Lang, Fronk, Fahrmeir:

Function estimation with locally adaptive dynamic models

Sonderforschungsbereich 386, Paper 263 (2001)

Online unter: http://epub.ub.uni-muenchen.de/

Projektpartner

Function estimation with locally adaptive dynamic models

Stefan Lang, Eva–Maria Fronk and Ludwig Fahrmeir

University of Munich, Ludwigstr. 33, 80539 Munich

Summary

We present a nonparametric Bayesian method for fitting unsmooth and highly oscillating functions, which is based on a locally adaptive hierarchical extension of standard dynamic or state space models. The main idea is to introduce locally varying variances in the state equations and to add a further smoothness prior for this variance function. Estimation is fully Bayesian and carried out by recent MCMC techniques. The whole approach can be understood as an alternative to other nonparametric function estimators, such as local or penalized regression with variable bandwidth or smoothing parameter selection. Performance is illustrated with simulated data, including unsmooth examples constructed for wavelet shrinkage, and by an application to sales data. Although the approach is developed for classical Gaussian nonparametric regression, it can be extended to more complex regression problems.

Keywords: adaptive smoothing, MCMC, nonparametric Bayesian regression, random walk priors, unsmooth functions, variable smoothing parameter

1 Introduction

Nonparametric methods for fitting smooth curves, such as kernel, local or spline regression, are now widely available and accepted. However these

methods can have bad performance when estimating unsmooth functions which have jumps, edges, or are highly oscillating. Nonparametric regression approaches that adapt to such spatial heterogeneity have gained considerable interest. Local regression with variable bandwidth (Fan and Gijbels, 1995) or adaptive ridging (Seifert and Gasser, 2000) and wavelet shrinkage regression (Donoho and Johnstone, 1994) are two prominent approaches.

In this paper, we present a Bayesian nonparametric approach, which is more closely related to spline fitting with locally adaptive penalties. Abramovich and Steinberg (1996) generalize the common penalized least squares criterion for smoothing splines with a global smoothing parameter by introducing a variable smoothing parameter into the roughness penalty. For estimation, they propose a two-step procedure: First a smoothing spline is fitted with a constant smoothing parameter chosen by generalized cross-validation. Then an estimate for the variable smoothing parameter is constructed, based on the derivatives of this pilot estimate, and is plugged into their locally adaptive penalty to fit the smoothing spline in a second step. Ruppert and Carroll (2000) propose P-splines based on a truncated power series basis and difference penalties on the regression coefficients with locally adaptive smoothing parameters. The latter are obtained by linear interpolation from a smaller number of smoothing parameters, defined for a subset of knots and estimated by generalized cross-validation.

Our approach is fully Bayesian and uses a two-stage prior for the unknown regression function. The first stage are first or second order random walk models as a discretized Bayesian version of the common roughness penalty for smoothing splines. The second stage consists of analogous smoothness priors for varying variances of the random walk model errors used in the first stage. Varying variances correspond to variable smoothness parameters and make the prior more flexible for modelling functions with differing curvature. A similar idea appears in stochastic volatility models (e.g., Shephard and Pitt, 1997), where variances of observation errors follow a similar stochastic process prior.

Alternative Bayesian models that allow for varying variances are robustified state space models with (discrete or continuous) mixtures of normals as error distributions, see Carter and Kohn (1996a, 1996b) for a fully Bayesian approach or Fahrmeir and Künstler (1999) for posterior mode estimation. These models have good performance for regression functions with jumps. We compare them to our approach in a simulation study in Section 4. Lang and Brezger (2001) propose such normal mixtures for spatially adaptive Bayesian P-spline fitting.

Other competing methods are regression splines with adaptive Bayesian knot or variable selection, see Smith and Kohn (1996, 1997), Denison, Mallick and Smith (1998) and Biller (2000). The number and the location of knots are unknown random variables, and function estimation becomes flexible through model averaging.

The rest of the paper is organized as follows. Locally adaptive dynamic models with varying variances are introduced in Section 2. Inference is fully Bayesian and uses recent MCMC techniques, combining Gibbs sampling with efficient block moves (Rue, 2001) for functions and a Metropolis-Hastings (MH) algorithm with conditional prior proposals (Knorr-Held, 1999) for varying variances. Details are given in Section 3. Performance is investigated in Section 4 by a simulation study. In particular, we compare our approach with fitting based on models with variances changing independently according to a mixture of normals and with results in Ruppert and Carroll (2000). The method is illustrated with an application to sales data analysed previously in the literature.

Although we focus here on a simple Gaussian observation model suitable for one-dimensional curve fitting, the basic idea can be adapted to more general settings. Some of the resulting extensions, for example to non-Gaussian observations or to surface estimation, are mentioned in the conclusions.

2 Locally adaptive dynamic models

Consider first the classical smoothing problem for a response variable, where observations $y = (y_1, y_2, \ldots, y_T)'$ are assumed to be the sum

$$y_t = \alpha_t + \epsilon_t, \qquad t = 1, \ldots, T \tag{1}$$

of a smooth trend function or regression curve, evaluated at the observation or design points t, and independent Gaussian errors $\epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2)$. We denote the vector of the function evaluations by $\alpha = (\alpha_1, \ldots, \alpha_T)'$ and use the same symbol for the whole curve. For simplicity, we assume equidistant design points. Extensions to non–equally spaced designs are shortly outlined at the end of the section. In a standard dynamic or state space modelling approach for estimation of the unknown function α, more exactly its evaluations α_t, the observation model (1) is supplemented by a Gaussian random walk of first order

$$\alpha_t \quad = \quad \alpha_{t-1} + u_t, \qquad\qquad u_t \sim \mathrm{N}(0; q^2) \tag{2}$$

or of second order

$$\alpha_t \quad = \quad 2\alpha_{t-1} - \alpha_{t-2} + u_t, \quad u_t \sim \mathrm{N}(0; q^2), \tag{3}$$

denoted as RW(1) respectively as RW(2). The errors u_t are mutually independent and independent of observation errors ϵ_t. In addition we will assume diffuse priors for initial values α_1, α_2. From a Bayesian point of view, the random walk models (2) and (3) define smoothness priors on first and second differences $\alpha_t - \alpha_{t-1} = u_t$ respectively $\alpha_t - 2\alpha_{t-1} + \alpha_{t-2} = u_t$ that help to regularise the estimation problem. For given variances σ_ϵ^2 and q^2 the famous linear Kalman filter and smoother computes the posterior

means $\hat{\alpha}_t = \mathrm{E}(\alpha_t \mid y_1, \ldots, y_T)$ as optimal smoothers, together with posterior variances. Since the posterior is Gaussian, mean and mode coincide, and therefore the estimates $\hat{\alpha}_t$, $t = 1, \ldots, T$, can also be obtained by maximizing the posterior. Taking logarithms, this leads to the classical optimal smoothing problem already considered by Whittaker (1923): Choose $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_T)$ as the minimizer of

$$\sum_{t=1}^{T}(y_t - \alpha_t)^2 + \frac{\sigma_\epsilon^2}{q^2}\sum_{t=2}^{T}(\alpha_t - \alpha_{t-1})^2 \tag{4}$$

for model (2), and

$$\sum_{t=1}^{T}(y_t - \alpha_t)^2 + \frac{\sigma_\epsilon^2}{q^2}\sum_{t=3}^{T}(\alpha_t - 2\alpha_{t-1} + \alpha_{t-2})^2 \tag{5}$$

for model (3). From (4) and (5), the close correspondence to spline smoothing becomes clear: The ratio $\lambda = \sigma_\epsilon^2/q^2$ is a *global* smoothing parameter and the penalty terms are the discretized versions of corresponding penalty terms for quadratic and cubic smoothing splines, see Kohn and Ansley (1987) for details. Already with a moderate number of observations, estimates $\hat{\alpha}_t$ are practically undistinguishable from spline smoothing estimates.

