

RESEARCH

Open Access



# Robust adaptive filtering using recursive weighted least squares with combined scale and variable forgetting factors

Branko Kovačević<sup>1</sup>, Zoran Banjac<sup>2\*</sup> and Ivana Kostić Kovačević<sup>3</sup>

## Abstract

In this paper, a new adaptive robustified filter algorithm of recursive weighted least squares with combined scale and variable forgetting factors for time-varying parameters estimation in non-stationary and impulsive noise environments has been proposed. To reduce the effect of impulsive noise, whether this situation is stationary or not, the proposed adaptive robustified approach extends the concept of approximate maximum likelihood robust estimation, the so-called M robust estimation, to the estimation of both filter parameters and noise variance simultaneously. The application of variable forgetting factor, calculated adaptively with respect to the robustified prediction error criterion, provides the estimation of time-varying filter parameters under a stochastic environment with possible impulsive noise. The feasibility of the proposed approach is analysed in a system identification scenario using finite impulse response (FIR) filter applications.

**Keywords:** Adaptive filtering, FIR filtering, M robust parameter estimation, Recursive weighted least squares, Variable forgetting factor, Recursive variance estimation, Non-stationary noise, Impulsive noise

## 1 Introduction

Adaptive filtering represents a common tool in signal processing and control applications [1–6]. An overview of methods for recursive parameter estimation in adaptive filtering is given in the literature [5–7]. There is, unfortunately, no recursive parameter estimation that is uniformly best. Recursive least squares (RLS) algorithm has been applied commonly in adaptive filtering and system identification, since it has good convergence and provides for small estimation error in stationary situations and under assumption that the underlying noise is normal [5–7]. In this context, however, two problems arise.

First, in the case of time varying parameters, forgetting factor (FF) can be used to generate only a finite memory, in order to track parameter changes [7, 8]. For a value of FF smaller than one, one can estimate the trend of non-stationarity very fast but with higher estimate variance, owing to smaller memory length. On the other hand, with a FF close to unity, the algorithm has wider

memory length and needs rather a relatively long time to estimate the unknown coefficients. However, these coefficients are estimated accurately in stationary situations. Moreover, RLS with fixed value of FF (FFF) is not effective for tracking time-varying parameters with large variations. This makes it necessary to incorporate an adaptive mechanism in the estimator, resulting in the concept of variable FF (VFF). Several adaptation procedures have been discussed by changing the memory length of signal [7–13]. In particular, the methods referred as the parallel adaptation RLS algorithm (PA-RLS) and extended prediction error RLS-based algorithm (EPE-RLS) have a good adaptability in non-stationary situations [9–13]. In addition, both methods assume that the variance of interfering noise is known in advance.

The second problem arises in an application where the required filter output is contaminated by heavy tailed distributed disturbance, generating outliers [14–20]. Namely, the classical estimation algorithms optimise the sum of squared prediction errors (residuals) and, as a consequence, give the same weights to error signals, yielding a RLS type procedure. However, an adequate information about the statistics of additive noise is not

\* Correspondence: zoran.banjac@viser.edu.rs

<sup>2</sup>School of Electrical and Computer Engineering, 283 Vojvode Stepe St., Belgrade, Serbia

Full list of author information is available at the end of the article

included in RLS computation. Possible approaches to robust system identification introduce a non-linear mapping of prediction errors. Although in the statistical literature, there are few approaches to robust parameter estimation, M robust approach (the symbol M means approximate maximum-likelihood) is emphasised, due to its simplicity for practical workers [15–23]. Robustified RLS algorithm, based on M robust principle, the so-called robustified recursive least square method (RRLS), uses the sum of weighted prediction errors as the performance index, where the weights are functions of prediction residuals [24–26]. However, M estimators provide for the solutions of the location parameters estimation problem [21–23]. As a consequence, in a situation of non-stationary noise signal with the time-varying variance, their efficiency should be bad [7, 24]. Therefore, a significant part of RRLS algorithm is the estimation of unknown noise variance or the so-called scale factor [15, 16, 24, 26]. A suitable practical robust solution is the median estimator based on absolute median deviations, named median of absolute median deviations (MAD) estimator [21–23]. However, RRLS algorithm of M robust type using MAD scale factor estimation is also found to be non-effective for tracking of time-varying parameters [7, 11, 24, 26]. For these reasons, neither of the stated algorithms alone can solve the both mentioned problems.

