MAP-BASED ESTIMATION OF THE PARAMETERS OF NON-STATIONARY GAUSSIAN PROCESSES FROM NOISY OBSERVATIONS

Alexander Krueger, Reinhold Haeb-Umbach

Department of Communications Engineering, University of Paderborn,
33098 Paderborn, Germany

ABSTRACT

The paper proposes a modification of the standard maximum a posteriori (MAP) method for the estimation of the parameters of a Gaussian process for cases where the process is superposed by additive Gaussian observation errors of known variance. Simulations on artificially generated data demonstrate the superiority of the proposed method. While reducing to the ordinary MAP approach in the absence of observation noise, the improvement becomes the more pronounced the larger the variance of the observation noise. The method is further extended to track the parameters in case of non-stationary Gaussian processes.

Index Terms— MAP parameter estimation, noisy observations

1. INTRODUCTION

MAP estimation of mean and variance of a stationary Gaussian process is probably one of the most popular textbook examples of Bayesian parameter estimation [1]. In practice, however, the Gaussian process is often not directly observable, but superposed by additive noise, caused e.g. by measurement or estimation errors. The problem occurs for example in the enhancement of noisy speech, where the enhancement algorithms require an estimate of the parameters of the noise process. These in turn have to be estimated from the noisy speech, i.e. the target process (here: acoustical environmental noise) whose parameters are to be estimated, is not directly observable, but only noisy observations, in this case observations corrupted by speech. Another example are tracking problems in the presence of unknown or time-variant dynamical model parameters [2, 3]. Here, the model parameters have to be estimated alongside the process state vector, both from noisy observations. While the expectation maximization algorithm is a general approach to compute maximum likelihood estimates of the hidden process parameters [4], in speech enhancement minimum statistics based methods [5] may be employed for the estimation of the noise power spectral density.

Here, we present a new general estimation concept taking the Bayesian perspective, which might e.g. be applied for the solution of the above-mentioned problems. Besides taking advantage of a priori knowledge about the parameters gained from previous observations, it exploits two usually valid fundamental assumptions: the uncorrelatedness of the desired process and the superposed noise and the availability of the variance of the noise process. While the latter might not be explicitly available in many applications, it is often possible to compute an estimate thereof.

The paper is organized as follows. In Section 2 we present the idea of MAP based parameter estimation, where we recapitulate the conventional MAP approach in Section 2.1, present the modifications for noisy observations in Section 2.2 and explain the extension for non-stationary processes in Section 2.3. Section 3 provides simulation results and the paper is concluded by Section 4.

2. MAP-BASED PARAMETER ESTIMATION FROM NOISY OBSERVATIONS

Assume that we are given a real-valued stationary white Gaussian stochastic process $\{V_m\}_{m\in\mathbb{N}}$, where the random variable V_m has the following Gaussian probability density function (pdf)

$$p_{V_m}(v_m) = \mathcal{N}\left(v_m; \mu_V, \sigma_V^2\right) \quad \text{for } m \in \mathbb{N}, \tag{1}$$

with μ_V and σ_V^2 denoting the mean and variance of V_m, respectively.

Assuming that at some time instant m we are given some estimates $\hat{\mu}_{V,m}$ and $\hat{\sigma}^2_{V,m}$ of the pdf parameters, we are interested in improving the estimates based on the noisy observation

$$\hat{v}_{m+1} := v_{m+1} + e_{m+1}. \tag{2}$$

Here, v_{m+1} is a realization of V_{m+1} and e_{m+1} denotes a realization of a zero-mean Gaussian observation error E_{m+1}, which is uncorrelated to V_{m+1}. The pdf of the error is assumed to be given by

$$p_{E_{m+1}}(e_{m+1}) = \mathcal{N}\left(e_{m+1}; 0, \sigma^2_{E_{m+1}}\right), \tag{3}$$

employing a time-variant known variance $\sigma^2_{E_{m+1}}$.