Priors (2), (3) and penalties (4), (5) can also be written in matrix notation. Priors for $\alpha = (\alpha_1, \ldots, \alpha_T)'$ are multivariate Gaussian,

$$p(\alpha|q^2) \propto \exp(-\frac{1}{2q^2}\alpha' K \alpha), \tag{6}$$

where the precision matrix K has band structure. For a RW(1) model, with a diffuse prior for the initial value, it is tridiagonal and given by

$$K = D_1' D_1,$$

where D_1 is the $(T-1) \times T$ upper two-diagonal matrix with entries (-1,1) defining the vector of first differences $D_1\alpha = (\alpha_2 - \alpha_1, \ldots, \alpha_T - \alpha_{T-1})'$. For a RW(2) model, $K = D_2' D_2$ is pentadiagonal, with $D_2\alpha = D_1(D_1\alpha)$ defining the vector of second differences. Note that $rank(K) = T - 1$ or $T - 2$, respectively.

The basic idea for estimation of nonsmooth functions is to replace the constant variance q^2 in (2) and (3) by locally varying variances q_t^2 that are considered as evaluations of a variance function q. This corresponds to replacing the global smoothing parameter by a *local* smoothing parameter $\lambda_t = \sigma_\epsilon^2/q_t^2$. To estimate the unknown variance function automatically together with the unknown curve α, we reparametrize by

$$h_t = \log(q_t^2) \iff q_t^2 = \exp(h_t)$$

and add a second smoothness prior in form of first or second order differences for $h = (h_k, \ldots, h_T)'$, implying also a smoothness prior for $q = (q_k^2, \ldots, q_T^2)'$. The index k depends on the choice of the prior for α, for a RW(1) $k = 2$ and for a RW(2) $k = 3$. Thus we will arrive at the following *locally adaptive dynamic models.*

Observation model for $y = (y_1, \ldots, y_T)'$:

$$y_t = \alpha_t + \epsilon_t, \quad \epsilon \sim \mathrm{N}(0, \sigma_\epsilon^2) \tag{7}$$

Smoothness priors for $\alpha = (\alpha_1, \ldots, \alpha_T)'$:

$$\alpha_t = \alpha_{t-1} + u_t, \quad \text{or} \quad \alpha_t = 2\alpha_{t-1} - \alpha_{t-2} + u_t \tag{8}$$

with $u_t \sim \mathrm{N}(0, \exp(h_t))$.
Smoothness priors for $h = (h_k, \ldots, h_T)$, $k = 2, 3$:

$$h_t = h_{t-1} + \eta_t, \quad \text{or} \quad h_t = 2h_{t-1} - h_{t-2} + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2). \tag{9}$$

We assume mutually independent errors ϵ_t, u_t and η_t and diffuse priors for initial values α_1, α_2, h_k, h_{k+1}. The model definition is completed by the common assumption of independent inverse Gamma hyperpriors

$$\sigma_\epsilon^2 \sim \mathrm{IG}(a_1, b_1), \quad \sigma_\eta^2 \sim \mathrm{IG}(a_2, b_2) \tag{10}$$

for the variances σ_ϵ^2, σ_η^2. By appropriate choice of a_1, b_1, a_2, b_2, these hyperpriors are made highly dispersed.

Again, smoothness priors for α and h can be written in multivariate form. For α, we obtain

$$p(\alpha|h) \propto \exp(-\frac{1}{2}\alpha' K \alpha)$$

where $K = D_d' Q D_d$, $d = 1, 2$, and the diagonal matrix Q has entries $1/q_t^2 = 1/\exp(h_t^2)$. Priors for h are completely analogous to (6), i.e.

$$p(h|\sigma_\eta^2) \propto \exp(-\frac{1}{2\sigma_\eta^2} h' L h)$$

where $L = D_d' D_d$, $d = 1, 2$.

As mentioned in the introduction, models with independently changing variances may be another alternative for fitting regression functions with jumps. Reparameterizing q_t^2 to $\gamma_t q^2$, we assume a continuous mixture of normals

$$u_t|\gamma_t, q^2 \sim N(0, \gamma_t q^2) \tag{11}$$

for the error distributions. Assuming i.i.d. inverse gamma priors

$$\gamma_t \sim IG(\nu/2, \nu/2), \tag{12}$$

the marginal distribution of the errors is a Student distribution with ν degrees of freedom. The case $\nu = 1$ of a Cauchy distribution is of special interest as a robust prior and is used for the rest of this paper. The multivariate (conditional) Gaussian form is the same as for variances with RW-priors. For the scale parameter q^2, we make the usual prior assumption $q^2 \sim IG(a, b)$.

Figure 1 illustrates the improvement in function estimation that can be achieved through locally adaptive models with variable smoothing parameter. The underlying "true" function is the so called Doppler function, which is highly oscillating near the origin, see Section 4.2. The function estimates in Figure 1 are taken from the simulation experiments in Section 4.2, and both runs correspond to the median in the $log_{10}(MSE)$ boxplots in Figure 9 b). The advantages of locally adaptive models are obvious: Whereas variable variances can adapt to changes in curvature of the underlying function, models with global variance are too rough in the areas where the function is smooth and have problems in following rapid oscellations near the origin. Section 4 also compares our approach to models with independently varying variances. Again, our models show inproved estimation properties.

Although we will focus on the locally adaptive model (7)-(10), some extensions are immediate: First, we may generalize the observation model (7) and the state equation (8) to the standard form $y_t = z_t'\alpha_t + \epsilon_t,\ \alpha_t = F_t\alpha_{t-1} + u_t$ of linear Gaussian state space models. Assuming again varying variances for u_t and appropriate hyperpriors as in (9), we obtain *locally adaptive state space models*. Secondly, *unequally spaced observations* can be dealt with by adjusting the variances $q_t^2 = \exp(h_t)$. For example in the case of a first order random walk, the necessary modification is $q_t^2 = \Delta_t \exp(h_t)$, where Δ_t is the difference between the t-th and the $(t-1)$-th design point.