In this article, we design a new robust adaptive finite impulse response (FIR) system for dealing with these problems simultaneously. To alleviate the effects of non-stationary and impulsive noise, this algorithm extends the concept of M robust estimation to adaptive M robust algorithm with the estimation of both filter parameters and unknown noise variance simultaneously. The estimated noise variance, together with the robustified extended prediction error criterion, calculated on the sliding data frame of proper length, is used to define a suitable robust discrimination function, as a normalised measure of signal non-stationarity. In addition, the VFF is introduced by the linear mapping of robust discrimination function. This, in turn, enables the tracking of time-varying filter parameter under the impulsive noise. Simulation results demonstrate the effectiveness of the proposed algorithm, by the comparison with the conventional recursive least squares (RLS) using VFF based on the standard EPE criterion, and the adaptive M robust-based algorithm with only scale factor (RRLS).

## 2 Problem formulation

A commonly used adaptive FIR filtering scenario, expressed as a system identification problem, is presented in Fig. 1. Here,  $x(k)$  is the random input signal,  $d(k)$  is the required filter output,  $n(k)$  is the noise or disturbance and  $e(k)$  is the prediction error or residual. The

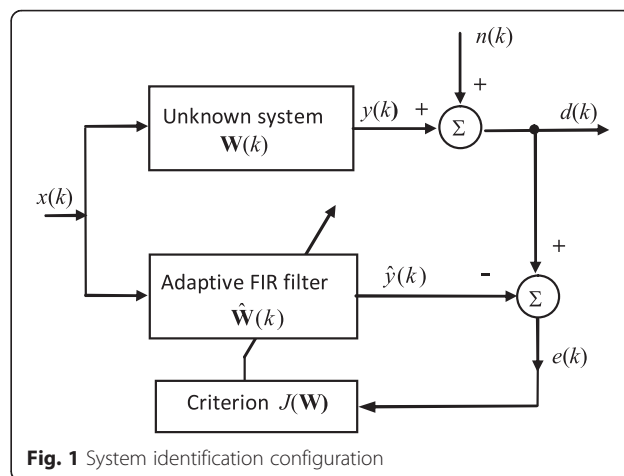


Fig. 1 System identification configuration

filter parameter vector,  $\mathbf{W}(k)$ , can be estimated recursively by optimising the prespecified criterion. The RLS algorithm with FF approaches the problem of estimation of non-stationary (time-varying) signal model parameters by minimising the sum of exponentially weighted squared residuals [5–7, 25, 26]. On the other hand, robust estimates are insensitive to outliers, but are inherently non-linear. Moreover, most robust regression procedures are minimization problems [15–26]. Specifically, M robust estimates are derived as minimization of the sum of weighted residuals, instead of the quadratic performance criterion in the classical RLS computation [21–23]. To combine these two approaches, let us define a new criterion as the sum of weighted residuals

$$J_k(\mathbf{W}, s) = \frac{1}{k} \sum_{i=1}^k \rho^{k-i} \phi\left(\frac{e(i, \mathbf{W})}{s}\right) \tag{1}$$

where the prediction error (residual) signal is given by

$$e(i, \mathbf{W}) = d(i) - \mathbf{X}^T(i) \mathbf{W} \tag{2}$$

with the regression vector,  $\mathbf{X}(i)$ , and the parameter vector,  $\mathbf{W}$ , being defined by

$$\begin{aligned} \mathbf{X}(k) &= [x(k), x(k-1), \dots, x(k-n+1)]^T; \mathbf{W} \\ &= [w_1, w_2, \dots, w_n]^T \end{aligned} \tag{3}$$

