Highest-density forecast regions for non-linear and non-normal time series models

Rob J. Hyndman

Monash University, Australia

Journal of Forecasting, **14**, 431–441 (1995)

Abstract

Forecast regions are a common way to summarize forecast accuracy. They usually consist of an interval symmetric about the forecast mean. However, symmetric intervals may not be appropriate forecast regions when the forecast density is not symmetric and unimodal. With many modern time series models, such as those which are non-linear or have non-normal errors, the forecast densities are often asymmetric or multimodal. The problem of obtaining forecast regions in such cases is considered and it is proposed that highest-density forecast regions be used. A graphical method for presenting the results is discussed.

Keywords:

non-linear time series, non-normal time series, highest density regions, forecast intervals, threshold models.

1 Introduction

Forecast regions are crucial to practical forecasting—they convey in a simple way the accuracy of a forecast. When the forecast densities are normal, as with ARIMA models (see, for example, Brockwell and Davis, 1991, p.175), they can easily be summarized by a forecast interval symmetric about the mean. For example, an approximate 95% forecast interval is the mean plus and minus two standard deviations. However, when a model is non-linear or non-normal, the forecast densities are usually no longer normal and are frequently asymmetric or multimodal (Al-Qassam and Lane, 1989; Moeanaddan and Tong, 1990; Tong, 1990). In this case, forecast regions symmetric around the mean can be quite misleading. Instead, forecast regions need to be constructed so as to convey the shape of the forecast density.

While there has been much activity in studying model properties, methods of estimation, diagnostics and point forecasts for many of the models that have been suggested (see, for example, De Gooijer and Kumar, 1992), forecast distributions and corresponding forecast regions tend to have been overlooked. In particular, the relative merits of different forecast regions do not seem to have been discussed previously in the forecasting literature. (It is not an issue for linear normal models because the forecast densities are normal and so all the obvious methods of constructing forecast regions yield identical results.) Often the methods that are appropriate for linear normal models are carried across to the non-linear and non-normal context. For example, Tong and Moeanaddin (1988) consider forecasts from a threshold autoregressive model but give forecast intervals based on a number of standard deviations about the mean. More frequently, the variance or some other measure of accuracy is tabulated but no forecast intervals are given (e.g., Ray, 1988).

In this article, the use of highest-density regions (HDRs) in forecasting is explored and compared with other forecast regions. Broemeling and Cook (1992) use highest density forecast regions but only for one-step ahead forecast regions and in a special case where the forecast density is a t-distribution. In this case, the HDR is equivalent to the usual interval around the mean.

We argue that HDRs are a more effective summary of the forecast distribution than other common forecast regions because they portray any asymmetry and multimodality that is in the forecast density. A simple Monte Carlo method for estimating HDRs and other forecast regions for any time series model is outlined. The graphical representation of the regions is also discussed.

Two examples are given which demonstrate the problems that can occur with the standard forecast

regions and the additional information which can be obtained from HDRs. The first example uses simulated data from an EXPAR model; the second example uses a SETAR model with real data.

2 Selecting a forecast region

Let $\{X(t_i)\}$ denote the time series observed at times $t_1 < t_2 < \cdots < t_n$. We are interested in the random variable, $X(t_m)$, given all observations up to and including time t_n, $n < m$. Denote this by $X_{m|n} = [X(t_m)|X(t_1), \ldots, X(t_n)]$ and its density by $p_{m|n}$. Let $\mu_{m|n}$ denote the mean of $p_{m|n}$ and $Q_{m|n}(p)$ denote its pth quantile. It is assumed that $p_{m|n}(x)$ is a continuous function of x.

A forecast region is a region of the sample space of $X_{m|n}$ which summarises, in some sense, the density $p_{m|n}$. Such a region could be constructed in an infinite number of ways. The three methods considered in the introductory example are, perhaps, the most obvious ways to construct a $100(1-\alpha)\%$ forecast region. That is,

1. The interval symmetric about the mean. That is,

$$R_\alpha = \mu_{m|n} \pm w.$$

2. the interval defined between the $\alpha/2$ and $(1-\alpha/2)$ quantiles. That is,

$$R_\alpha = [Q_{m|n}(\alpha/2), Q_{m|n}(1-\alpha/2)].$$

3. The highest-density region. That is

$$R_\alpha = \{x : p_{m|n}(x) \geq f_\alpha\}$$

Here w and f_α are chosen such that $\Pr(X_{m|n} \in R_\alpha) = 1-\alpha$.