Our approach is motivated by the standard MAP parameter estimation, which, however, assumes the observations to be free of errors. We give a short review of this estimation procedure, before we address the necessary modifications to consider observation errors.

2.1. Error-Free Observations

In the case of an error-free observation $\hat{v}_{m+1} = v_{m+1}$, the improved parameters $\hat{\mu}_{V,m+1}$ and $\hat{\sigma}^2_{V,m+1}$ are obtained according to the standard MAP parameter estimation approach by maximizing the joint posterior pdf of μ_V and σ_V^2 given the new observation v_{m+1}

$$\begin{aligned} &p^{(m+1)}_{\mu_V,\sigma_V^2|V_{m+1}}(\mu,\sigma^2|v_{m+1}) \\ &\propto p^{(m)}_{\mu_V,\sigma_V^2}(\mu,\sigma^2) \cdot p^{(m)}_{V_{m+1}|\mu_V,\sigma_V^2}\left(v_{m+1}|\mu,\sigma^2\right), \end{aligned} \tag{4}$$

which is according to Bayes' rule proportional to the product of the joint prior parameter pdf $p^{(m)}_{\mu_V,\sigma_V^2}(\mu,\sigma^2)$ and the Gaussian observation likelihood

$$p^{(m)}_{V_{m+1}|\mu_V,\sigma_V^2}\left(v_{m+1}|\mu,\sigma^2\right) \propto \sigma^{-1}\exp\left(-\frac{(v_{m+1}-\mu)^2}{2\sigma^2}\right). \tag{5}$$

In order to be able to apply the estimation procedure recursively for subsequent time instants, the posterior pdf must be in the same family as the prior pdf. Thus, the prior pdf is chosen to be a conjugate prior for the likelihood (5), which may be written as the product [1]

$$p^{(m)}_{\mu_V,\sigma^2_V}(\mu,\sigma^2) = p^{(m)}_{\mu_V|\sigma^2_V}(\mu|\sigma^2)\cdot p^{(m)}_{\sigma^2_V}(\sigma^2) \qquad (6)$$

of the conditional pdf of μ_V given σ^2_V

$$p^{(m)}_{\mu_V|\sigma^2_V}(\mu|\sigma^2) = \mathcal{N}\left(\mu;\hat{\mu}_{V,m},\frac{\sigma^2}{\kappa_m}\right) \qquad (7)$$

$$\propto (\sigma)^{-1}\exp\left(-\frac{\kappa_m\left(\mu-\hat{\mu}_{V,m}\right)^2}{2\sigma^2}\right) \qquad (8)$$

and the marginal pdf of σ^2_V

$$p^{(m)}_{\sigma^2_V}(\sigma^2) \propto \left(\sigma^2\right)^{-\frac{\nu_m}{2}-1}\exp\left(-\frac{\nu_m\lambda^2_{V,m}}{2\sigma^2}\right), \qquad (9)$$

which is a scaled inverse chi-square distribution. The four hyper parameters $\hat{\mu}_{V,m}$, $\kappa_m > 0$, $\nu_m > 0$ and $\lambda^2_{V,m}$ may be interpreted as the location and scale of μ_V as well as the degrees of freedom and scale of σ^2_V, respectively. Typically, $\nu_m = \kappa_m$ represents the number of observations underlying the distribution.

Using (8) and (9) the prior distribution (6) may be formulated as

$$p^{(m)}_{\mu_V,\sigma^2_V}(\mu,\sigma^2) \propto \left(\sigma^2\right)^{-\frac{(\nu_m+3)}{2}}\exp\left(-\frac{\nu_m\lambda^2_{V,m}+\kappa_m\left(\mu-\hat{\mu}_{V,m}\right)^2}{2\sigma^2}\right). \qquad (10)$$

