

# Nonlinear autoregressive time series with multivariate Gaussian mixtures as marginal distributions

C.A. Glasbey

Biomathematics and Statistics Scotland

JCMB, King's Buildings, Edinburgh, EH9 3JZ, Scotland

July 27, 2000

## Abstract

A new form of nonlinear autoregressive time series is proposed to model solar radiation data, by specifying joint marginal distributions at low lags to be multivariate Gaussian mixtures. The model is also a type of multiprocess dynamic linear model, but with the advantage that the likelihood has closed form.

**Key words:** Causal graph, Latent variable, Markov process, Multiprocess dynamic linear model, Solar radiation

## 1 Introduction

Knowledge of the statistical characteristics of time series of solar radiation has many uses, one of which is in the design and evaluation of solar energy systems. Models are needed both to summarise data and to simulate large sets of stochastically-similar data for input for system models. Fig 1 shows 30-second averages of radiation, measured for 8 hours per day for one month, in Edinburgh, obtained as part of a EU-funded project (Graham et al., 1996). In this plot, the 16 hours of missing data per day has been omitted from the time scale. Fig 1 also shows expected radiation values, which were derived as a multiple of the elevation of the sun at that location on the earth's surface at specified times (Palz and Greif, 1996). We found that by dividing each observed radiation value ( $R_t$ ) by its expected value ( $E_t$ ), we were able to remove daily and seasonal patterns in both the mean and variability of solar radiation. The resulting variable is a measure of clearness, as it takes a value greater than unity when the sky is clear and a value less than unity when the sky is cloudy. Its marginal distribution is strongly bimodal. Further, by taking a square-root transformation, so that

$$C_t = \sqrt{\frac{R_t}{E_t}},$$

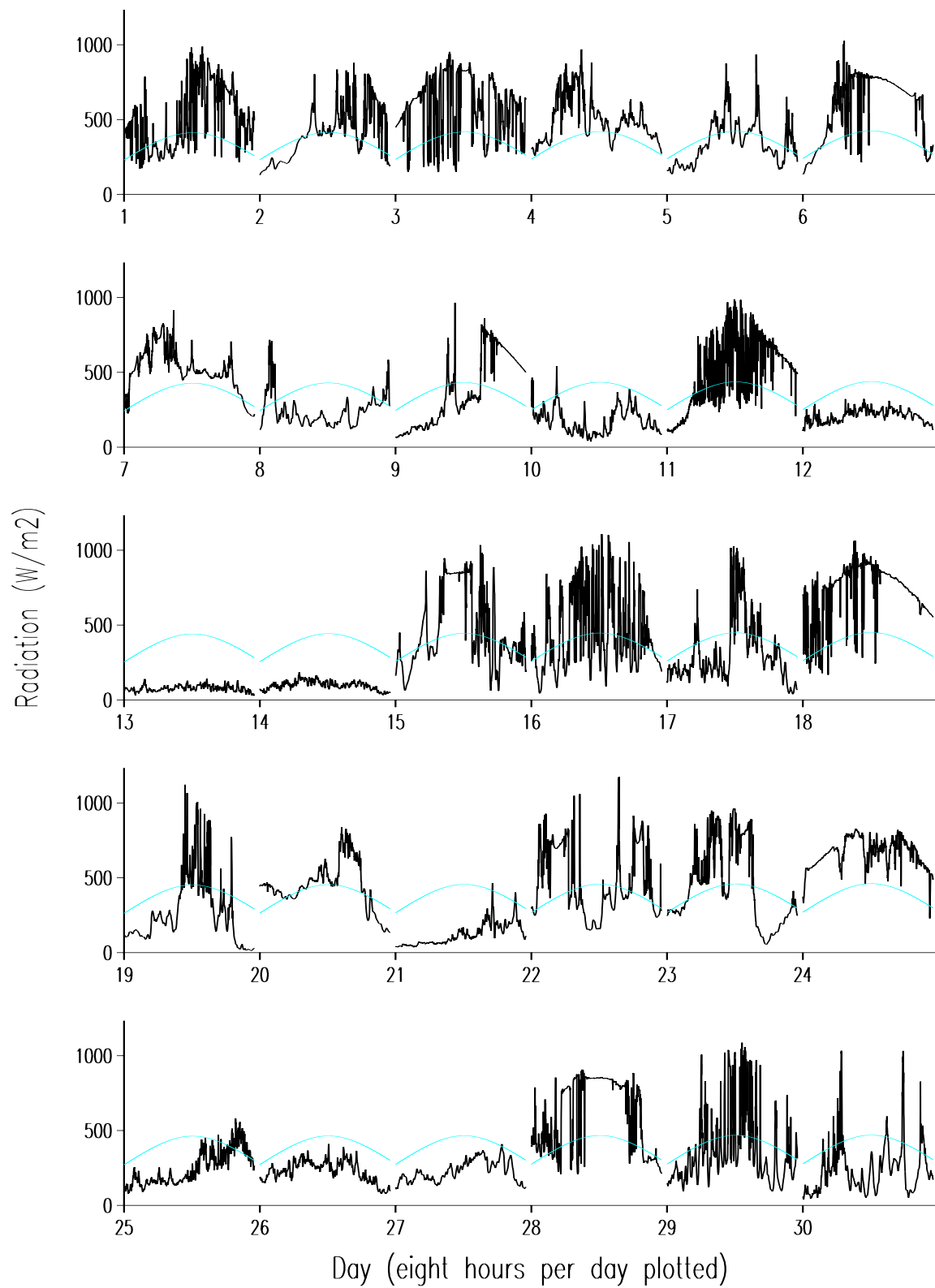


Figure 1: *30-second averages of solar radiation plotted against time, for May 1993: — observed, ··· expected.*