3 Nonparametric Bayesian inference via MCMC

Fully Bayesian inference is based on the posterior of the unknowns given the data $p(\alpha, h, \sigma_\epsilon^2, \sigma_\eta^2|y)$. Due to the hierarchical structure of the model summarized in (7) - (10) its unnormalized form is easily derived as

$$p(\alpha, h, \sigma_\epsilon^2, \sigma_\eta^2|y) \propto p(y|\alpha, \sigma_\epsilon^2)p(\alpha|h)p(h|\sigma_\eta^2)p(\sigma_\epsilon^2)p(\sigma_\eta^2).$$

To sample from the posterior, we use a hybrid MCMC algorithm. The parameters $\alpha = (\alpha_1, \ldots, \alpha_T)$ are sampled as an entire block from their Gaussian full conditionals $p(\alpha|\cdot)$ given the other parameters and the data. From the Gaussian observation model and the prior for α we obtain

$$p(\alpha|\cdot) \propto \exp\left(-\frac{1}{2\sigma_\epsilon^2}(y-\alpha)'(y-\alpha)\right)\exp\left(-\frac{1}{2}\alpha' K \alpha\right)$$

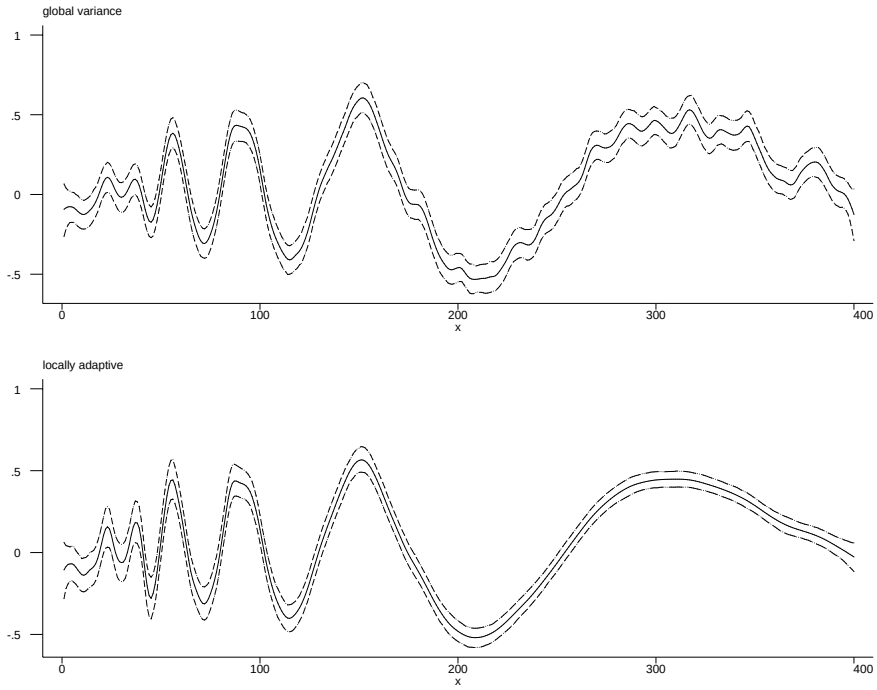

Figure 1: *Illustration of the improvement of locally adaptive models with variable smoothing parameter (lower panel) compared to models with a global smoothing parameter (upper panel). Shown are the respective posterior means together with 80 % pointwise credible intervals.*

From this it is easily derived that the posterior is multivariate Gaussian $\alpha|\cdot \sim N(\mu, \Sigma)$ with

$$\mu = \frac{1}{\sigma_\epsilon^2}\Sigma\, y, \qquad \Sigma = \left(\frac{1}{\sigma_\epsilon^2}I + K\right)^{-1}.$$

Since the precision matrix $P = I/\sigma_\epsilon^2 + K$ is a band matrix, posterior samples for α can be efficiently drawn by using a Cholesky decomposition of P as suggested in Rue (2001). The resulting $O(T)$ forward-backward sampling scheme is closely related to the block move Gibbs samplers of Carter and Kohn (1994) and Frühwirth-Schnatter (1994).

Full conditionals for the variance function $h = (h_k, \ldots, h_T)'$, $k = 2, 3$, are not in closed form. We use an MH-algorithm with conditional prior proposals of Knorr-Held (1999) for drawing from the full conditionals $p(h_{rs}|\cdot)$ for blocks $h_{rs} = (h_r, \ldots, h_s)'$, partitioning h into a sequence of subvectors. MH steps consist of drawing a proposal h_{rs}^* from the conditional prior and accepting it

with probability

$$\min\left(1, \frac{\prod_{t=s}^{r} p(\alpha_t|\alpha_{l<t}, h_t^\star)}{\prod_{t=s}^{r} p(\alpha_t|\alpha_{l<t}, h_t)}\right).$$

The conditional distributions $p(\alpha_t|\alpha_{l<t}, h_t)$ are Gaussian, defined by the priors (2) or (3). The conditional distribution of h_{rs} given the rest h_l, $l \notin \{r, \ldots, s\}$, is a multivariate Gaussian distribution. Its mean and covariance matrix can be written in terms of the precision matrix L of h. Let L_{rs} denote the submatrix of L, given by the rows and columns numbered r to s and let $L_{1,r-1}$ and $L_{s+1,T}$ denote the matrices left and right of L_{rs}. Then the (conditional) mean μ_{rs} and the covariance matrix Σ_{rs} are given by

$$\mu_{rs} = \sigma_\eta^2 \begin{cases} -{L_{rs}}^{-1} L_{s+1,T} h_{s+1,T} & r = k \\ -{L_{rs}}^{-1} L_{k,r-1} h_{k,r-1} & s = T \\ -{L_{rs}}^{-1}(L_{k,r-1} h_{1,r-1} + L_{s+1,T} h_{s+1,T}) & \text{else} \end{cases}$$

and

$$\Sigma_{rs} = \sigma_\eta^2 {L_{rs}}^{-1},$$

respectively. Details about efficient computation of the mean μ_{rs} and the choice of the block size $s - r + 1$ can be found in Fahrmeir and Lang (2001).

The full conditionals for the variance parameters σ_ϵ^2 and σ_η^2 are inverse gamma distributions given by

$$\sigma_\epsilon^2 \mid y, \alpha \sim \text{IG}\left(a_1 + \frac{T}{2},\, b_1 + \frac{1}{2}\sum_{t=1}^{T}(y_t - \alpha_t)^2\right)$$

and

$$\sigma_\eta^2 \mid h \sim \text{IG}\left(a_2 + \frac{rank(L)}{2},\, b_2 + \frac{1}{2}h'Lh\right).$$

Thus, updating of σ_ϵ^2 and σ_η^2 can be done by simple Gibbs steps.