Since its mode satisfies

$$\left(\hat{\mu}_{V,m},\frac{\nu_m}{\nu_m+3}\lambda^2_{V,m}\right) = \underset{\mu,\sigma^2}{\operatorname{argmax}}\, p^{(m)}_{\mu_V,\sigma^2_V}(\mu,\sigma^2), \qquad (11)$$

given the initial estimate $\hat{\sigma}^2_m$, the scale $\lambda^2_{V,m}$ is chosen such that $\hat{\sigma}^2_m$ maximizes the prior:

$$\lambda^2_{V,m} = \frac{\nu_m+3}{\nu_m}\hat{\sigma}^2_{V,m}. \qquad (12)$$

Using the prior pdf (10) and the likelihood (5), it can be shown after some basic manipulations [1] that the posterior pdf (4) has the same form as the prior pdf (10), where the four parameters κ_m, ν_m, $\hat{\mu}_{V,m}$ and $\lambda^2_{V,m}$ have to be correspondingly replaced by

$$\kappa_{m+1} = \kappa_m + 1, \quad \nu_{m+1} = \nu_m + 1 \qquad (13)$$

$$\hat{\mu}_{V,m+1} = \hat{\mu}_{V,m} + \frac{1}{\kappa_m+1}\left(v_{m+1}-\hat{\mu}_{V,m}\right) \qquad (14)$$

$$\lambda^2_{V,m+1} = \frac{1}{\nu_m+1}\left(\nu_m\lambda^2_{V,m} + \frac{\kappa_m}{\kappa_m+1}\left(v_{m+1}-\hat{\mu}_{V,m}\right)^2\right). \qquad (15)$$

The posterior pdf (4) is, according to (11), maximized by $\hat{\mu}_{V,m+1}$ and $\hat{\sigma}^2_{V,m+1} = \frac{\nu_{m+1}}{\nu_{m+1}+3}\lambda^2_{V,m+1}$.

For the following parameter estimation employing the observation v_{m+2} this posterior pdf is interpreted as new priori pdf $p^{(m+1)}_{\mu_V,\sigma^2_V}(\mu,\sigma^2)$ allowing a recursive processing.

2.2. Noisy Observations

In the case of a noisy observation $\hat{v}_{m+1}$ the observation likelihood changes to

$$p^{(m)}_{\hat{V}_{m+1}|\mu_n,\sigma^2_n}\left(\hat{v}_{m+1}|\mu,\sigma^2\right) \propto \left(\sigma^2+\sigma^2_{E_{m+1}}\right)^{-\frac{1}{2}}\exp\left(-\frac{\left(\hat{v}_{m+1}-\mu\right)^2}{2\left(\sigma^2+\sigma^2_{E_{m+1}}\right)}\right). \qquad (16)$$

Since the prior pdf (6) is not a conjugate prior for the likelihood (16), the posterior pdf has a different form than the prior pdf and can be shown to satisfy

$$p^{(m+1)}_{\mu_V,\sigma^2_V|\hat{V}_{m+1}}(\mu,\sigma^2|\hat{v}_{m+1}) \propto \left(\sigma^2+\sigma^2_{E_{m+1}}\right)^{-\frac{1}{2}}\left(\sigma^2\right)^{-\frac{(\nu_m+3)}{2}} \cdot\exp\left(-\frac{\tilde{\kappa}_{m+1}\left(\sigma^2\right)\left[\mu-\tilde{\mu}_{V,m+1}\left(\sigma^2\right)\right]^2}{2\sigma^2}\right) \cdot\exp\left(-\frac{\nu_m\lambda^2_{V,m}+\frac{\kappa_m\sigma^2\left(\hat{v}_{m+1}-\hat{\mu}_{V,m}\right)^2}{(\kappa_m+1)\sigma^2+\kappa_m\sigma^2_{E_{m+1}}}}{2\sigma^2}\right), \qquad (17)$$

where

$$\tilde{\kappa}_{m+1}\left(\sigma^2\right) := \kappa_m + \frac{\sigma^2}{\sigma^2+\hat{\sigma}^2_{E_{m+1}}} \qquad (18)$$

$$\tilde{\mu}_{V,m+1}\left(\sigma^2\right) := \hat{\mu}_{V,m} + \frac{\sigma^2\left(\hat{v}_{m+1}-\hat{\mu}_{V,m}\right)}{\left(\kappa_m+1\right)\sigma^2+\kappa_m\sigma^2_{E_{m+1}}} \qquad (19)$$

resemble (13) and (14), but now are functions of σ^2. Due to its much more complex form compared to the prior pdf (10), the search of the mode of the posterior pdf (17) cannot be adopted from the error-free case in a straight forward manner.