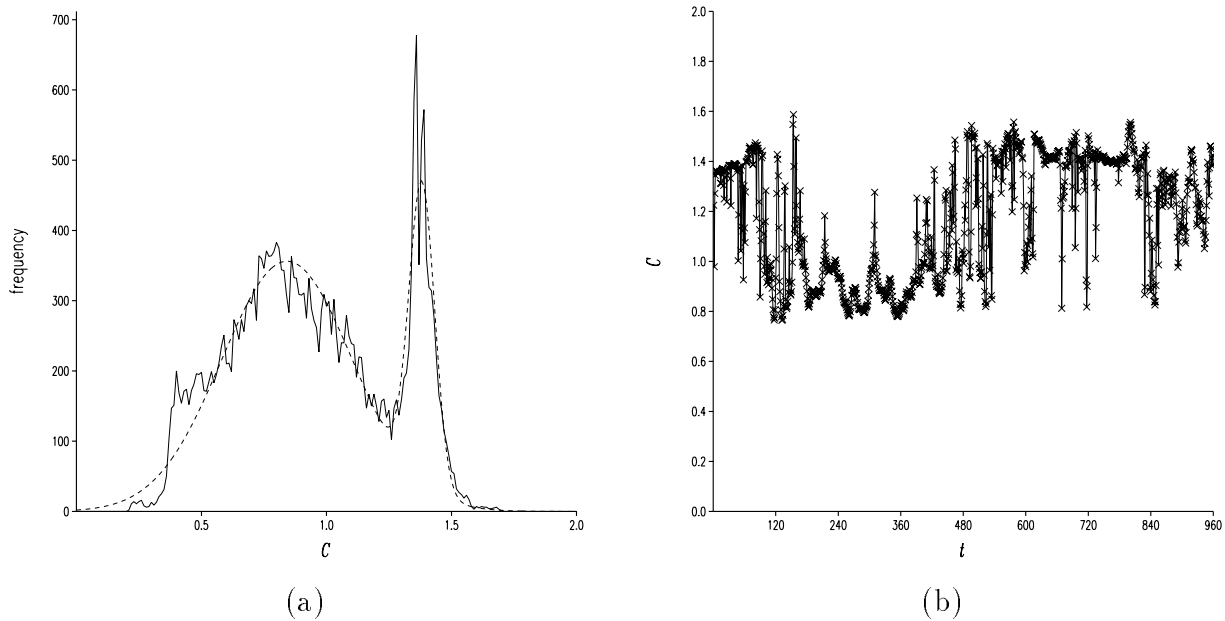


Figure 2: *Plots of  $C$ , the square-root of a clearness index: (a) distribution of  $C$  for data in Fig 1: — histogram, - - - maximum likelihood fit of mixture of two Gaussian distributions ( $\hat{\mu} = (0.84, 1.38)^T$ ,  $\hat{\sigma} = (0.261, 0.051)^T$ ,  $\hat{\pi} = (0.810, 0.190)^T$ ), (b) time series plot for May 1.*

the distribution was found to be well approximated by a mixture of two Gaussians, as can be seen in Fig 2(a). The two modes in the distribution of  $C$  are produced by cloudy times, when radiation is indirect, and cloud-free times, when radiation is direct (Owczarek, 1997). For illustration, Fig 2(b) shows a time series plot of  $C$  for a single day. We see that  $C$  can change substantially within a 30-second period. In applications involving photovoltaic cells, such changes would have an immediate effect of power output, and so it is important to model them. Therefore, we choose to model at this time scale rather than averaging the data, this being the shortest time scale at which we were able to record with our instrumentation.

Published analyses of solar radiation data have concentrated on first- and second-moment properties and on mechanistic models. Autoregressive models of solar radiation at a single site have been developed (Graham and Hollands, 1990; Aguiar and Collares Pereira, 1992; Beyer et al., 1995), models for mean instantaneous sky luminance distribution have been introduced and validated (Brunger and Hooper, 1993; Perez et al., 1993a; Ineichen et al., 1994), and the stochastic effects of cloud cover have been considered (Perez et al., 1993b; Long and Ackerman, 1995; Tovar et al., 1995; Morf, 1998). Here, we explore models for the data in Fig 2(b) and for 29 subsequent days. For simplicity, we ignore the mild dependence between days and assume that these are 30 independent realisations of a stationary time series. (The data are available by FTP at <ftp://ftp.bioss.sari.ac.uk/pub/chris/>). In §2 we review existing nonlinear time series models. Then, in §3 we propose a new model which is more appropriate for our data. Finally, in §4 we discuss the results.

## 2 Existing models

There are two broad categories of nonlinear time series models, multiprocess dynamic linear models and nonlinear autoregressive processes, both of which are capable of generating bimodal marginal distributions such as in Fig 2(a). Nonlinear models are capable of generating many other nonlinear features of time series, but we focus on multimodality, and consider each model category in turn, and its relevance to our data.

The first category is *multiprocess dynamic linear models* (MDLM) (West and Harrison, 1997, §12.3), alternatively termed *state-space models with switching*. If we initially consider the marginal distribution of  $C$ , then a standard way to represent a mixture model is to introduce an integer-valued indicator (also termed latent, state or exogenous) variable,  $I$ . Then, for a mixture of  $m$  normal distributions,

$$(C_t|I_t) \sim N(\mu_{I_t}, \sigma_{I_t}^2), \quad P(I_t = i) = \pi_i \quad i = 1, \dots, m,$$

where component  $i$  has mean  $\mu_i$ , variance  $\sigma_i^2$ , and occurs with probability  $\pi_i$ . To introduce temporal dependence, in a first-order MDLM it is assumed that the pair,  $(C, I)$ , is jointly Markovian, which we denote by

$$(C_t, I_t | \text{past}) \sim (C_t, I_t | C_{t-1}, I_{t-1})$$

(Billio and Monfort, 1998). This dependence structure can be represented by a directed causal graph, as shown in Fig 3(a).