In the case of a hierarchical t-formulation outlined in (11) and (12) we have to update the independent weights γ_t instead of variance parameters h_t. In contrast to the h_t's the full conditionals for γ_t are known distributions. The inverse gamma distribution of $\gamma_t|\cdot$ is given by

$$p(\gamma_t|\cdot) \sim IG(\frac{\nu}{2} + \frac{1}{2}, \frac{\nu}{2} + \frac{u_t^2}{2q^2})$$

where u_t is the error term in (8). Similar to the γ_t's the full conditional of q^2 is again an inverse gamma distribution, i.e.

$$q^2 \sim IG(a + \frac{rank(K)}{2}, b + \frac{1}{2}\alpha' K \alpha),$$

with $K = D_d' Q D_d$ and the diagonal matrix $Q = diag(1/\gamma_k, \ldots, 1/\gamma_T)$, $k = 2, 3$. Once again, updating of γ_t and q^2 can be done by simple Gibbs steps.

4 Simulations und Application

To gain experience and to study performance, we applied our locally adaptive approach to simulated and real data. Section 4.1 reports on results for unsmooth functions constructed for wavelet shrinkage by Donoho and Johnstone (1994). In Section 4.2 we present and compare results for highly oscillating functions given in Ruppert and Carroll (2000) and others. Section 4.3 contains an application to a time series of sales data from West and Harrison (1989).

4.1 Unsmooth functions

To demonstrate the practicability of our approach for unsmooth functions we use a variant of the blocks function considered by Donoho and Johnstone (1994). The true function consists of 256 time points and is shown in Figure 2 together with noisy observations generated according to the Gaussian observation model (1). For simulations we used $\sigma_\epsilon^2 = 0.95^2$ which corresponds to a signal to noise ratio of 4. We simulated 250 replications of the model. Usually the best results were obtained with a RW(1) prior for α and a RW(1) prior for varying variances (denoted in the following as rw1vrw1). The presentation is therefore restricted to this case. For comparison we also present results obtained by using the hierarchical t-formulation in (11) and (12) (trw1), and using a simple RW(1) with global variance (rw1).

Figure 3 shows function estimates averaged over the 250 replications for rw1vrw1 (a), trw1 (b) and rw1 (c). For comparison the true function is always included in the plots (solid lines). Figure 4 displays boxplots of $log_{10}(MSE)$ where the emirical mean squared error is defined by $MSE = 1/T \sum_{t=1}^{T}(\alpha_t - \hat{\alpha_t})^2$. Figure 3 reveals that both approaches with locally adaptive variances perform more or less equally well but are less biased than the simple approach with a global variance. In terms of the MSE the hierarchical t-formulation trw1 slightly outperforms the approach rw2vrw1 where variances are dependent. Additionally, Figure 5 shows the average of estimated variances and weights for rw1vrw1 and trw1 based on the respective posterior mean estimates. Obviously, the discontinuities of the blocks function are well detected in both approaches.

We also investigated the coverage of pointwise credible intervals. In a Bayesian approach based on MCMC simulation techniques credible intervals are estimated by computing the respective quantiles of the sampled function evaluations. For a nominal level of 80% the average coverage is approximately 81% for the approach with global variance and 87% for both approaches with locally adaptive variances. This indicates that our Bayesian approach yields rather conservative credible intervals. We should stress, however, that in the case of a RW(2) prior for α the average coverage is below the nominal level for all estimators although for the approaches with locally adaptive variances the average coverage is very close to the nominal level (78%).

4.2 Highly oscillating functions

In order to demonstrate the usefullness of our approach for estimating highly oscillating curves, we mainly refer to Ruppert and Carroll (2000) who propose P-splines based on a truncated power series basis and quadratic penalties on the regression coefficients with locally adaptive smoothing parameters. In their first simulation example they used the functions shown in Figure 6. Figure a) corresponds to a function with low spatial variability and Figure c) to a function with severe spatial variability. Figures b) and d) show noisy observations generated according to the Gaussian observation model (1) with $\sigma_\epsilon^2 = 0.2^2$. As for the blocks function we simulated 250 replications of both models. In contrast to unsmooth functions, like the blocks function investigated in Section 4.1, for highly oscillating but smooth functions the best results are usually obtained with a RW(2) prior for α and still a RW(1) prior for h (rw2vrw1). In analogy to the previous section we compare our approach with the hierarchical t-formulation in (11) and (12) (trw2), and using a simple RW(2) prior with global variance (rw2).

Figures 7 and 8 compare the true functions (solid lines) with the average function estimates (dashed lines) for low respectively severe spatial variability. Fe focus on rw2vrw1 (a), trw2 (b) and rw2 (c). Figure 9 display boxplots of $log_{10}(MSE)$. Figure 10 shows for rw2vrw1 and trw2 the average estimated variances and weights based and the respective posterior mean estimates (Doppler with severe spatial variability only). From Figures 7 - 10 we can draw the following conclusions: For low spatial variability the estimators with global and locally adaptive variance perform equally well in terms of the bias. However, in terms of the MSE our proposed esimator rw2vrw1 is clearly superior to the estimator with global variance and the hierarchical t-formulation. For severe spatial variability the estimators rw2vrw1 and trw2 obviously reduce the estimation bias compared to the estimator rw2 with global variance. The lowest MSE is obtained with rw2vrw1. The estimators trw2 and rw2 seem to have more or less identical MSE's.

For the Doppler function with low spatial variability, Ruppert and Carroll (2000) obtained a value of approximately -1.5 for the median of $log_{10}(MSE)$. Both their global and local penalty estimator perform equally well in this situation. For the function with severe spatial variability their local penalty estimator has superior performance compared to their global penalty estimator with a median value of approximately -1.25 for $log_{10}(MSE)$. They claim that their estimator performs slightly better than the Bayesian method of Smith and Kohn (1996) and the stepwise selection method of Stone et al. (1997). All of our estimators outperform their P-splines approach by far.

The Doppler functions were also used in a simulation study in Lang and Brezger (2001). For low spatial variability the Bayesian P-splines approach therein and the approach presented in this paper perform equally well. For severe spatial variability, however, the estimator rw2vrw1 performs slightly better than the best estimator in Lang and Brezger (2001).

In analogy to the blocks function we also investigated the coverage of pointwise credible intervals. Average coverage rates for the various estimators range from 84% to 85% for the function with low spatial variability and from 81% to 83% for the function with severe spatial variability. Hence, for all estimators the average coverage is slightly above the nominal level of 80%.

4.3 CP6 sales data

The monthly CP6 sales data (West and Harrison, 1989) shown on the left in Figure 11 indicate an additive outlier and a change of the slope in December 1955 as well as further change points in January 1957 and 1958. The best fit for the trend α displayed in Figure 11 was obtained here by a second order random walk for α and a first order random walk for h. Goodness of fit is measured by the deviance information criteria (DIC) recently proposed for Bayesian model comparison by Spiegelhalter et al. (2001). Adaption to the change points and to smooth trends between them seems to be quite adequate. Also the changepoints are detected by the peaks in the fit of the variance function. The estimates are based on a sample size of 105000 iterations and a burn in period of 5000 iterations. For estimation we used every 100th sampled parameter after the burn in period. The block sizes for sampling the h_t's were chosen randomly in every iteration between 10 and 15. The computation took only a few seconds on a Pentium III computer with 850 Mhz. To demonstrate the mixing behaviour of our proposed block move samplers, we display in Figure 12 the sampling paths for a particular parameter α_t and the corresponding variance parameter $\exp(h_t)$. Obviously, the mixing is quite satisfactory. Similar sampling paths have been obtained for the other parameters.