We propose to search the mode by first approximating (18) and (19) by constants as follows

$$\kappa_{m+1} := \tilde{\kappa}_{m+1}\left(\hat{\sigma}^2_{V,m}\right) \qquad (20)$$

$$\hat{\mu}_{V,m+1} := \tilde{\mu}_{V,m+1}\left(\hat{\sigma}^2_{V,m}\right). \qquad (21)$$

This approximation is motivated by the fact that both quantities (18) and (19) are bounded by

$$\kappa_m < \tilde{\kappa}_{m+1}\left(\sigma^2\right) < \kappa_m + 1 \qquad (22)$$

$$\hat{\mu}_{V,m} < \tilde{\mu}_{V,m+1}\left(\sigma^2\right) < \hat{\mu}_{V,m} + \frac{1}{\left(\kappa_m+1\right)}\left(v_m-\hat{\mu}_{V,m}\right), \qquad (23)$$

where the lower and upper bounds are reached for $\sigma^2 \to 0$ and $\sigma^2 \to \infty$, respectively. Since the mode of the prior with respect to σ^2 is located at $\hat{\sigma}^2_{V,m}$, it is reasonable to take this value as the prediction for the value of the mode of the posterior pdf (17) and evaluate (18) and (19) at this location.

Using the approximations (20) and (21) in (17) yields an approximated posterior, which we denote by $\hat{p}^{(m+1)}_{\mu_V,\sigma^2_V|\hat{V}_{m+1}}$ and for which

$$\underset{\mu}{\operatorname{argmax}}\,\hat{p}^{(m+1)}_{\mu_V,\sigma^2_V|\hat{V}_{m+1}}(\mu,\sigma^2|\hat{v}_{m+1}) = \hat{\mu}_{V,m+1} \quad \forall\sigma^2\in\mathbb{R}^+ \qquad (24)$$

holds. Thus, the search for its mode is equivalent to the search of the minimum of the following function

$$f:\mathbb{R}^+\to\mathbb{R},$$

$$f(\psi) := -2\ln\left(\hat{p}^{(m+1)}_{\mu_V,\sigma^2_V|\hat{V}_{m+1}}\left(\hat{\mu}_{V,m+1},\psi|\hat{v}_{m+1}\right)\right) \qquad (25)$$

$$= (K_5+3)\ln(\psi) + \ln(K_0+\psi) + \frac{K_1}{\psi} + \frac{K_2}{K_3\psi+K_4} \qquad (26)$$

with constants defined as

$$K_0 := \sigma^2_{E_{m+1}},\ \ K_1 := \nu_m \lambda^2_{V,m},\ \ K_2 := \kappa_m \left(\hat{v}_{m+1} - \hat{\mu}_{V,m}\right)^2, \tag{27}$$

$$K_3 := 2\left(\kappa_m + 1\right)\ \ K_4 := 2\kappa_m \sigma^2_{E_{m+1}},\ \ K_5 := \nu_m. \tag{28}$$