Usually, a stronger assumption is made in MDLM, that  $I$  is unaffected by past values of  $C$ . Therefore

$$(I_t | \text{past}) \sim (I_t | I_{t-1}), \quad (C_t | I_t, \text{past}) \sim (C_t | C_{t-1}, I_{t-1}, I_t), \quad (1)$$

as represented in Fig 3(b). Therefore,  $I$  alone is a Markov chain, but  $C$  alone is not. Typically the conditional distribution of  $C$  is taken to be Gaussian, such as

$$(C'_t | I_t, \text{past}) \sim N(\phi_{I_{t-1}, I_t} C'_{t-1}, 1 - \phi_{I_{t-1}, I_t}^2), \quad \text{where} \quad C'_t = \frac{C_t - \mu_{I_t}}{\sigma_{I_t}}, \quad (2)$$

and  $\phi_{i,j}$  denotes the autocorrelation coefficient at lag one when the process is in states  $i$  and  $j$  at times  $t-1$  and  $t$ . The variance term,  $1 - \phi^2$ , ensures that  $C'$  has unit variance. It follows that the bivariate marginal distribution at lag one is a mixture of bivariate Gaussian distributions,

$$\begin{bmatrix} C_{t-1} \\ C_t \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_i \\ \mu_j \end{bmatrix}, \begin{bmatrix} \sigma_i^2 & \phi_{i,j} \sigma_i \sigma_j \\ \phi_{i,j} \sigma_i \sigma_j & \sigma_j^2 \end{bmatrix} \right) \quad \text{with probability } \pi_i P_{i,j}, \quad i, j = 1, \dots, m, \quad (3)$$

where  $P$  is the matrix of transition probabilities of  $I$ , and  $\pi$  is the stationary distribution among states. The joint likelihood for  $(C_1, C_2, \dots, C_T)^T$  does not have a simple form, and modified Kalman filters (West and Harrison, 1997), EM algorithms (Hamilton, 1990) and Markov chain Monte Carlo methods (Carter and Kohn, 1996) have been used to evaluate it. Note, however, that if a yet stronger assumption is made, that

$$(C_t | I_t, \text{past}) \sim (C_t | I_t),$$

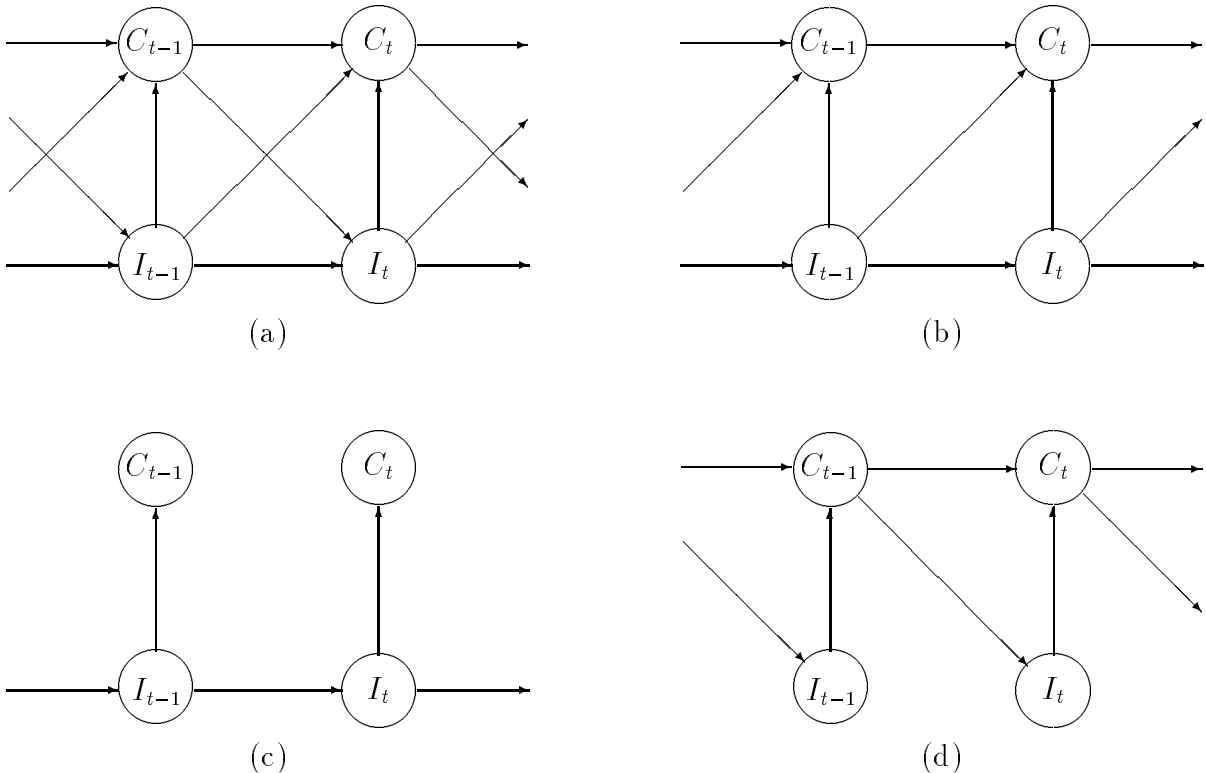


Figure 3: *Causal graph showing dependence structure in different time series models, with  $C_t$  denoting the observed variable at time  $t$  and  $I_t$  denoting an unobserved categorical indicator variable: (a) fully-connected multiprocess dynamic linear model, (b) commonly-used multiprocess dynamic linear model, (c) hidden Markov model, (d) first-order nonlinear autoregressive model.*

as shown in Fig 3(c), then we have what is termed a *hidden Markov model* (MacDonald and Zucchini, 1997), for which the exact likelihood can be evaluated (Leroux and Puterman, 1992).