Here, the quantities  $s$  and  $\rho$  represent the scale and forgetting factors, respectively. Moreover,  $\mathbf{W}$  is the unknown filter parameter vector that has to be estimated. In addition,  $\phi(\cdot)$  is a robust score, or loss, function, which has to suppress the influence of impulsive noise, generating outliers. Having in mind the importance of reducing the influence of outliers contaminating the Gaussian noise samples,  $\phi(\cdot)$  should be similar to a quadratic function in the middle, but it has to increase more slowly in the tails than the quadratic one. In addition, its first derivative,  $\psi(\cdot) = \phi'(\cdot)$ , the so-called

influence function in the statistical literatures, has to be bounded and continuous [21–23, 25, 26]. The first property provides that single outlier will not have a significant influence, while the second one provides that patchy or grouped outliers will not have a big impact. A possible choice, for example, is the Huber's robust loss function, with the corresponding influence function [21].

$$\psi(x) = \min\left\{\frac{|x|}{\sigma^2}, \frac{\Delta}{\sigma}\right\} \text{sgn}(x) \tag{4}$$

Here,  $\text{sgn}(\cdot)$  is the signum function,  $\sigma$  is the noise standard deviation and  $\Delta$  is a free parameter. This parameter can be adopted in such a way to provide for required efficiency robustness under the zero-mean white normal noise model [21–23]. The non-linear transformation of data based on (4) is known in the statistical literature as winsorization [21–23].

Taking the first partial derivate of (1) with respect to the elements of  $\mathbf{W}$  in (3), say  $\mathbf{W}_j, j = 1, 2, \dots, n$ , being equal to zero, we see that the minimization of (1) reduces to finding the solution of  $n$  non-linear algebraic relations:

$$\sum_{i=1}^k \rho^{k-i} \mathbf{X}_{ij} \psi\left(\frac{e(i, \mathbf{W})}{s}\right) = 0, j = 1, 2, \dots, n \tag{5}$$

where  $\mathbf{X}_{ij}$  is the element in the  $j$ th column of the row vector  $\mathbf{X}^T(i)$  in (3), while  $\psi(\cdot)$  is the first derivative of  $\phi(\cdot), \psi(\cdot) = \phi'(\cdot)$ . Of course, for non-linear  $\psi(\cdot)$ , (5) must be solved by iterative numerical methods, and two suitable procedures are Newton-Raphson's and Ditter's algorithms, respectively, [27, 28]. Here, a slightly different approach is proposed using a weighted least-squares (WLS) approximation of (5). In this approach, the relation (5) is replaced by the following approximation

$$\sum_{i=1}^k \mathbf{X}_{ij} \beta(k, i) e(i, \mathbf{W}) \approx 0, j = 1, 2, \dots, n \tag{6}$$

where the exponentially weighted robust term is given by

$$\beta(k, i) = \rho^{k-i} \omega(i, \mathbf{W}_0) \tag{7}$$

while its robust part is defined by

$$\omega(i, \mathbf{W}_0) = \begin{cases} \frac{\psi\left(\frac{e(i, \mathbf{W}_0)}{s}\right)}{\frac{e(i, \mathbf{W}_0)}{s}} & \text{if } d(i) \neq \mathbf{X}^T(i) \mathbf{W}_0 \text{ and } s \neq 0 \\ 1 & \text{if } d(i) = \mathbf{X}^T(i) \mathbf{W}_0 \text{ and/or } s = 0 \end{cases} \tag{8}$$

Here,  $\mathbf{W}_0$  is an initial estimate of the parameter vector,  $\mathbf{W}$ , which can be obtained, for example, by using the conventional non-recursive LS estimator [5, 25, 26]. The

solution of (6), say  $\hat{\mathbf{W}}(k)$ , represents a one-step non-recursive suboptimal M robust estimate of  $\mathbf{W}$  in (3).