5 Conclusions

The results in Section 4 provide strong empirical evidence that locally adaptive dynamic models are a promising and conceptually simple approach for nonparametric estimation of unsmooth curves. In particular, the results for unsmooth and highly oscillating curves, such as the Doppler, are very encouraging.

In applications, a decision has to be made about the order of random walk priors. For a subjective choice, we recommend to use a RW(2) for regression functions with possibly high curvature but without jumps. For regression functions with jumps a RW(1) prior should be considered as an alternative. For the variance function, a RW(1) model is usually a good choice. Obviously, some data driven method for model choice, in particular giving support for deciding about the types of random walks, would be helpful. The recently proposed DIC criterion (Spiegelhalter et al., 2001) is a rather general tool in connection with MCMC techniques, which we routinely compute from

the MCMC output. It is a generalization of the classical AIC criterion for complex hierarchical models. More practical experience is needed, however, before DIC can be used as a standard tool for Bayesian model comparison. In particular, reasonable estimates of the MC error of DIC are needed. Therefore, we currently consider DIC values as a guideline which has to be used in connection with subjective priors.

Apart from the extensions already mentioned at the end of Section 2, the following generalizations could offer a field for future research: First, the Gaussian observation model (7) can be replaced by non Gaussian observation models. In particular, choice of distribution from the exponential family defines a large class of locally adaptive modifications of standard dynamic generalized linear models (e.g. Fahrmeir and Tutz, 2001, Ch.8) or generalized additive mixed models (Fahrmeir and Lang, 2001). A further possibility is the introduction of varying variances in the observation model, as in stochastic volatility models (Taylor, 1986). Another generalization concerns Markov random fields for spatial data analysis. Here local adaption for unsmooth surfaces could be achieved by introducing unknown weights or scale factors in pairwise difference priors (see e.g. Besag, Green, Higdon and Mengersen, 1995, Section 3), together with spatial smoothness priors for them, and estimating these weights simultaneously with the surfaces. Aykroyd (1998) follows this idea, additional research on performance and applicability is necessary, however.

Acknowledgements: This research has been supported by grants from the German National Science Foundation, Sonderforschungsbereich 386. We thank the referee for valuable comments and suggestions which led to substantial improvements of the first version of this paper.

References

Abramovich, F. and Steinberg, D. M. (1996), Improved inference an nonparametric regression using L_k-smoothing splines. Journal of Statistical Planning and Inference, 49, 327-341.

Aykroyd, R. (1998), Bayesian estimation for homogeneous and inhomogeneous Gaussian random fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20, 533-539.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995), Bayesian computation and stochastic systems, Statist. Sci., 10, 3-66.

Biller, C. (2000), Adaptive Bayesian regression splines in semiparametric generalized linear models. J. Comp. and Graph. Stat., 12, 122-140.

Carter, C. K. and Kohn, R. (1994), On Gibbs sampling for state space models. Biometrica, 81, 541-553.

Carter, C. K. and Kohn, R. (1996a), Markov chain Monte Carlo in conditionally Gaussian state space models. Biometrika 83, 589-601.

Carter, C.K. and Kohn, R. (1996b), Robust Bayesian nonparametric regression. In Statistical Theory and Computational Aspects of Smoothing, eds. Hrdle, W. and Schimek, M.G., Physica-Verlag, Heidelberg, 1996.

Denison, D. G. T., Mallick, B. K. and Smith, A. F. M. (1998), Automatic Bayesian curve fitting. J. R. Statist. Soc. B, 60, 333–350.

Donoho, D. L. and Johnstone, I. M. (1994), Ideal spatial adaption by wavelet shrinkage. Biometrika, 81, 425-455.

Fahrmeir, L. and Künstler, R. (1999), Penalized likelihood smoothing in robust state space models. Metrika 49, 173-191.

Fahrmeir, L. and Lang, S. (2001), Bayesian inference for generalized additive mixed models based on Markov random field priors. Appl. Statist. (JRSS C), 50, 201-220.

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

Fan, J. and Gijbels, I. (1995), Data–driven bandwidth selection in local polynomial filting: variable bandwidth and spatial adaption. J. R. Statist. Soc. B 57, 371-394.

Fruehwirth-Schnatter, S. (1994), Data augmention and dynamic linear models. Journal of Time Series Analysis, 15, 183-202.

Harvey, A. C. (1989), Forecasting, structural time series models and the Kalman filter, Cambridge University Press, Cambridge.

Knorr–Held, L.(1999), Conditional prior proposals in dynamic models. Scand. J. Statist., 26, 129-144.

Kohn, R. and Ansley, C. F. (1987), A new algorithm for spline smoothing based on smoothing a stochastic process. SIAM J. of Sci. and Statist. Comp., 8, 33-48.

Lang, S. and Brezger, A. (2001), Bayesian P-splines. SFB 386 Discussion Paper 336, University of Munich.

Rue, H. (2001), Fast sampling of Gaussian Markov random fields. J. R. Statist. Soc., 63, 325-338.

Ruppert, D. and Carroll, R.J. (2000), Spatially adaptive penalties for spline fitting. Australian and New Zealand Journal of Statistics, 42, 205-223.

Seifert, B. and Gasser T. (2000), Data adaptive ridging in local polynomial regression. J. Comp. and Graph. Stat. 9, 338-360.

Shephard, N., and Pitt, P. (1997), Likelihood analysis on non-Gaussian measurement time series. Biometrika 84, 653-667.

Smith, M. and Kohn, R. (1996), Nonparametric regression using Bayesian variable selection. Journal of Econometrics, 75, 317-343.

Smith, M. and Kohn, R., (1997), A Bayesian approach to nonparametric bivariate regression. J. Am. Statist. Ass., 92, 1522-1535.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde (2001), Bayesian deviance, the effective number of parameters, and the comparison of arbitrary complex models. Preprint.

Stone, C. J., Hansen, M. and Kooperberg, C. and Troung (1997), Polynomial splines and their tensor products in extended linear modeling (with discussion) Ann. Statist., 25, 1371-1470.

Taylor, St. J. (1986), Modelling financial time series. John Wiley, Chichester.

Whittaker, E. T. (1923), On a new method of graduation. Proceedings of Edinburgh Mathmatical Society, 41, 63-75.