For our data, Fig 4 shows bivariate histograms of  $C$  at lags 1 and 120, corresponding to time delays of thirty seconds and one hour. In this application, we find histograms such as these more informative than the plots of conditional means advocated by Tjøstheim (1994). Although the histograms look as though they may be approximated by mixtures of bivariate Gaussian distributions, components with means  $(\mu_i, \mu_j)^T$  for  $\mu_i \neq \mu_j$ , do not seem to appear at lag 1, as they would need to do if the MDLM were appropriate, in accord with (3). One quick way to fit the model specified by (1) and (2) is by maximising a quasi-likelihood of pairs of observations,

$$Q(C) = \prod_{t=2}^T P(C_{t-1}, C_t)$$

(Hjort and Omre, 1994), by fitting distribution (3) to the data in Fig 4(a). We reduced the number of parameters, by assuming that  $\phi_{i,j} = \sqrt{\phi_i \phi_j}$  and that the process is time reversible, so  $\pi P$  is symmetric. A numerical optimisation algorithm was run repeatedly from multiple, randomly-chosen starting values, and the best results were chosen. Table 1 shows the maximised values of  $Q$  for a range of values of  $m$ , from which 3 components appear to be sufficient. Table 2

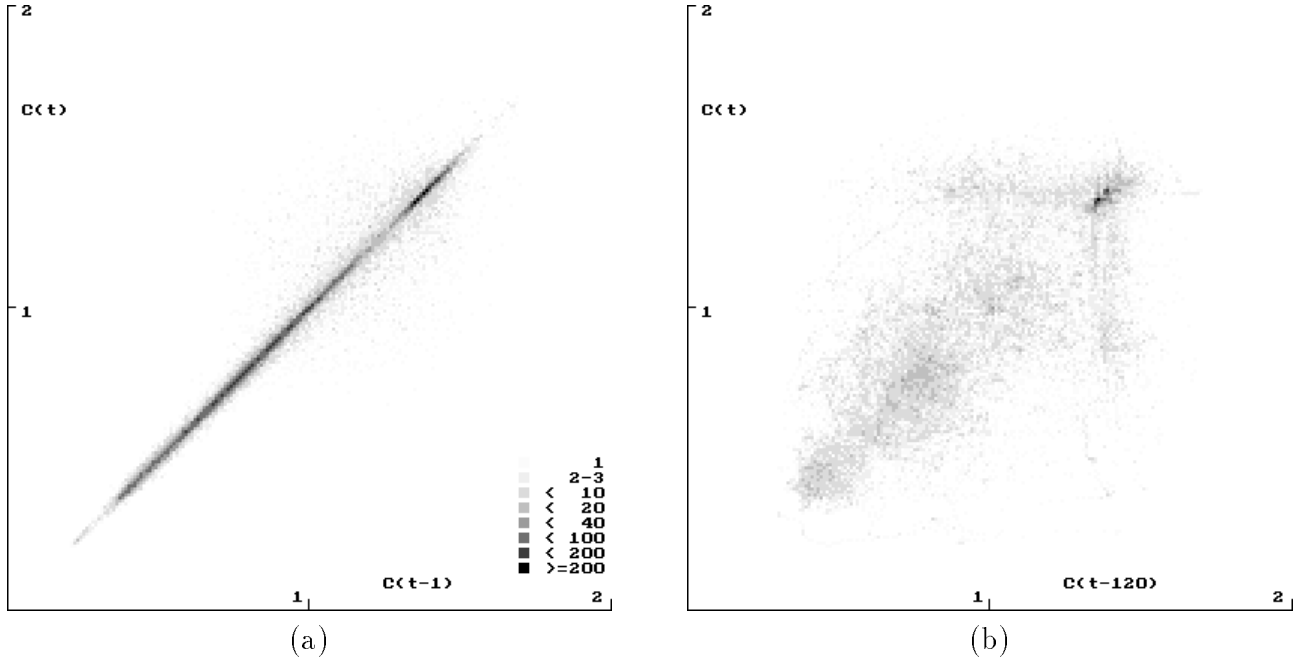


Figure 4: *Bivariate histograms of  $C$  at lags 1 and 120.*

$m$	2	3	4	5	6
$\frac{1}{T} \log Q$	3.84	4.04	4.05	4.07	4.07

Table 1: *Maximised values of the quasi-likelihood for a range of values of the number of mixture components in MDLM.*

shows the resulting parameter estimates when  $m = 3$  and Fig 5 shows 95% ellipses for each of the nine constituent distributions superimposed on the bivariate histogram at lag one. Two of the three pairs of off-diagonal elements in  $\pi P$  are estimated to be zero. These are shown as grey ellipses in Fig 5, and imply that in this model the chain,  $I$ , is reducible. It can also be seen that none of the off-diagonal distributions matches the data, so we conclude that this MDLM is not an appropriate model.