Application of the non-recursive M robust scheme (6)–(8) requires the non-linear residual transformation,  $\psi(\cdot)$ , and scaling factor,  $s$ , to be defined in advance. But, in general, the standard deviation,  $\sigma$ , in (4) is not known beforehand and has to be estimated somehow. A commonly used robust estimate of  $\sigma$  in the statistical literature is the median scheme, based on the absolute median deviations [21–23]

$$s = \frac{\text{median}|e_i - \text{median}(e_i)|}{0.6745}, i = 1, 2, \dots, L \tag{9}$$

where  $L$  denotes the length of sliding data frame. The divisor 0.6745 in (8) is used because the MAD scale factor estimate,  $s$ , is approximately equal to the noise standard deviation,  $\sigma$ , if the sample size,  $L$ , is large and if samples actually arise from a normal distribution [21]. Moreover, because  $s \approx \sigma, \Delta$  is usually taken to be 1.5 [21–23]. This choice will produce much better results, in comparison to the RLS method, when the corresponding noise probability density function (pdf) has heavier tails than the Gaussian one. Furthermore, it will remain good efficiency of RLS when the pdf is exactly normal [21–23].

The weighting term,  $\omega$ , in (8) is not strictly related to the popular robust MAD estimation of the scale factor,  $s$ . This estimate guarantees that  $s \neq 0$ , but in the general case, a scale factor estimator may not guarantee that the estimate,  $s$ , should become equal to zero. This is the reason why the condition  $s = 0$  is included in (8). Particularly, the application of a recursive robustified scale factor estimation requires the initial guess,  $s(0)$ , to be given beforehand. A common choice is  $s(0) = 0$ , but in the first few steps, the obtained estimate of scale factor can be equal to zero, so that the unit value of  $\omega$  in (8) has to be chosen.

The proposed suboptimal M robust estimator (6)–(8) is numerically simpler than the ones oriented towards solving the non-linear optimization problem in (5), but it still remains complex computation. Namely, this method does not have an attractive recursive form and, therefore, is not computationally feasible as the RLS type estimators. Moreover, M robust approach is conservative and may degrade without further adaptation [15–20]. Starting from the proposed non-recursive M robust estimator (6)–(8), a simple and practically applicable recursive M robust parameter estimation procedure with both adaptive robustified scale and variable forgetting factors is derived in the next paragraph. Some alternative approaches for the scale factor adaptation can be found in the literature [15, 16, 24]. Moreover, the application of the EPE-based VFF for solving different practical problems is also discussed in

the literature [12, 17–20, 29, 30]. However, similarly to sample mean and sample variance, the standard EPE approach is non-robust towards outliers [21–23]. Therefore, in the next chapter, an alternative M robust approach for generating VFF adaptively is proposed.

### 3 A new recursive robust parameter estimation algorithm with combined scale and forgetting factors

The solution of (6) can be also represented in the computationally more feasible recursive form, using the well-known algebraic manipulations (for more details, see Appendix 1), [25, 26]. This results in the parameter estimation algorithm

$$\hat{\mathbf{W}}(k) = \hat{\mathbf{W}}(k-1) + \mathbf{K}(k)e(k, \hat{\mathbf{W}}(k-1)) \quad (10)$$

$$e(k, \hat{\mathbf{W}}(k-1)) = d(k) - \mathbf{X}^T(k)\hat{\mathbf{W}}(k-1) \quad (11)$$

$$\begin{aligned} \mathbf{K}(k) &= \frac{\mathbf{M}(k)\mathbf{X}(k)\omega(k)}{1 + \mathbf{X}^T(k)\mathbf{M}(k)\mathbf{X}(k)\omega(k)}; \quad \mathbf{M}(k) \\ &= \frac{1}{\rho}\mathbf{P}(k-1) \end{aligned} \quad (12)$$

$$\mathbf{P}(k) = (\mathbf{I} - \mathbf{K}(k)\mathbf{X}^T(k))\mathbf{M}(k) \quad (13)$$

Here, the term  $\omega(k)$  is defined by (8), when the initial estimate,  $\mathbf{W}_0$ , is replaced by the preceding estimate,  $\hat{\mathbf{W}}(k-1)$ , while the prediction error (residual) in (11) is given by (2), when the unknown parameter vector,  $\mathbf{W}$ , is substituted by  $\hat{\mathbf{W}}(k-1)$ .