For symmetric and unimodal distributions such as the normal distribution, these regions are identical. But for non-normal forecast distributions, the regions can all be different. Of the three choices above, only the third choice is flexible enough to convey both multimodality and asymmetry in the forecast density. Any region restricted to a single interval will mask multimodality in the forecast density; any interval symmetric about the mean (or median or mode) will will not convey any skewness or other asymmetry in the forecast density.

Furthermore, intervals symmetric about the mean are usually constructed using the mean plus or minus a multiple of the standard deviation. In this case, the probability coverage of the intervals may vary. Of course, it is possible to construct intervals which are symmetric about the mean and which have constant probability coverage, but this is rarely done and the resulting intervals will still mask asymmetry and multimodality.

The usual purpose of forecast regions also suggests that the third choice above is the most natural criterion. A forecast region is intended to summarize the forecast density by delineating a comparatively small region of the sample space with a given probability. It can be thought of as a summary of the region in which the process will most likely fall. Therefore it is desirable that the forecast region consist of points where the density is relatively high. Highest density regions are defined as the $100(1-\alpha)\%$ region such that every point inside the region has probability density at least as large as every point outside the region. This is equivalent to the region occupying the smallest possible volume in the sample space (Box and Tiao, 1973). In this sense, highest density regions represent the canonical choice amongst all possible regions.

Highest density regions are common in Bayesian analysis where they are applied to the posterior density. In that context they are known under various other names including "credible sets" (Edwards et al., 1963), "highest posterior density regions" (Box and Tiao, 1973) and "Bayesian confidence sets" (Lindley, 1965).

Example 1: EXPAR model

As a motivating (but artifical) example, consider the simple first-order EXPAR model

$$X(t) = -0.3X(t-1) - 0.8X(t-1)\exp(-X^2(t-1)) + e(t) \tag{1}$$

where $\{e(t)\}$ is a normal white noise sequence with variance 0.0064. This model is similar to one considered by Al-Qassam and Lane (1989) and is used here because it demonstrates rather unusual behaviour which exaggerates the problem of choosing an appropriate forecast density. However, the same difficulties that arise with this model occur with many non-linear and non-normal models, although usually to a lesser extent.

A realization of this process is plotted in the left half of the plots in Figure 2. The time series tends to oscillate between positive and negative values. Occasionally, there are two consecutive positive