The second category of nonlinear time series model is *nonlinear autoregressive* (NLAR) processes (Tong, 1990, §3). First-order processes, denoted NLAR(1), are defined by

$$C_t = f(C_{t-1}, e_t), \tag{4}$$

for a specified function,  $f$ , where  $e_t$  is an independent error term. In particular, Jones (1978) showed that amplitude-dependent exponential autoregressive models (EXPAR) can produce bimodal marginal distributions. More recently, work on NLAR models capable of producing multimodal marginal distributions has concentrated on those that can be formulated using indicator variables, such as:

$$(C_t | \text{past}) \sim N\left(\mu_{I_t} + \phi_{I_t}(C_{t-1} - \mu_{I_t}), \sigma_{I_t}^2(1 - \phi_{I_t}^2)\right), \quad p(I_t = i) = g_i(C_{t-1}), \quad i = 1, \dots, m, \tag{5}$$

for some functions  $g_1, \dots, g_m$ . The causal graph for this model is shown in Fig 3(d). It can be

$i$	$\hat{\mu}_i$	$\hat{\sigma}_i$	$\hat{\phi}_i$	$\widehat{\pi_i P_{i,j}}$		
				$j = 1$	2	3
1	0.79	0.238	0.9994	0.673	0.005	0.0
2	1.19	0.209	0.8771	0.005	0.187	0.0
3	1.38	0.038	0.9956	0.0	0.0	0.129

Table 2: *Parameter estimates for three-component MDLM, obtained by maximising the quasi-likelihood.*

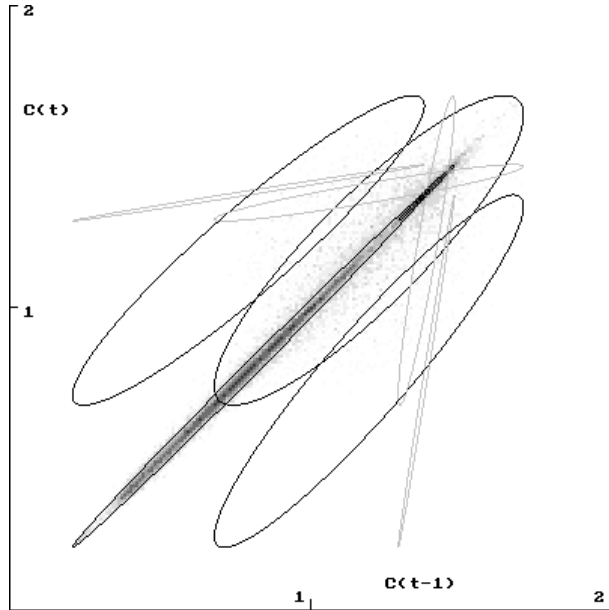


Figure 5: *Bivariate histogram of  $C$  at lag 1, with 95% ellipses for each of the nine constituent distributions in the MDLM model in table 2 superimposed. The two pairs of off-diagonal distributions for which the term in  $\widehat{\pi P}$  is zero are shown as grey ellipses (density scale is given in Fig 4(a)).*

seen to be another special case of the fully-connected MDLM in Fig 3(a), but quite different from Figs 3(b) and (c), and the  $C$ -process alone is Markovian, but  $I$  alone is not. This model has strengths and weaknesses when compared with MDLM. One of the strengths is that, as a result of the Markov property, the likelihood has closed form:

$$L(C) = P(C_1, C_2) \prod_{t=3}^T P(C_t | C_{t-1}). \quad (6)$$

We can assess the appropriateness of the Markovian assumption with our data, without making any further model assumptions. If we discretise  $C$  using the histogram bins, then we can simply use the histogram in Fig 4(a) as a transition matrix, after rescaling each column to unity. (Note, this is only possible because we have a large dataset and we are only considering a first-order process.) By raising this transition matrix to the power 120 and multiplying by the marginal

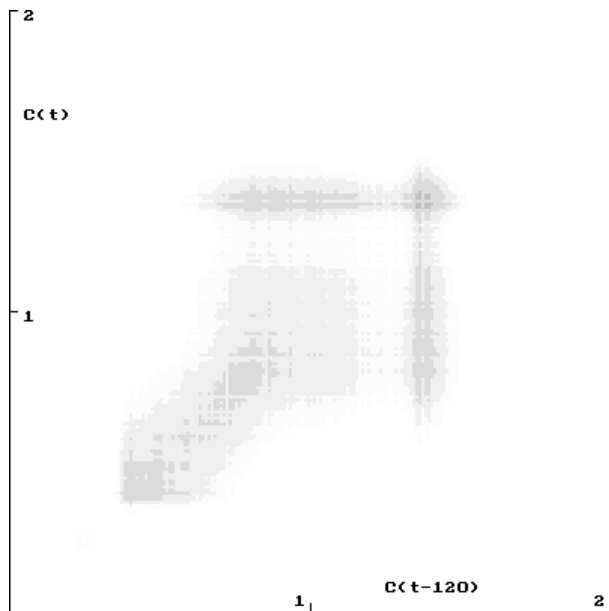


Figure 6: *Predicted bivariate histogram of  $C$  at lag 120, assuming the process is Markovian (density scale is given in Fig 4(a)).*

distribution, we obtain Fig 6, which shows the bivariate histogram at lag 120 that this model predicts. The agreement with Fig 4(b) can be seen to be good, reproducing all the main features except for the long-term persistence in values of  $C$  around 1.4. This lends support to the assumption that  $C$  is Markovian.

Several choices of  $g$  in (5) have been proposed. Self-exciting threshold autoregressive (SETAR) processes are defined by

$$g_i(C) = \begin{cases} 1 & \text{if } T_{i-1} \leq C < T_i \\ 0 & \text{otherwise} \end{cases} \quad i = 1, \dots, m,$$

for limits  $-\infty \equiv T_0 < T_1 < \dots < T_m \equiv \infty$  (Moeanaddin and Tong, 1990). Wong and Li (2000) proposed mixture autoregressive (MAR) models, which generalise the mixture transition distribution (MTD) models of Le et al. (1996), and for which probabilities are constant, given by

$$g_i(C) = \pi_i \quad i = 1, \dots, m.$$

Müller et al. (1997) proposed a form of smooth threshold autoregressive (STAR) process by specifying that

$$g_i(C) = \frac{e^{-(C-\mu'_i)^2/2\sigma^2}}{\sum_j e^{-(C-\mu'_j)^2/2\sigma^2}}, \quad (7)$$

which involves additional parameters  $\mu'_1, \dots, \mu'_m$  and  $\sigma^2$ , and a similar model was proposed in the neural network literature by Jordon and Jacobs (1994). However, unlike MDLM, in general these models do not have marginal distributions that are Gaussian mixtures, nor are their marginal distributions generally known. One exception is the MAR model, but this is too simple to account for the observed patterns in Fig 4(a).