Application of recursive M robust estimation algorithm (10)–(13) assumes the non-linear transformation,  $\psi(\cdot)$  in (4), as well as the scale factor,  $s$  in (8), and the forgetting factor,  $\rho$  in (12), to be known. Since the scale factor,  $s$ , represents an estimate of the unknown noise standard deviation,  $\sigma$ , and the argument of non-linearity  $\psi(\cdot)$  in (8) is the normalised residual, the non-linear transformation  $\psi(\cdot)$  in (8) is defined by (4) with the unit variance, i.e.  $\sigma = 1$ . Particularly, if one chooses the linear transformation in (8),  $\psi(x) = x$ , this results in the unit weight,  $\omega(k) = 1$  in (8), and algorithm (10)–(13) reduce to the standard RLS algorithm with FF defined by (10) and (11), where the corresponding matrices, instead of (12) and (13), are given by [5, 25, 26]

$$\mathbf{K}(k) = \mathbf{P}(k-1)\mathbf{X}(k)[\rho + \mathbf{X}^T(k)\mathbf{P}(k-1)\mathbf{X}(k)]^{-1} \quad (14)$$

$$\mathbf{P}(k) = \frac{1}{\rho} \left( \mathbf{P}(k-1) - \frac{\mathbf{P}(k-1)\mathbf{X}(k)\mathbf{X}^T(k)\mathbf{P}(k-1)}{\rho + \mathbf{X}^T(k)\mathbf{P}(k-1)\mathbf{X}(k)} \right) \quad (15)$$

In addition, if one defines the Huber's non-linearity in (4) by using the non-normalised argument on the right-hand side of (4), yielding

$$\psi(x) = \min(|x|, \Delta\sigma)\text{sign}(x) \quad (16)$$

and approximate the first derivative,  $\psi'(\cdot)$ , by the weighted term  $\omega(x) = \psi(x)/x$  in (8), together with the application of the winsorised residual,  $\psi(e)$ , instead of original one,  $e$ , in the parameter update equation (10), algorithm (10)–(13) can be rewritten as

$$\hat{\mathbf{W}}(k) = \hat{\mathbf{W}}(k-1) + \mathbf{P}(k)\mathbf{X}(k)\psi[e(k)] \quad (17)$$

$$\mathbf{P}(k) = \mathbf{P}(k-1) - \frac{\mathbf{P}(k-1)\mathbf{X}(k)\mathbf{X}^T(k)\mathbf{P}(k-1)\psi'[e(k)]}{1 + \psi'[e(k)]\mathbf{X}^T(k)\mathbf{P}(k-1)\mathbf{X}(k-1)} \quad (18)$$

where the prediction error (residual),  $e$ , is given by (11).

The obtained algorithm in (3), (11), (16)–(18), represents the standard M robust RLS (RRLS), where the common approach is to estimate the unknown noise standard deviation,  $\sigma$ , by the MAD based scale factor in (9). This algorithm can be exactly derived by applying the Newton-Raphson iterative method for solving the non-linear optimization problem in (1), with the unit parameters  $s$  and  $\rho$ , respectively. Here, the non-linearity,  $\psi$ , in (16) represents the first derivative of the loss function,  $\phi$ , in (1) [24–26]. It should be noted that the parameter update equation in (17) is non-linear, in contrast to the linear parameter update equation in (10). However, both procedures for generating the weighting matrix sequences,  $\mathbf{P}(k)$ , in (12), (13), and (18), respectively, are non-linear. Later, it will be shown that the scheme for generating the weighting matrix is very important for achieving the practical robustness.

The proposed parameter estimation algorithm (10)–(13) are derived from the M robust concept that is conservative, so the quality of parameter estimates may degrade without further adaptations of  $s$  and  $\rho$  variables.