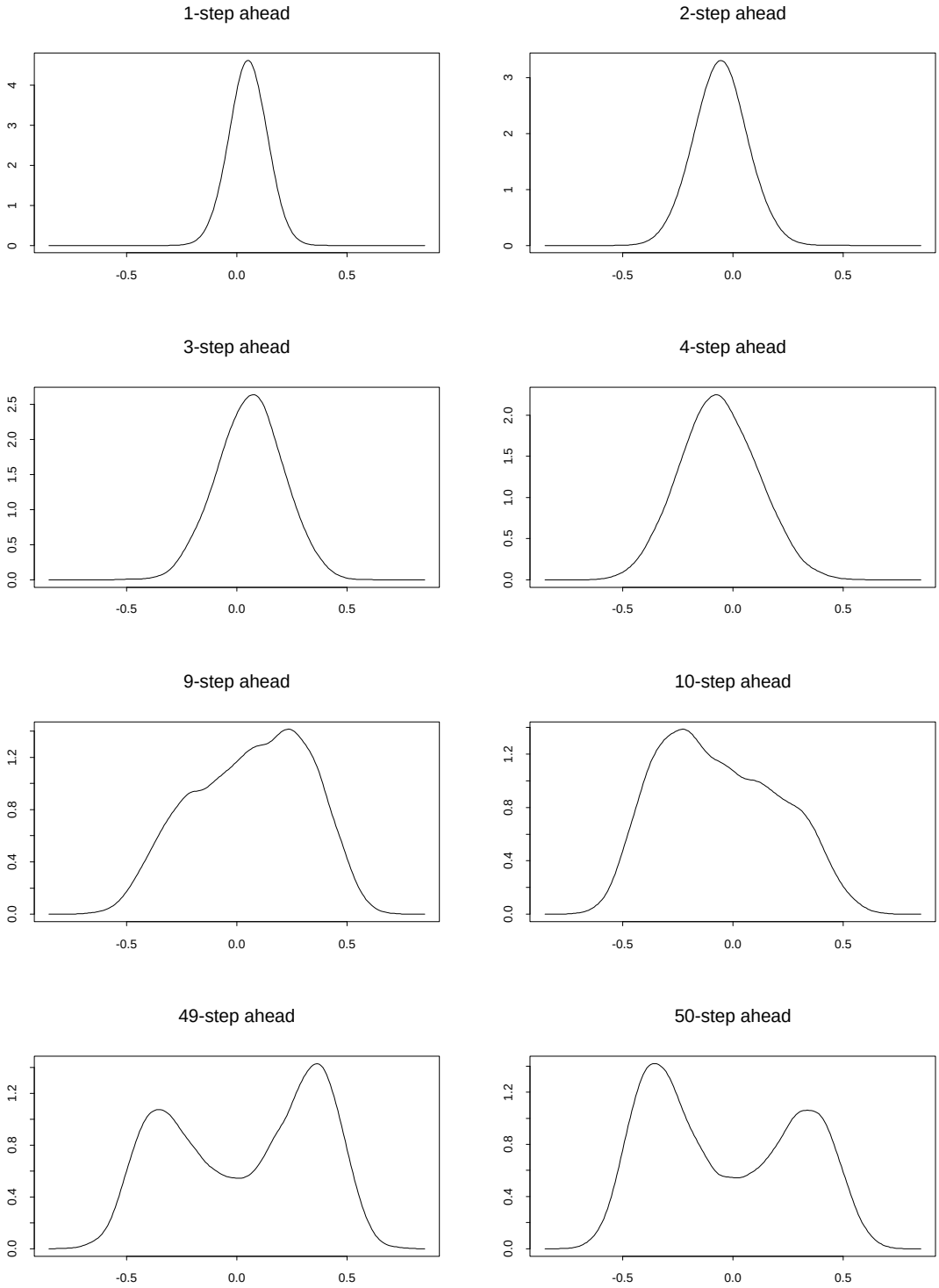


Figure 1: *Forecast densities for the EXPAR model*

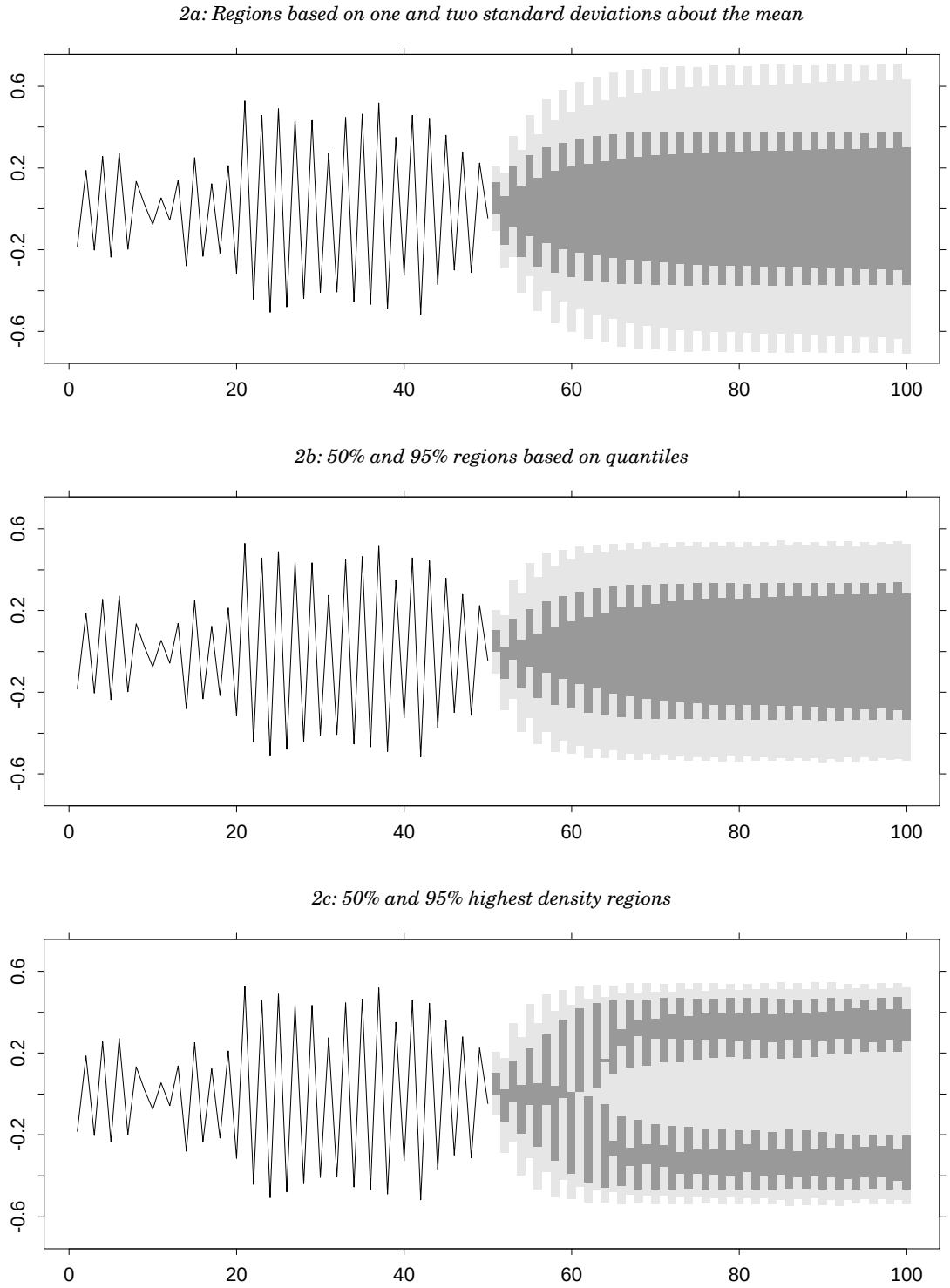


Figure 2: *A realization from the EXPAR model with forecast regions calculated using three different criteria. Only the highest density regions reveal the bimodality of the forecast density.*

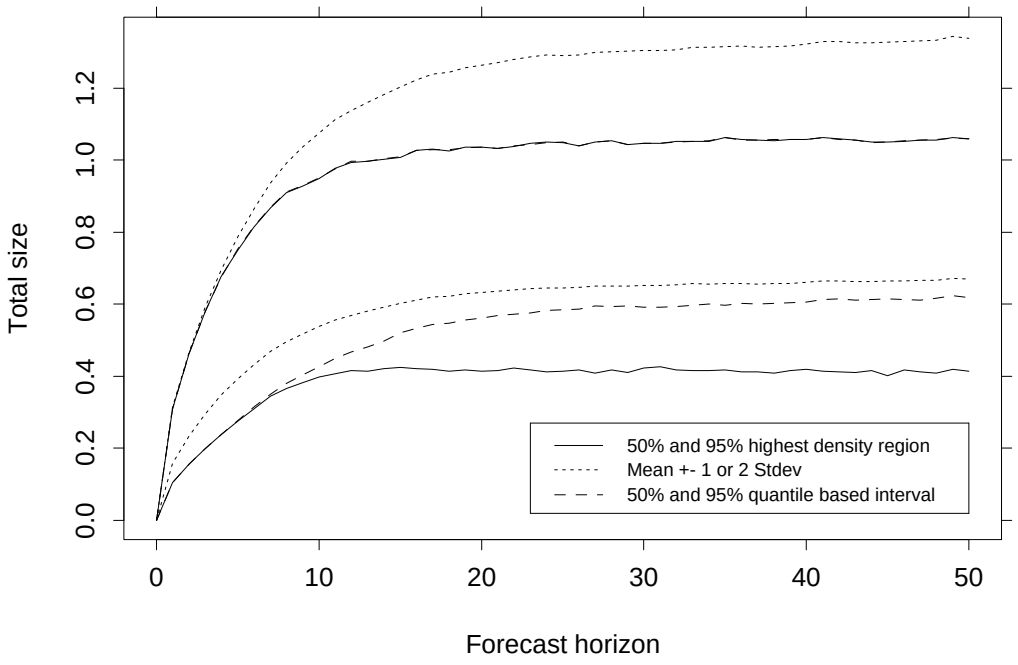


Figure 3: *The total size of the forecast regions shown in Figure 2. The highest density region will, by definition, always produces the smallest forecast regions. The lower three lines correspond to the 50% regions and the upper three lines to the 95% regions. Note that the 95% quantile interval and the 95% HDR region are so close in size that the lines are barely distinguishable.*

(or negative) values and the series switches phase. This occurred at time 9 in this realization. Figure 1 shows some of the forecast densities for the next 50 forecast horizons conditional on the last observed value $X(50) = -0.0462$. Details of how these densities were computed are given in Section 3.1. Note that the densities become increasingly bimodal for longer forecast horizons, thus reflecting the increasing probability of the process switching phase.

Figures 2a and 2b give forecast regions obtained by commonly used methods. Figure 2a shows the mean of the forecast density and the regions formed by one and two standard deviations about the mean. Figure 2b shows the 95% forecast region formed from the 2.5% and 97.5% quantiles and the 50% forecast region formed from the quartiles. Both methods give forecast regions which reflect the oscillations characteristic of the process. But they also mask the bimodality of the forecast distributions.