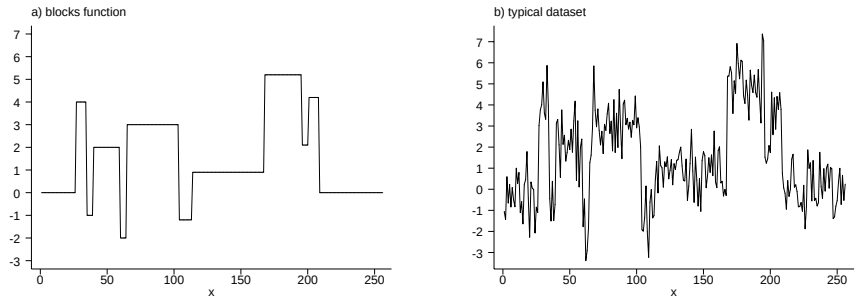

Figure 2: *Blocks function: Figure a) displays the true function and Figure b) a typical replication.*

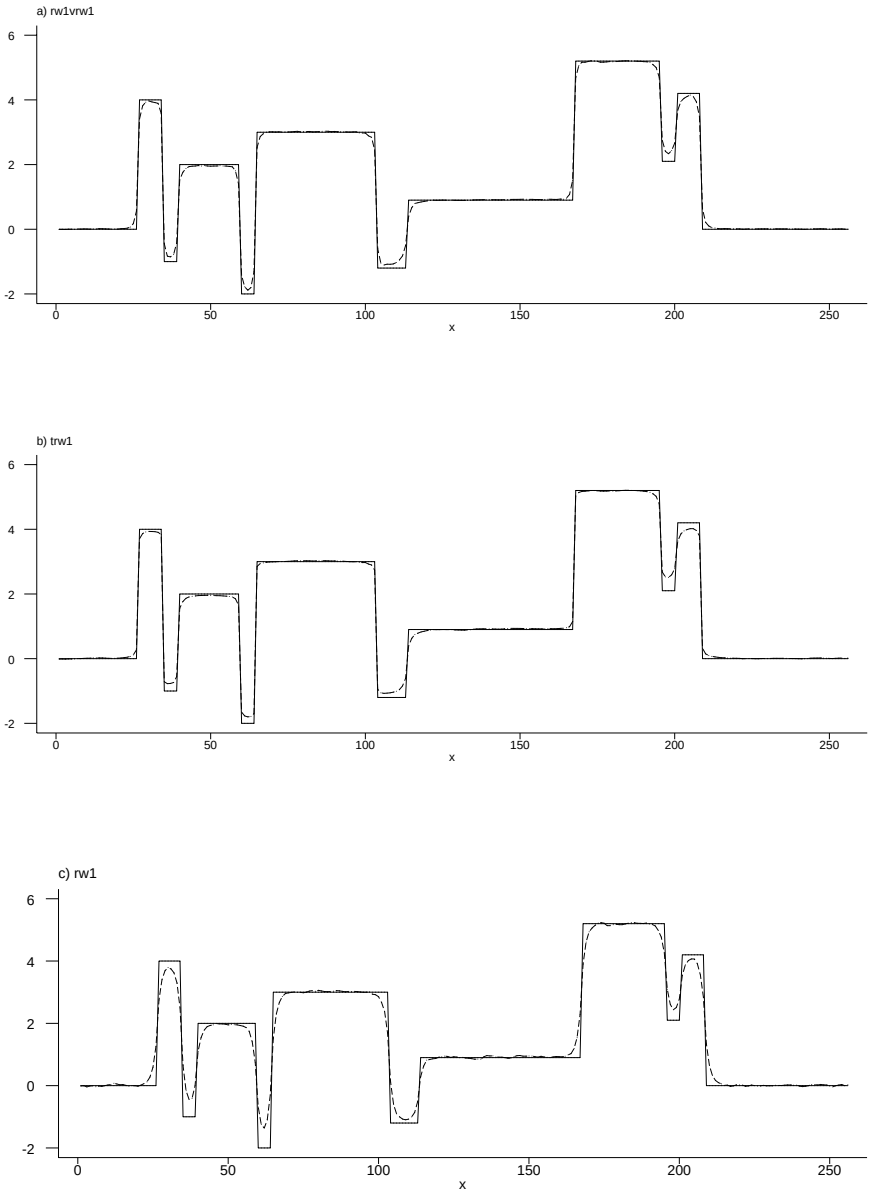

Figure 3: *Blocks function: average estimates (dahsed lines) and true function (solid lines).*

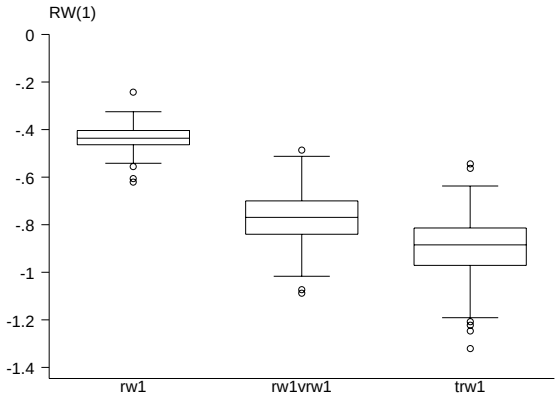

Figure 4: *Blocks function: boxplots of* $log_{10}(MSE)$ *for the three estimators.*

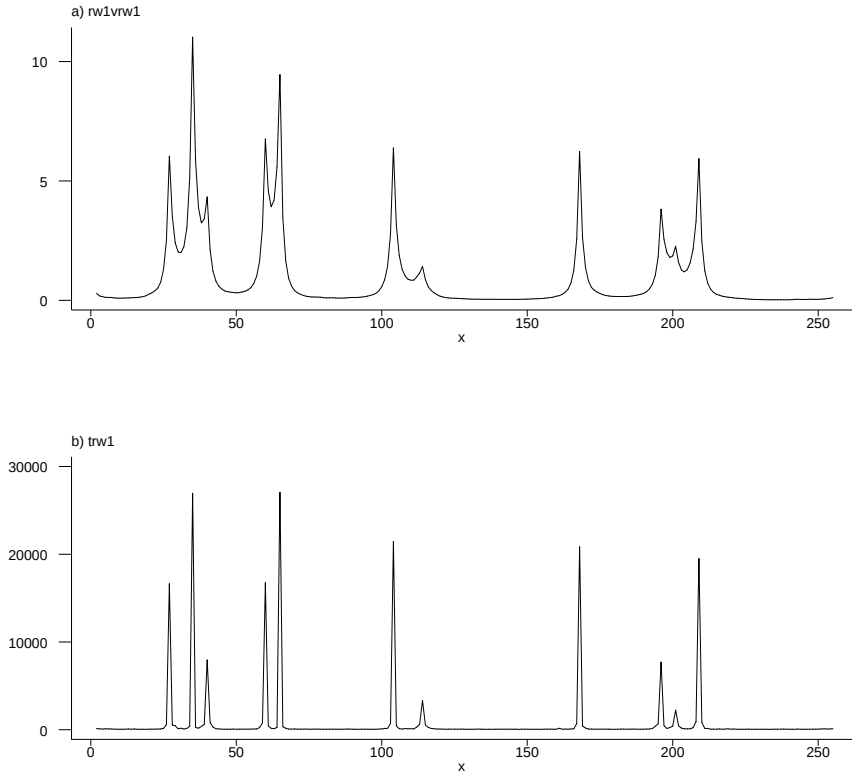

Figure 5: *Block function: average estimates of the variance or weights for rw1vrw1 and trw1, respectively.*