### 3 New class of models

The lack of MDLM and NLAR models consistent with the bivariate histogram at lag one in Fig 4(a), leads us to propose a new class of NLAR model. We specify that  $C$  is a stationary Markov process, and that  $(C_{t-1}, C_t)^T$  has a symmetric bivariate marginal distribution. Joe (1997, pp. 244-246) briefly considers such models. We assume that the bivariate distribution is so chosen that the process is irreducible. It follows that, unlike most NLAR processes, the marginal distribution of  $C$  is known, it is specified by the marginal density of the bivariate distribution, and also the process is time reversible, which is consistent with the evidence in Fig 4. Further, we specify that

$$\begin{bmatrix} C_{t-1} \\ C_t \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_i \\ \mu_i \end{bmatrix}, \sigma_i^2 \begin{bmatrix} 1 & \phi_i \\ \phi_i & 1 \end{bmatrix} \right) \quad \text{with probability } \pi_i, \quad i = 1, \dots, m.$$

Note that, although we have not included any component distributions with unequal means,  $(\mu_i, \mu_j)^T$ , switching between components can still occur, because  $(C_{t+1}, C_t)^T$  is not constrained to be drawn from the same mixture component as  $(C_t, C_{t-1})^T$ . It follows that  $(C_t | \text{past})$  is given by (5), but now

$$g_i(C) = \frac{\frac{\pi_i}{\sigma_i} e^{-(C-\mu_i)^2/2\sigma_i^2}}{\sum_j \frac{\pi_j}{\sigma_j} e^{-(C-\mu_j)^2/2\sigma_j^2}} \quad i = 1, \dots, m. \quad (8)$$

Therefore, we have a STAR model similar to (7), but because we use a common set of parameters in (5) and (8), we obtain the simplifications already discussed. As an aside, we observe that our new model is also the Gibbs sampler of a bivariate symmetric distribution. Jennison (1993) considered such a process, involving Gaussian mixtures, in discussing convergence rates of Markov chain Monte Carlo methods.

The model specified by (5) and (8) was fit to our data by maximising the likelihood (6) for a range of values of  $m$ . Again, a numerical optimisation algorithm was run repeatedly from multiple, randomly-chosen starting values, and the best results were chosen. Table 3 shows the maximised likelihoods, from which we see that 4 components appear to be sufficient. Parameter estimates for the 4-component model are given in Table 4, together with standard errors obtained from a numerical approximation to the Hessian matrix. Fig 7(a) shows 95% ellipses for each of the four constituent distributions superimposed on the bivariate histogram at lag one. We can see that the first and second components model the indirect radiation mode in the distribution, the fourth component models the direct radiation mode, and the third component is needed to describe large fluctuations in clearness as clouds pass in front of the sun at times when cloud cover is partial. Fig 7(b) shows the predicted density at lag 120, which is very similar to Fig 6 and in reasonable agreement with the observed histogram in Fig 4(b). All the main aspects in the histogram are reproduced, except for the long-term persistence in values of  $C$  around 1.4, a feature that can also be discerned in Fig 1. However, it should be noted that the parameter estimates are somewhat different from those in Fig 2(a) and the fit to the univariate histogram is less good. We attribute this to the likelihood being dominated by conditional densities,  $P(C_t | C_{t-1})$ , and so maximum likelihood estimators are not robust against slight model misspecification.

$m$	2	3	4	5	6
$\frac{1}{T} \log L$	3.19	3.28	3.33	3.33	3.33

Table 3: *Maximised values of the likelihood for a range of values of the number of mixture components in the new NLAR model.*

$i$	$\hat{\mu}_i$	$\hat{\sigma}_i$	$\hat{\phi}_i$	$\hat{\pi}_i$
1	0.92 (0.028)	0.308 (0.0140)	0.9997 (0.00003)	0.468 (0.0141)
2	0.98 (0.026)	0.316 (0.0151)	0.9966 (0.00038)	0.294 (0.0093)
3	1.20 (0.005)	0.171 (0.0025)	0.5564 (0.01480)	0.167 (0.0094)
4	1.37 (0.001)	0.030 (0.0009)	0.9985 (0.00010)	0.072 (0.0060)

Table 4: *Maximum likelihood parameter estimates (standard errors in brackets) for new NLAR model with four components.*