For multimodal forecast densities, a forecast region consisting of several intervals—one for each mode—is necessary to portray the characteristic shape of the distribution. Such a region is the highest density region (HDR) defined in Section 2. Figure 2c displays 50% and 95% HDRs for this

example. The darker region is the 50% region and the lighter shade denotes the 95% region. These represent the regions of the sample space where the density is greatest. Beyond forecast horizon 14, the modes are sufficiently large that the 50% region consists of two intervals. The smaller interval represents the smaller mode corresponding to those processes that are out of phase with the majority of realizations. The forecast densities are slowly converging to the stationary density consisting of two equal-sized modes.

Figure 3 shows the total size of the forecast regions displayed in Figure 2.

3 Estimation

3.1 Estimating forecast densities

To compute any forecast region it is necessary to first calculate or estimate the forecast density $p_{m|n}$ or some of its properties such as the first two moments or some quantiles. For some models, these can be found analytically. As noted above, for ARIMA models, the forecast densities are normal and the mean and variance can be readily calculated. Also, for many non-linear and non-normal models, the one-step forecast densities are known analytically. But, in general, forecast densities for non-linear and non-normal models are not known.

In the special case of a non-linear and non-normal discrete time autoregression model, Davies et al. (1988) outline a method for finding multi-step forecast densities using recursive numerical quadrature. This idea is extended by Moeanaddan and Tong (1990) and related methods for finding moments are discussed by Pemberton (1987), Tong and Moeanaddin (1988) and Al-Qassam and Lane (1989). However, computing time for these recursive algorithms makes them impracticable except in the simplest first order cases.

By far the quickest and most generally applicable way to obtain an estimate of the multi-step forecast density and its moments is via simulation. The distribution of $X_{m|n}$ may be studied by generating many independent realizations of the process between times t_n and t_m using the observed values of the series, $X(t_1), \ldots, X(t_n)$, as starting values (see, e.g., Granger and Teräsvirta, 1993). Error terms in the model are replaced by a pseudo-random sequence. This approach is of general applicability and is widely used in practice; e.g., it is implemented in the computer package STAR (Tong, 1990) which fits discrete-time threshold models. An alternative approach, which may be

particularly useful if the forecaster wants to make no assumptions about the error density, is to bootstrap the fitted residuals.

Let $x_i(t_m)$ denote the value of the ith realization of $X(t_m)$ calculated in this way. The $x_i(t_m)$ are collected for each $m > n$ and can be used to estimate the moments and other properties of $X_{m|n}$. In addition, the density, $p_{m|n}$, can be estimated empirically from $\{x_i(t_m)\}_{i=1}^n$. Substantial computer memory is saved by updating moment and density estimates after each realization of the process is generated rather than storing all simulated values until the end. Then the empirical density estimates are used to obtain highest density forecast regions.

This approach was used to obtain Figures 1 and 2. In this case, $N = 10000$ and a simple kernel density estimate was used (see, for example, Silverman, 1986 or Scott, 1992).

3.2 Estimating HDRs

The obvious method for calculating an HDR would be to use numerical quadrature with the empirical density, $\hat{p}_{m|n}(x)$ to find the highest density region of given probability. This approach, with some variations, is considered by Wright (1986) and Hyndman (1990).

There is, however, a much quicker way of calculating a HDR due to Hyndman (1996). Note that f_α is the α-quantile of Y where $Y = p(X_{m|n})$ is the random variable formed by transforming $X_{m|n}$ by its own density. So an estimate of R_α can be computed by estimating f_α. We let $\hat{f}_\alpha$ be the α sample quantile of the values in $\{\hat{p}_{m|n}(x_i(t_m))\}_{i=1}^N$ and define $\hat{R}_\alpha$ by $\{x : \hat{p}_{m|n}(x) \geq \hat{f}_\alpha\}$. Then $\hat{f} \to f_\alpha$ and $\hat{p}_{m|n}(x) \to p_{m|n}(x)$ as $N \to \infty$ and so $\hat{R}_\alpha \to R_\alpha$ as $N \to \infty$.

Clearly, this algorithm results in a region in which the actual percentage of points contained in the HDR is equal to the required probability coverage, $1 - \alpha$. The method avoids any numerical quadrature and is relatively quick even for long range forecasting. Furthermore, it is easy to implement with existing software. (The densities and highest density regions displayed in Figures 1 and 2 were calculated in a few minutes using Splus 3.1, © 1993, Statistical Sciences Inc., on a DECstation 5000/25.) If the true forecast density, $p_{m|n}$, is known or easily calculated, it can be used instead of an estimate. Further details of the properties of this algorithm are given in Hyndman (1996). Splus software to compute HDRs using this algorithm is available on the S archive of statlib@lib.stat.cmu.edu.

As is common in forecasting, we have not taken into account the effect of estimating the parameters in the model on the computation of the HDRs. But the same problem exists for all other forecasting regions based on time series models. Even for the standard normal ARIMA models, uncertainty in the parameter estimates is rarely considered when constructing forecast intervals (e.g., Brockwell and Davis, 1991).