Since f is continuous and a sum of a strictly monotonically increasing and a strictly monotonically decreasing function, for which

$$\lim_{\psi \to 0} f\left(\psi\right) = \lim_{\psi \to \infty} f\left(\psi\right) = \infty \tag{29}$$

holds, it can be shown that it has exactly one minimum, which is a local one. For that reason searching the minimum of f can be equivalently formulated as searching the (only positive) root of its derivative f', which is obtained by applying some basic derivation rules as

$$f'(\psi) = \frac{1}{\psi^2\left(K_0 + \psi\right)\left(K_3\psi + K_4\right)^2} \cdot g\left(\psi\right) \tag{30}$$

with $g\left(\psi\right)$ defined by

$$\begin{aligned} g\left(\psi\right) := &\left[\left(K_5 + 3\right)\psi - K_1\right]\left(K_0 + \psi\right)\left(K_3\psi + K_4\right)^2 \\ &+ \psi^2\left[\left(K_3\psi + K_4\right)^2 - 2K_2K_3\left(K_0 + \psi\right)\right]. \end{aligned} \tag{31}$$

As the first factor of f' is positive for positive ψ, it is sufficient to find the positive root of g, which in principle may be determined analytically, since g is a fourth order polynomial. However, for reasons of simplicity we propose a numerical approach instead. It can be verified that $g(0) < 0$ and $g\left(b_{\mathrm{U}}\right) > 0$ with

$$b_{\mathrm{U}} := \max\left(\frac{K_1}{K_5 + 3}, \frac{2K_2 + K_3K_0 - K_4}{K_3}\right). \tag{32}$$

Due to the continuity of g, its single positive root must be located within the interval $\left[0, b_{\mathrm{U}}\right]$ and may thus be effectively computed by using a combination of a bisection and Newton approach.

Starting from an initial approximate solution ψ_0 obtained by the bisection method, according to the Newton method improved solutions ψ_j, $j = 1, 2, ...$ may be iteratively computed by

$$\psi_j = \psi_{j-1} - \frac{g\left(\psi_{j-1}\right)}{g'\left(\psi_{j-1}\right)}. \tag{33}$$

We propose to stop the iterations, if the relative difference between the j-th and $(j-1)$-th solution

$$\delta_j := \frac{\left|\psi_j - \psi_{j-1}\right|}{\left|\psi_{j-1}\right|} \tag{34}$$

falls below a lower threshold δ. The estimate of the mode of the posterior pdf $\hat{p}^{(m+1)}_{\mu_V, \sigma^2_V | \hat{V}_{m+1}}(\mu, \sigma^2 | \hat{v}_{m+1})$ with respect to σ^2 is then computed from

$$\hat{\sigma}^2_{V,m+1} := \psi_J, \tag{35}$$

where J is the smallest index, for which $\delta_J < \delta$ holds.

Having found the mode of the approximate posterior pdf $\hat{p}^{(m+1)}_{\mu_V, \sigma^2_V | \hat{V}_{m+1}}$, the goal is to approximate this pdf by a pdf having the same form as the prior pdf (6). For this purpose, it is reasonable to set

$$\nu_{m+1} = \kappa_{m+1} \tag{36}$$

and to use (6) for the approximation of the posterior pdf, where the parameters κ_m, $\hat{\mu}_{V,m}$, $\hat{\sigma}^2_{V,m}$ and ν_m are replaced by (20), (21), (35) and (36), respectively.

It is important to note that this method simplifies to the standard MAP approach for $\sigma^2_{E_m} = 0$.

2.3. Modifications for Non-Stationary Processes

We now assume the Gaussian stochastic process $\left\{V_m\right\}_{m \in \mathbb{N}}$ to be non-stationary, which can be expressed by substituting the time-invariant mean μ_V and variance σ^2_V in (1) by its time-variant counterparts μ_{V_m} and $\sigma^2_{V_m}$.

In order to estimate these time-variant parameters, we propose a simple modification of the approach proposed in the previous section. The idea is to introduce a forgetting mechanism which keeps only the information obtained by the most recent N observations. This is accomplished by keeping the values of ν_m and κ_m constant at the value of N, while the rest of the estimation process remains unaltered. The choice of N depends on the desired trade-off between estimation accuracy and tracking ability.