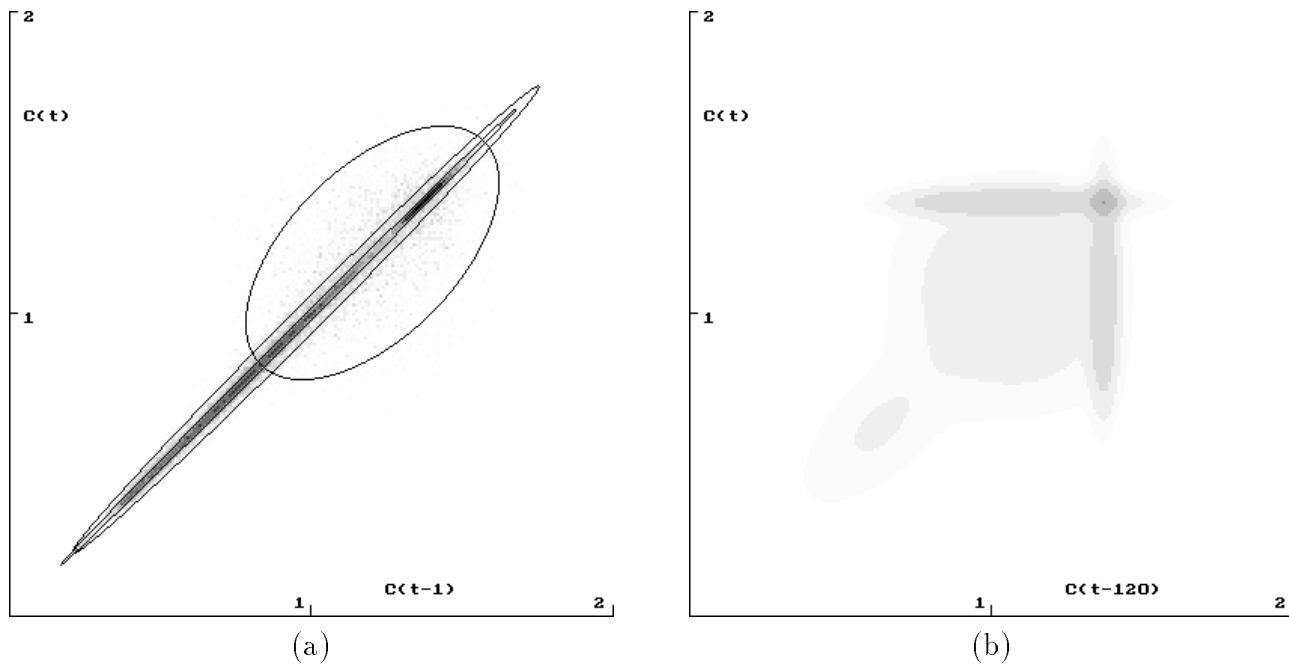


Figure 7: *Bivariate histograms of  $C$ : (a) at lag 1, with 95% ellipses for each of the four constituent distributions in the new NLAR model in table 4 superimposed, (b) at lag 120, as predicted by new NLAR model (density scale is given in Fig 4(a)).*

A new  $p$ th-order NLAR process can be defined in the same way, by specifying the conditional dependence,

$$(C_t \mid \text{past}) \sim (C_t \mid C_{t-1}, \dots, C_{t-p}),$$

and the  $(p + 1)$ -dimensional marginal distribution,

$$(C_t, C_{t-1}, \dots, C_{t-p})^T \sim N(\mu_i \mathbf{1}, V_i) \quad \text{with probability } \pi_i, \quad i = 1, \dots, m,$$

where  $\mathbf{1}$  is a unit vector. We restrict  $V_i$  to be a  $(p + 1) \times (p + 1)$  positive-definite, symmetric, Toeplitz matrix, to ensure that  $C$  is stationary. When  $m = 1$  the process is a  $p$ th-order autoregressive process, and for  $m > 1$  it is a particular form of a mixture of  $m$  autoregressive processes. In theory, it would also be possible to include component distributions with unequal means,  $(\mu_i, \mu_j)^T$ , but the drawbacks would be having to deal with a complicated set of constraints to ensure that  $C$  was stationary and a large increase in the number of parameters.

## 4 Discussion

We have identified a nonlinear time series model for 30-second averages of solar radiation recorded at a single site in Edinburgh. This provides a simple way to summarise such data and is easy to simulate from, as is needed for input to solar energy system models. The model is one of a new class of NLAR process, obtained by specifying symmetric marginal distributions at low lags. The resulting NLAR processes are time-reversible with known marginal distributions. Further, if the joint distributions are mixtures of Gaussians, the models are a subclass of MDLM. They generalise autoregressive processes and, unlike most MDLM, their likelihoods have closed form expressions. However, the new models also have some drawbacks. There is some inelegance in that, for  $m > 1$ , the models are non-nested as  $p$  increases. Also, joint and conditional distributions for these new models at other lags are not mixtures of Gaussians, nor do they have closed form expressions, as illustrated in the Appendix. As a related issue, it does not seem to be possible to embed the discrete time process in continuous time. Nor are we aware of any way to extend the model to spatial or spatio-temporal situations, as we would wish to do to simulate variations in solar radiation over a region.

## Acknowledgements

I am grateful to Alastair Hunter for introducing me to this application area and collecting the data under CEC Contract No. JOU2-CT92-0018, and to Carmen Fernandez for early discussion of the work. The work was supported by funds from the Scottish Executive Rural Affairs Department.

## References

- Aguiar, R. and Collares Pereira, M. (1992). TAG – a time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation data. *Solar Energy*, 49:167–174.
- Beyer, H. G., Bohlen, M., and Schumacher, J. (1995). GENG: A solar radiation data generator for the simulation system INSEL. In Freiesleben, W., Palz, W., Ossenbrink, H. A., and Helm, P., editors, *Proceedings of 13th European Photovoltaic Solar Energy Conference*, volume 1, pages 982–985, Bedford. H. S. Stephens & Associates.
- Billio, M. and Monfort, A. (1998). Switching state-space models – likelihood function, filtering and smoothing. *Journal of Statistical Planning and Inference*, 68:65–103.
- Brunger, A. P. and Hooper, F. C. (1993). Anisotropic sky radiance model based on narrow field of view measurements of short wave radiance. *Solar Energy*, 51:53–64.
- Carter, C. K. and Kohn, R. (1996). Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika*, 83:589–601.
- Graham, R., Glasbey, C. A., and Hunter, A. G. M. (1996). *Consequences of decentralised PV on local network management. Final report: variation of solar energy across a region: spatio-temporal models*. SAC Report, Bush Estate, Penicuik EH26 0PH, Scotland.
- Graham, V. A. and Hollands, K. G. T. (1990). A method to generate synthetic hourly solar radiation globally. *Solar Energy*, 44:333–341.
- Hamilton, J. D. (1990). Analysis of times series subject to changes in regime. *Journal of Econometrics*, 45:39–70.
- Hjort, N. L. and Omre, H. (1994). Topics in spatial statistics (with discussion). *Scandinavian Journal of Statistics*, 21:289–357.
- Ineichen, P., Molineaux, B., and Perez, R. (1994). Sky luminance data validation – comparison of 7 models with 4 data-banks. *Solar Energy*, 52:337–346.
- Jennison, C. (1993). Discussion on the meeting on the Gibbs sampler and other Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, 55:54–56.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.
- Jones, D. A. (1978). Nonlinear autoregressive processes. *Proceedings of the Royal Society, London, Series A*, 360:71–95.
- Jordan, M. I. and Jacobs, R. A. (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, 8:181–214.
- Le, N. D., Martin, R. E., and Raftery, A. E. (1996). Modeling flat stretches, bursts, and outliers in time series using Mixture Transition Distribution Models. *Journal of the American Statistical Association*, 91:1504–1515.

