{as}}(i,i)}{p_n^{\text{as}}(i)},$$

$$p_n^{\text{epi},2}(i,2) = (1 - p_{n+1}(i|i,i))\,\frac{p_{n+1}^{\text{as}}(i,i)}{p_n^{\text{as}}(i)},$$

$$p_n^{\text{epi},2}(i,l) = (1 - p_{n+l-1}(i|i,i)) \times \left(\prod_{k=1}^{l-2} p_{n+k}(i|i,i)\right)\frac{p_{n+1}^{\text{as}}(i,i)}{p_{\mathbf{as}}(i)},$$

$$\text{for } l > 2, \qquad \text{(A3)}$$

for the 2nd-order model.

REFERENCES

Agresti, A. 1990. *Categorical data analysis.* John Wiley, New York.

Baraldi, C. 1994. *Modellierung von Zeitreihen durch stochastische Prozesse.* Dissertation, University of Basel, Switzerland.

Casella, G. and Berger, R. L. 1990. *Statistical inference.* Duxbury Press, Belmont.

Cox, D. R. and Miller, H. D. 1965. *The theory of stochastic processes.* Chapman and Hall, London.

Dixon, W. J. (ed.). 1992. *BMPD statistical software manual.* University Press of California, Princeton.

Hess, P. and Brezowski. H. 1977. *Katalog der Grosswetterlagen Europas.* Offenbach am Main, Selbstverlag des Deutschen Wetterdienstes.

Klaus, D. 1978. Perioden und Hauptkomponentenanalyse der Grosswetterlagen-Häufigkeiten Europas und der Ozeantemperaturfluktuationen des Nordatlantiks. *Meteorol. Rdsch.* **31**, 47–55.

Klaus, D. and Stein, G. 1978. Temporal variations of the European Grosswetterlagen and possible causes. *Geophys. Astrophys. Fluid Dynamics* **11**, 89–100.

Nicolis, C., Ebeling, W. and Baraldi, C. 1997. Markov processes, dynamic entropies and the statistical prediction of mesoscale weather regimes. *Tellus* **49A**, 108–118.

Salvisberg, E. 1996. Wetterlagenklimatologie-Möglichkeiten und Grenzen ihres Beitrages zur Klimawirkungsforschung im Alpenraum, Bern. *Geogr. Bernensia* **G51**.

Schüepp, M. 1979. *Klimatologie der Schweiz*, Band III. *Witterungsklimatologie.* Zürich: Schweizerische Meteorologische Anstalt, 93 pp.

Spekat, A., Heller-Schulze, B. and Lutz, M. 1983. Über Grosswetter und Markov–Ketten. *Meteorol. Rdsch.* **36**, 243–248.

Stefanicki, G., Talkner, P. and Weber, R. 1998. Frequency changes of weather types in the alpine region since 1945. *Theor. Appl. Climatol.* **60**, 47–61.

Talkner, P. 1999. Stochastic resonance in the semiadiabatic limit. *New Journal of Physics* **1**, 4.1–4.25 (http://www.njp.org).

Wanner, H., Rickli, R., Salvisberg, E., Schmutz, C. and Schüepp, M. 1997. Global climate change and variability and its influence on Alpine climate — concepts and observations. *Theor. Appl. Climatol.* **58**, 221–243.

Wanner, H., Rickli, R., Salvisberg, E. and Schüepp, M. 1998. 50 years of alpine weather statistics (AWS). *Meteor. Zeitschrift, N.F.* **7**, 99–111.

Wilks, D. S. 1995. *Statistical methods in atmospheric sciences.* Academic Press, New York.