3. SIMULATIONS

To verify the advantages of the proposed tracking method over the conventional MAP estimation method, which uses the update equations (13) - (15), we compared both methods in estimating the parameters of stationary as well as non-stationary Gaussian stochastic processes in the presence of noisy observations.

3.1. Stationary Case

The apparent difference between the two methods becomes clear in Figure 1, which shows the realization of a noisy observation process $\left\{\hat{V}_m\right\}_{1 \le m \le M}$ with $M = 10000$ and $\mu_V = \sigma^2_V = 1$ as well as the corresponding mean and variance estimates. The variance of the observation error $\sigma^2_{E_m}$ was randomly drawn from a uniform distribution on $\left[0, 16\sin^2\left(\frac{2\pi m}{M}\right)\right]$.

It can be observed that the conventional MAP approach tries to estimate the mean μ_V and variance $\sigma^2_V + \sigma^2_{E_m}$ of the noisy observed process $\hat{V}_m$ rather than of V_m, resulting in more fluctuating mean estimates and overestimated variance estimates compared to the proposed method. Note, in particular, the different scaling of the graphs showing the variance estimates in Figure 1c. It should be emphasized that correcting the variance estimates $\hat{\sigma}^{2(\mathrm{MAP})}_{V,m}$ obtained by the conventional MAP method by simply subtracting $\sigma^2_{E_m}$ is not reasonable, since it results in strongly fluctuating estimates, which may even become negative.

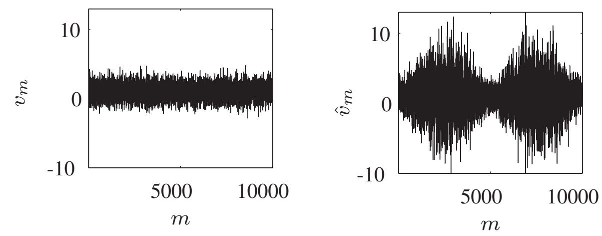
3.2. Non-stationary Case

The proposed method was further applied for the tracking of the time-variant parameters

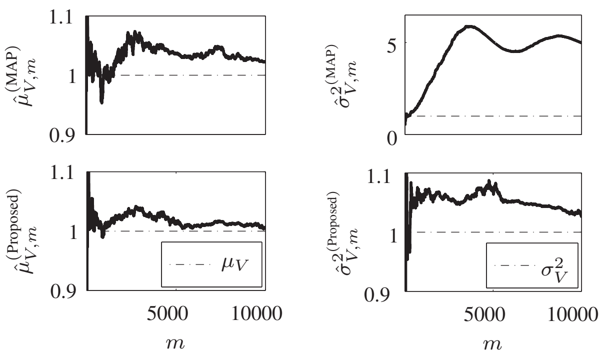
$$\mu_{V_m} = \cos\left(\frac{4\pi m}{M}\right),\ \ \sigma^2_{V_m} = 1 + \sin^2\left(\frac{\pi m}{M}\right),\ \ M = 2000. \tag{37}$$

of a non-stationary Gaussian random process V_m. The variance $\sigma^2_{E_m}$ was randomly drawn from a uniform distribution on $\left[0, c^2\sin^2\left(\frac{2\pi m}{M}\right)\right]$, where c controls the maximum variance of the superposed error E_m. For the proposed method, the values of ν_m and κ_m were kept constant at $N = 10$ to allow an appropriate tracking ability. A corresponding modification was applied to (13) of the conventional MAP approach.