4 Graphical presentation

Figure 2 demonstrates a graphical display of forecast regions involving dark and light shaded regions representing regions of different probability. This display is particularly appropriate for highest density regions as higher density shading is used for regions of higher probability density.

An alternative display would be to allow shading density to vary continuously in proportion to the forecast density. While this may provide a little more information concerning the overall shape of the density, it would not be then possible to simultaneously depict any specific forecast intervals.

The plot also enables a rapid assessment of how the shape of the density changes as the forecast lead time increases. This feature is particularly clear in the following example.

Example 2: Nicholson's blowfly data

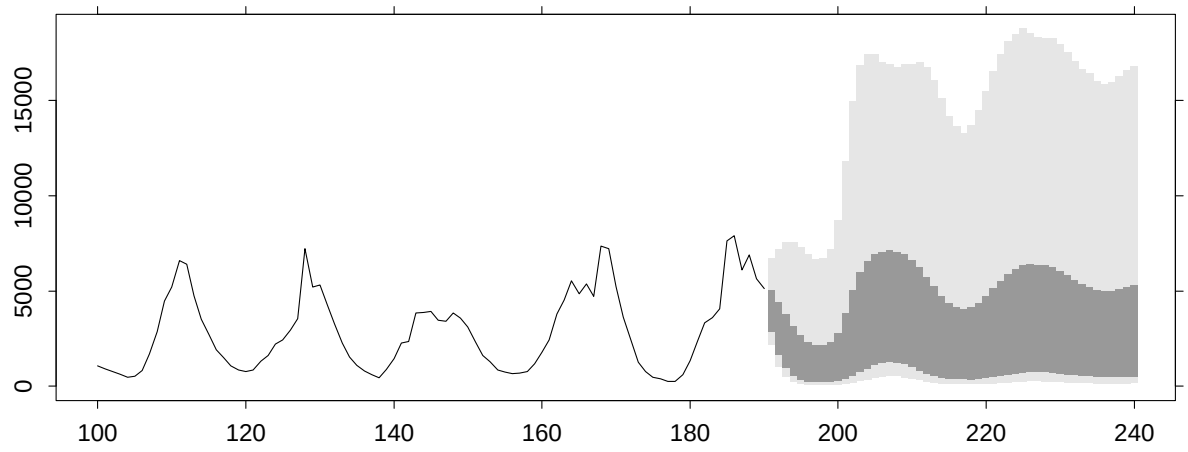
A.J. Nicholson's Australian blowfly data were obtained by recording every two days the population of caged blowflies on a strictly controlled diet (Nicholson, 1957). The data are given in Brillinger et al. (1980) and Tong (1990). Here we consider the first 190 points (380 days) of the series. The most striking feature of these data are aperiodic population cycles induced by the experimental conditions. A few of these cycles can be seen on the left of Figure 4.

Let X_t represent the log (base 10) of the population at time t. Tong (1988) proposed the following SETAR model for these data:

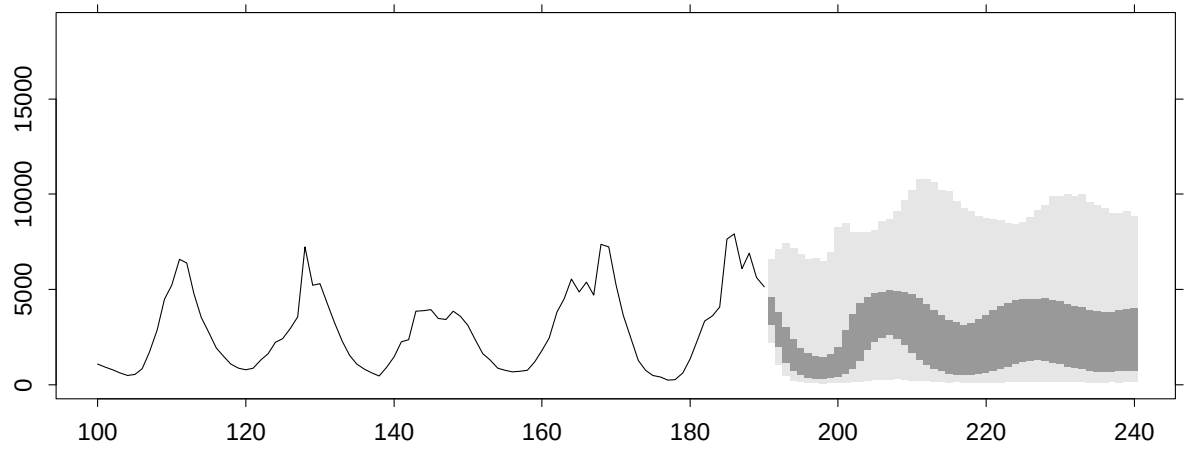
$$X_t = \begin{cases} 2.65 + 0.27X_{t-1} + e_t & X_{t-8} \leq 3.05 \\ 0.48 + 1.40X_{t-1} - 0.19X_{t-2} - 0.36X_{t-3} + e_t & X_{t-8} > 3.05 \end{cases}$$

where the variance of e_t is 0.0148.