- Leroux, B. G. and Puterman, M. L. (1992). Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics*, 48:545–558.
- Long, C. N. and Ackerman, T. P. (1995). Surface measurements of solar irradiance: a study of the spatial correlation between simultaneous measurements at separated sites. *Journal of Applied Meteorology*, 34:1039–1046.
- MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series*. Chapman and Hall, London.
- Moeanaddin, R. and Tong, H. (1990). Numerical evaluation of distributions in non-linear autoregression. *Journal of Time Series Analysis*, 11:33–48.
- Morf, H. (1998). The stochastic two-state solar irradiance model (STSIM). *Solar Energy*, 62:101–112.
- Müller, P., West, M., and MacEachern, S. (1997). Bayesian models for non-linear autoregressions. *Journal of Time Series Analysis*, 18:593–614.
- Owczarek, S. (1997). Vector model for calculation of solar radiation intensity and sums incident on tilted surfaces. Identification for the three sky condition in Warsaw. *Renewable Energy*, 11:77–96.
- Palz, W. and Greif, J., editors (1996). *European Solar Radiation Atlas: Solar Radiation on Horizontal and Inclined Surfaces*. Springer-Verlag, Berlin, 3rd edition.
- Perez, R., Seals, R., and Michalsky, J. (1993a). All-weather model for sky luminance distribution – preliminary configuration and validation. *Solar Energy*, 50:235–245.
- Perez, R., Seals, R., Michalsky, J., and Ineichen, P. (1993b). Geostatistical properties and modelling of random cloud patterns for real skies. *Solar Energy*, 51:7–18.
- Tjøstheim, D. (1994). Non-linear time series: a selective review. *Scandinavian Journal of Statistics*, 21:97–130.
- Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Clarendon Press, Oxford.
- Tovar, J., Olmo, F. J., and Alados Arboledas, L. (1995). Local-scale variability of solar radiation in a mountainous region. *Journal of Applied Meteorology*, 34:2316–2322.
- West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer, New York, 2nd edition.
- Wong, C. S. and Li, W. K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society, Series B*, 62:95–116.

## Appendix: dependence at lag two for first-order process

Consider

$$P(C_{t+1} | C_{t-1}) = \int P(C_{t+1} | C_t = x) P(C_t = x | C_{t-1}) dx.$$

From (5),

$$P(C_t | C_{t-1}) = \sum_{i=1}^m \frac{g_i(C_{t-1})}{\sqrt{2\pi\sigma_i^2(1-\phi_i^2)}} \exp\left[-\frac{\{C_t - \mu_i - \phi_i(C_{t-1} - \mu_i)\}^2}{2\sigma_i^2(1-\phi_i^2)}\right],$$

therefore

$$\begin{aligned} P(C_{t+1} | C_{t-1}) &= \sum_j \sum_i \int \frac{g_j(x)}{\sqrt{2\pi\sigma_j^2(1-\phi_j^2)}} \frac{g_i(C_{t-1})}{\sqrt{2\pi\sigma_i^2(1-\phi_i^2)}} \\ &\times \exp\left[-\frac{\{C_{t+1} - \mu_j - \phi_j(x - \mu_j)\}^2}{2\sigma_j^2(1-\phi_j^2)} - \frac{\{x - \mu_i - \phi_i(C_{t-1} - \mu_i)\}^2}{2\sigma_i^2(1-\phi_i^2)}\right] dx. \end{aligned}$$

Substituting for  $g_j(x)$  from (8), and rearranging terms, produces

$$\begin{aligned} P(C_{t+1} | C_{t-1}) &= \sum_j \sum_i \frac{1}{\sqrt{2\pi\sigma_j^2(1-\phi_j^2)}} \frac{g_i(C_{t-1})}{\sqrt{2\pi\sigma_i^2(1-\phi_i^2)}} \\ &\times \int \frac{\frac{\pi_j}{\sigma_j} e^{-(x-\mu_j)^2/2\sigma_j^2}}{\sum_k \frac{\pi_k}{\sigma_k} e^{-(x-\mu_k)^2/2\sigma_k^2}} \exp\left[-\frac{\{C_{t+1} - \mu_j - \phi_j(x - \mu_j)\}^2}{2\sigma_j^2(1-\phi_j^2)} - \frac{\{x - \mu_i - \phi_i(C_{t-1} - \mu_i)\}^2}{2\sigma_i^2(1-\phi_i^2)}\right] dx. \end{aligned}$$

We know of no analytic solution to this integral, of an exponentiated quadratic form in  $x$  divided by a sum of such exponentiated quadratic forms. It is evident that the solution will not be a mixture of Gaussian densities.