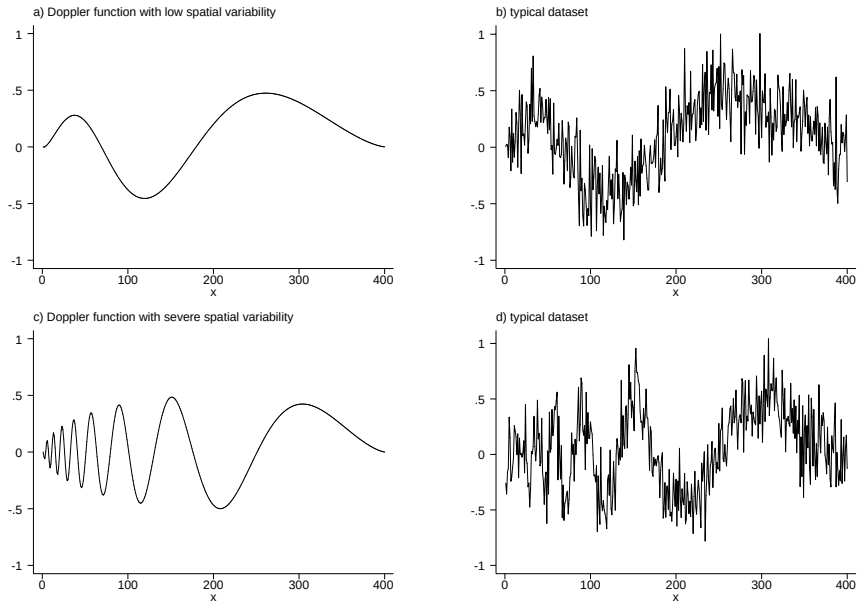

Figure 6: *Doppler functions: Panels a) and c) display the true functions and panels b) and d) typical replications.*

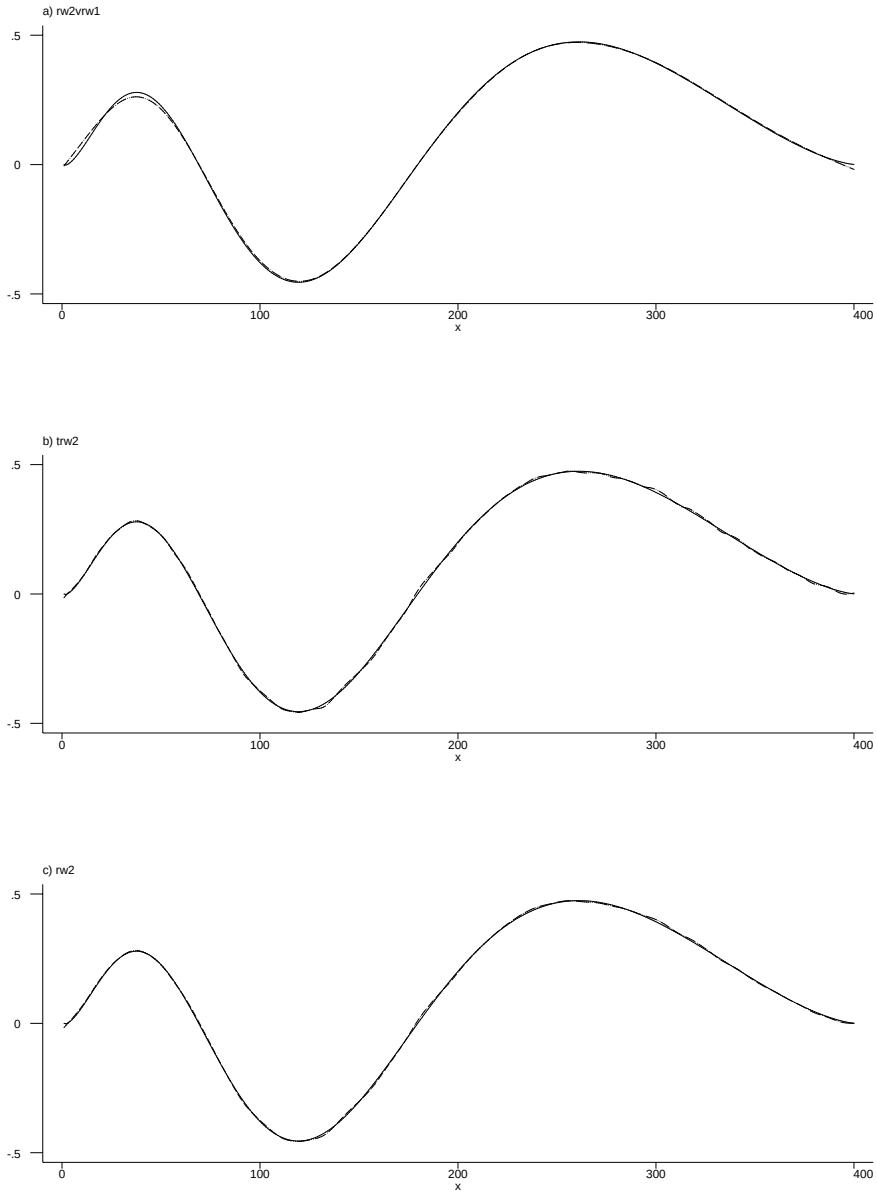

Figure 7: *Doppler function with low spatial variability: average estimates (dashed lines) and true function (solid lines).*