5a: Regions based on one and two standard deviations about the mean



5b: 50% and 95% regions based on quantiles



5c: 50% and 95% highest density regions

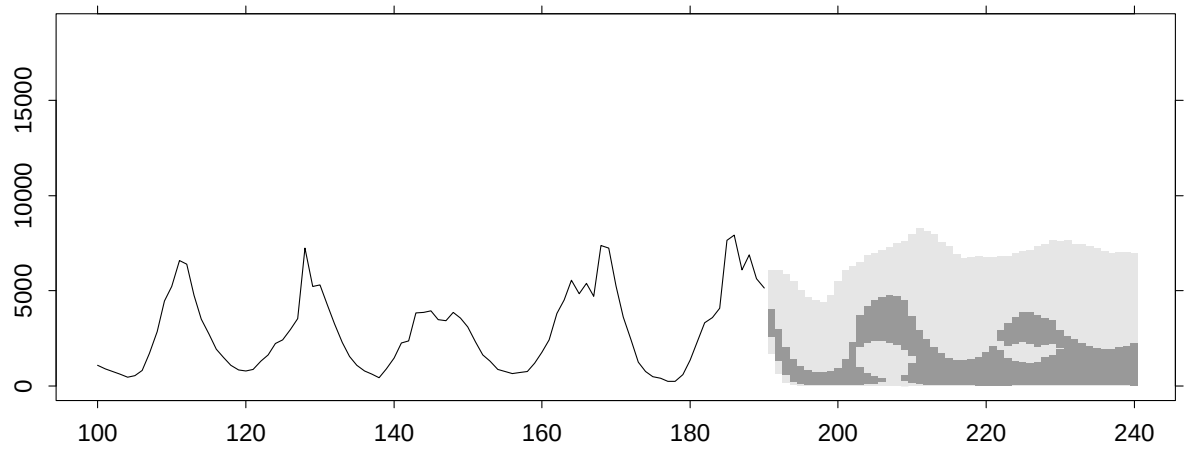


Figure 4: *Forecast regions for Nicholson's blowfly data using Tong's model.*

The forecast densities were calculated using the method outlined in Section 3.1 and transformed back to the scale of the original data.

Let $F(x)$ be the distribution function on the original scale and $G(y)$ be the distribution function on the log scale. We have estimated $G'(y)$ and wish to find $F'(x)$. Now $F(x) = Pr(X \leq x) = Pr(\log(X) \leq \log(x)) = G(\log(x))$. Hence $F'(x) = G'(\log(x))/x$. Replacing $G'(y)$ by its estimate, we obtain an estimate of $F'(x)$. This was then used to compute the highest density forecast regions and quantile-based regions. The forecast interval symmetric about the mean was computed for the logged data from the estimate of $G'(y)$ and the end-points of the interval were back-transformed to the original scale. The method preserves the probability coverage of the region (since quantiles are invariant under monotonic transformation) and ensures the intervals lie on the non-negative real line. The resulting forecast regions are given in Figure 4.

Note the bimodality in the forecast densities from time 203; the left hump becomes narrower until time 207 and then widens again so that by time 211 the humps corresponding to each mode have become sufficiently small that the 50% HDR contains both modes. The same phenomenon occurs in the next predicted cycle. This bimodality possibly represents the potential for the process to switch phase. For example, around time 207, the process is most likely to be near a maximum, with a small probability of it being near a minimum and even smaller probability of being half-way between the two. Of course, only highest density regions are able to display these complex changes in the shape of the forecast densities. It is not known if the potential to switch phase is a biological phenomenon or an artefact of the model.

Another point of interest here is the non-monotonicity of the forecast intervals: they are narrowest at the bottom of the trough in the cycle. Unlike linear forecasts, it is possible with non-linear models to obtain forecasts further ahead which are more reliable than short-term forecasts. This phenomenon is also documented by Tong and Moeanaddin (1988), Pemberton (1989) and Tong (1990, p.349).