To measure the accuracy of the estimated quantities we computed the root mean square errors RMSE_{μ_V} and $\mathrm{RMSE}_{\sigma^2_V}$ of the estimated means and variances, respectively, averaged over 200 experiments, which are given in Table 1 for different values of c. It



(a) Realizations of clean and noisy random processes



(b) Mean estimation (c) Variance estimation

Fig. 1: *Exemplary performance of conventional MAP and proposed approach on a stationary white Gaussian random process.*

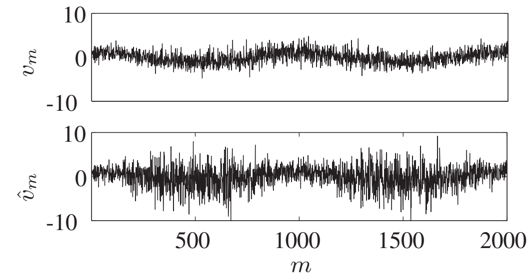
Table 1: *Root means square errors of estimated parameters*

c	0.5	1	2	3	4	6
$\text{RMSE}_{\mu_V}^{(\text{Proposed})}$	0.27	0.27	0.28	0.29	0.30	0.35
$\text{RMSE}_{\mu_V}^{(\text{MAP})}$	0.28	0.29	0.35	0.42	0.51	0.71
$\text{RMSE}_{\sigma_V^2}^{(\text{Proposed})}$	0.41	0.41	0.45	0.48	0.53	0.62
$\text{RMSE}_{\sigma_V^2}^{(\text{MAP})}$	0.52	0.49	0.88	1.99	3.65	8.40

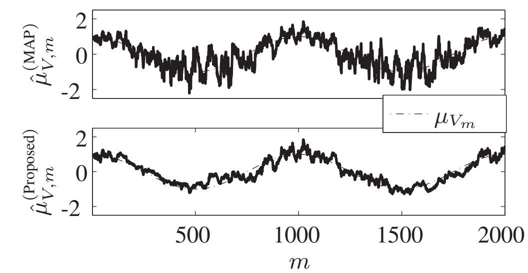
can be seen that in all cases the proposed method outperforms the conventional MAP procedure, where the superiority becomes more apparent for increasing values of c. This can be attributed to the fact that, as in the stationary case, the conventional MAP method estimates the parameters of $\hat{V}_m$ rather than that of V_m. Figure 2 shows the increased fluctuation of the mean estimates and overestimation of the variance obtained by the conventional MAP method compared to the proposed approach for a realization of V_m with $c = 4$. Note, once again, the different scaling used for the graphs in Figure 2c.

4. CONCLUSION

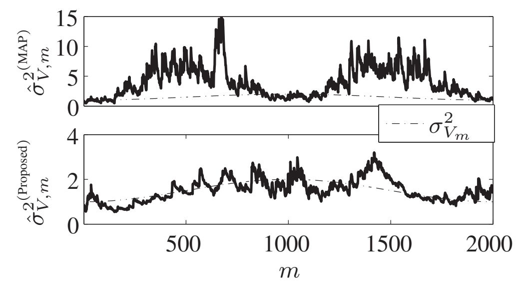
We have presented a modified MAP approach for estimating the parameters of non-stationary Gaussian random processes from noisy observations with zero mean Gaussian errors with known variance. In contrast to the conventional MAP method, the proposed technique allows to estimate the parameters of the concealed desired process by explicitly considering observation errors. It was exemplarily shown by simulations that the superiority of the presented method to conventional MAP increases with the variance of the observation errors.



(a) Realizations of clean and noisy random processes



(b) Mean estimation



(c) Variance estimation

Fig. 2: *Exemplary performance of conventional MAP and proposed approach on a non-stationary Gaussian random process.*

5. REFERENCES

[1] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis, Second Edition*, Chapman & Hall, July 2003.

[2] J. Droppo and A. Acero, "Noise robust speech recognition with a switching linear dynamic model," in *ICASSP*, 2004, vol. 1, pp. I–953–6 vol.1.

[3] A. Krueger and R. Haeb-Umbach, "A model based approach to joint compensation of noise and reverberation for speech recognition," in *Robust Speech Recognition of Uncertain or Missing Data*, chapter 10. Springer, to be published in 2011.

[4] L. Deng, J. Droppo, and A. Acero, "Recursive estimation of nonstationary noise using iterative stochastic approximation for robust speech recognition," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 568 – 580, nov. 2003.

[5] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, 2001.