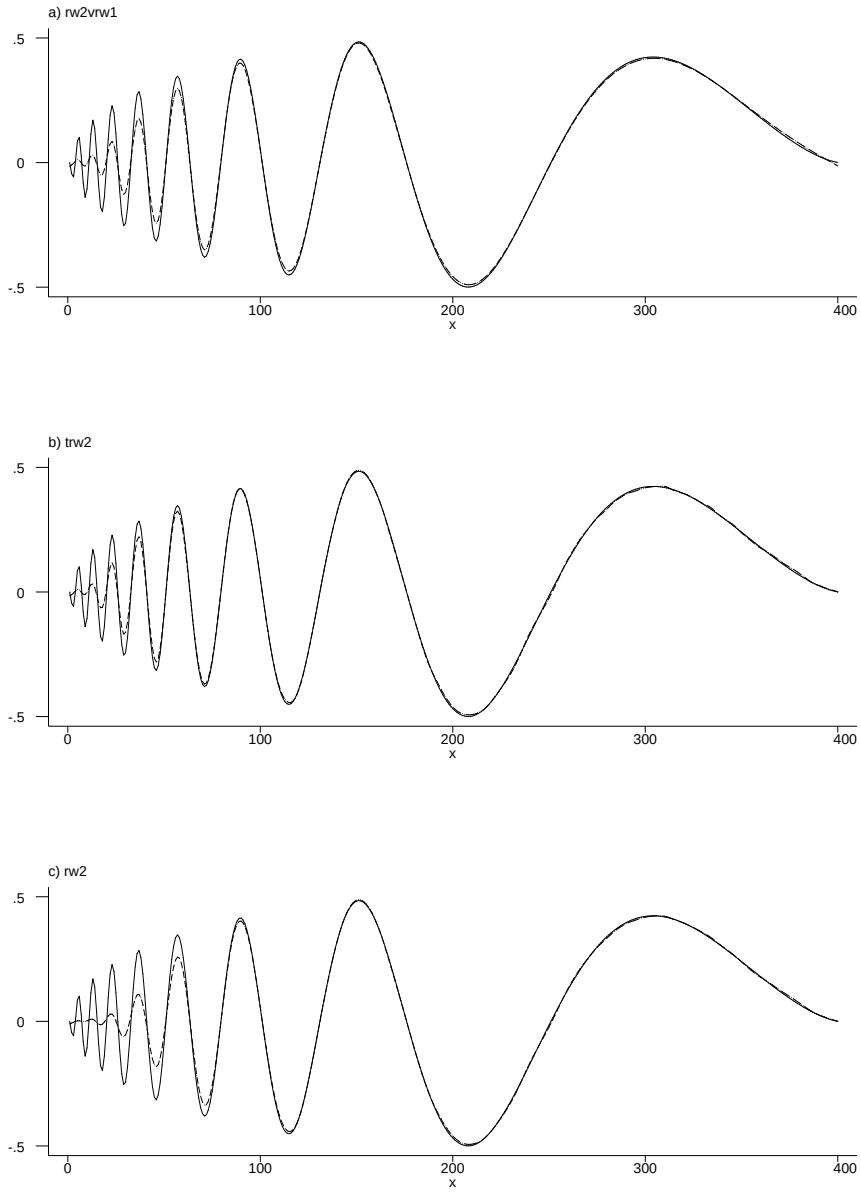

Figure 8: *Doppler function with severe spatial variability: average estimates (dashed lines) and true function (solid lines).*

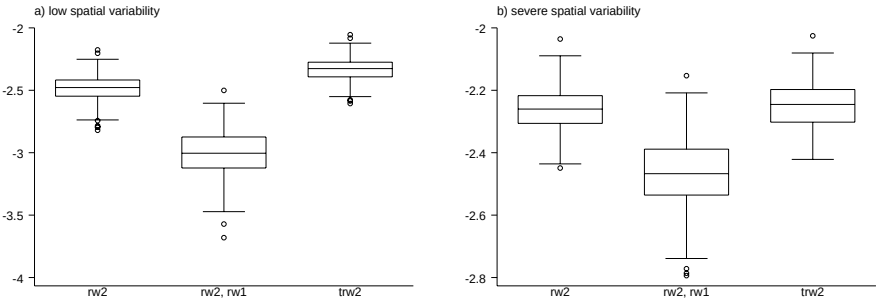

Figure 9: *Doppler function: boxplots of* $log_{10}(MSE)$ *for the three estimators.*

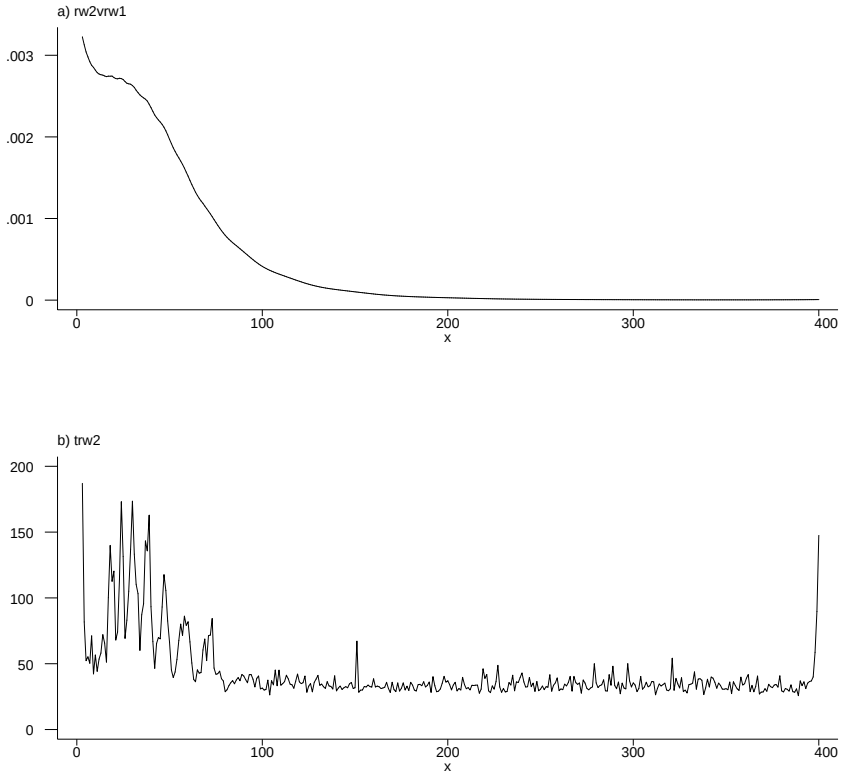

Figure 10: *Doppler function with severe spatial variability: average estimates of the variance or weights for rw2vrw1 and trw2, respectively.*

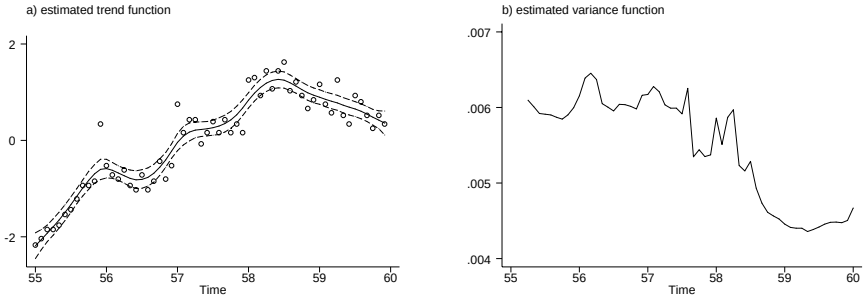

Figure 11: *CP6 sales data: Figure a) shows the posterior mean together with pointwise 80 % credible intervals for* α*. The observations are marked by circles. Figure b) displays the posterior mean estimates for the locally adaptive variances.*

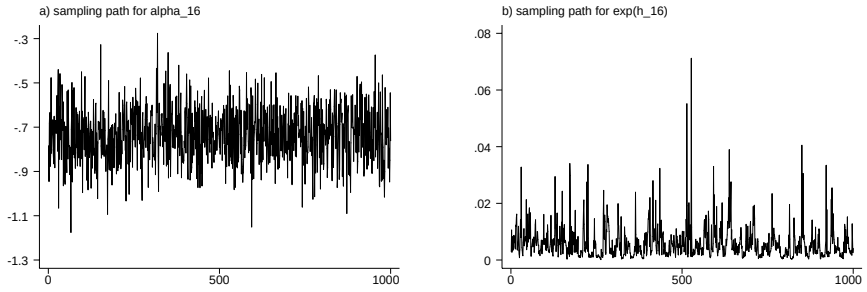

Figure 12: *CP6 sales data: Figure a) shows the sampling path for the parameter* α_{16} *and Figure b) for the variance parameter* $\exp(h_{16})$*.*