5 Conclusions

We have considered several advantages of highest density forecast regions in non-linear and non-normal time series modelling, namely that they are the region of smallest total size, that they represent the most likely region in which future observations will fall and that they can portray

asymmetry and multimodality. We have also seen that they can be relatively easily estimated.

There are some implications in this discussion for point forecasts and for measuring non-linear forecasting accuracy. Traditionally, we have used the mean or median for point forecasts and the mean square error, mean absolute percentage error or some similar statistic to measure forecast accuracy (e.g., Makridakis et al., 1983). However, if the forecast densities are strongly bi-modal, as in the EXPAR process considered here, the mean and median fall in a region of very low density and so do not represent where the process is 'likely' to be in the future. Similarly, their associated measures of accuracy may be quite misleading. In such cases, the mode would appear to be a better point forecast and a more appropriate measure of accuracy would be based on the height of the forecast density at the point of observation.

6 Acknowledgements

I would like to thank Dr Gary Grunwald and three referees for their comments in connection with this material.

References

Al-Qassam, M. S. and J. A. Lane (1989) Forecasting exponential autoregressive models of order 1, *Journal of Time Series Analysis*, **10**(2), 95–113.

Box, G. E. P. and G. C. Tiao (1973) *Bayesian inference in statistical analysis*, Addison-Wesley, Reading, MA.

Brillinger, D. R., J. Guckenheimer, P. E. Guttorp and G. Oster (1980) Empirical modelling of population time series data: the case of age and density dependent vital rates, in *Lectures on Mathematics in the Life Sciences*, vol. 13, pp. 65–90, American Mathematical Society.

Brockwell, P. J. and R. A. Davis (1991) *Time series: theory and methods*, Springer-Verlag, New York, 2nd ed.

Broemeling, L. D. and P. Cook (1992) Bayesian analysis of threshold autoregressions, *Commun. Statist. Theory Meth*, **21**(9), 2459–2482.

Davies, N., J. Pemberton and J. D. Petruccelli (1988) An automatic procedure for identification,

estimation and forecasting univariate self exciting threshold autoregressive models, *The Statistician*, **37**, 199–204.

De Gooijer, J. G. and K. Kumar (1992) Some recent developments in non-linear time series modelling, testing and forecasting, *International Journal of Forecasting*, **8**, 135–156.

Edwards, W., H. Lindman and L. J. Savage (1963) Bayesian statistical inference for psychological research, *Psychological Review*, **70**, 193–242.

Granger, C. W. J. and T. Teräsvirta (1993) *Modelling nonlinear economic relationships*, Oxford University Press, New York.

Hyndman, R. J. (1990) An algorithm for constructing highest density regions, unpublished manuscript available from the author.

Hyndman, R. J. (1996) Computing and graphing highest density regions, *The American Statistician*, **50**, 120–126.

Lindley, D. V. (1965) *Introduction to probability and statistics from a Bayesian perspective*, Cambridge University Press.

Makridakis, S., S. C. Wheelwright and V. E. McGee (1983) *Forecasting: methods and applications*, John Wiley and Sons, New York, 2nd ed.

Moeanaddan, R. and H. Tong (1990) Numerical evaluation of distributions in non-linear autoregression, *Journal of Time Series Analysis*, **11**(1), 33–48.

Nicholson, A. J. (1957) The self-adjustment of populations to change, *Cold Spring Harbour Symp. Quant. Biol.*, **22**, 153–173.

Pemberton, J. (1987) Exact least squares multi-step prediction from nonlinear autoregressive models, *Journal of Time Series Analysis*, **8**(4), 443–448.

Pemberton, J. (1989) Forecasting accuracy of non-linear time series models, Tech. rep., Worcester Polytechnic Institute, MA.

Ray, D. (1988) Comparison of forecasts: An empirical investigation, *Sankhyā B*, **50**, 258–277.

Scott, D. W. (1992) *Multivariate density estimation: theory, practice, and visualization*, John Wiley, New York.

Silverman, B. W. (1986) *Density estimation for statistics and data analysis*, Chapman and Hall, London.

Tong, H. (1988) Non-linear time series modelling in population biology: a preliminary case study, in *Nonlinear time series and signal processing*, vol. 106 of *Lecture Notes in Control and Information Sciences*, pp. 75–87, Springer-Verlag, Berlin.

Tong, H. (1990) *Non-linear time series: a dynamical system approach*, Oxford University Press, New York.

Tong, H. and R. Moeanaddin (1988) On multi-step non-linear least squares prediction, *The Statistician*, **37**, 101–110.

Wright, D. E. (1986) A note on the construction of highest posterior density intervals, *Applied Statistician*, **35**(1), 49–53.