#### 3.1 Adaptive robustified estimation of scale factor

As an adaptive robust alternative to the non-recursive robust MAD estimate in (9), the scale factor,  $s$ , can be estimated simultaneously with the filter parameter vector,  $\mathbf{W}$ . Namely, if  $\bar{p}(n)$  is the pdf of zero-mean Gaussian white noise,  $n(k)$ , in Fig. 1, with the unit variance, then the pdf of noise with some variance,  $\sigma^2$ , is given by  $p(n) = \bar{p}(n/\sigma)/\sigma$ . Thus, one can define an auxiliary performance index, in the form of the conditional maximum-likelihood (ML) criterion, [25, 26].

$$\begin{aligned} J(\sigma/\mathbf{W}) &= E \left\{ F \left( \frac{e(k, \mathbf{W})}{\sigma} \right) / \mathbf{W} \right\}; \quad F(n) \\ &= -\ln(p(n)) \end{aligned} \quad (19)$$

where  $e(\cdot)$  is the prediction error signal in (11) and  $E\{\cdot/\mathbf{W}\}$  represents the conditional mathematical expectation when the parameter vector,  $\mathbf{W}$ , is given. Furthermore,

one can use the Newton's stochastic algorithm for recursive minimization of the performance index in (19) [25, 26].

$$s(k) = s(k-1) - \left[ k \frac{\partial^2 J(s(k-1)/\hat{\mathbf{W}}(k-1))}{\partial \sigma^2} \right]^{-1} \left[ k \frac{\partial J(s(k-1)/\hat{\mathbf{W}}(k-1))}{\partial \sigma} \right] \quad (20)$$

where  $\hat{\mathbf{W}}(k)$  and  $s(k)$  are the corresponding estimates, at time instant  $k$ , of  $\mathbf{W}$  and  $\sigma$ , respectively. In addition, let us introduce the empirical approximation of the criterion (19) as

$$J_k(s/\hat{\mathbf{W}}) = \frac{1}{k} \sum_{i=1}^k F\left(\frac{e(i, \hat{\mathbf{W}})}{s}\right) \quad (21)$$

Under certain conditions, with  $k$  increasing,  $J_k$  in (21) approaches to  $J$  in (19). Moreover, since  $p(n) = \bar{p}(n/\sigma)/\sigma$ , one obtains from (19)

$$F(n) = \ln(\sigma) + f\left(\frac{n}{\sigma}\right); \quad f\left(\frac{n}{\sigma}\right) = -\ln\left(\bar{p}\left(\frac{n}{\sigma}\right)\right) \quad (22)$$

In addition, with large  $k$  and by using the optimality conditions, yielding

$$\frac{\partial J(s/\hat{\mathbf{W}})}{\partial \sigma} \approx \frac{\partial J_k(s/\hat{\mathbf{W}})}{\partial \sigma}; \quad \frac{\partial^2 J(s/\hat{\mathbf{W}})}{\partial \sigma^2} \approx \frac{\partial^2 J_k(s/\hat{\mathbf{W}})}{\partial \sigma^2}; \quad J_{k-1}(s/\hat{\mathbf{W}}) \approx 0 \quad (23)$$

one obtains from (20)–(23) an approximate optimal solution in the recursive form

$$ks(k) = (k-1)s(k-1) + e(k, \hat{\mathbf{W}}(k-1))g(e(k, \hat{\mathbf{W}}(k-1))/s(k-1)) \quad (24)$$

or equivalently

$$ks^2(k) = (k-1)s^2(k-1) + e^2(k, \hat{\mathbf{W}}(k-1))\omega(k) \quad (25)$$

Here, the robust weighting term,  $\omega(k)$ , is defined by (8), when  $\psi$  function is changed by  $g$  function, while  $\mathbf{W}_0$  is substituted by  $\hat{\mathbf{W}}(k-1)$  and  $s$  by  $s(k-1)$ , respectively. As mentioned before, in M robust estimation, we wish to design estimators that are not only quite efficient in the situations when the underlying noise pdf is normal but also remain high efficiency in situations when this pdf possesses longer tails than the normal one, generating the outliers [21–27]. Thus, we can define M robust estimator not exactly as the ML estimator based on the standard normal pdf  $\bar{p}(n)$ , with zero-mean and unit variance, but ML estimator corresponding to a pdf  $\bar{p}(n)$  that is similar to the standard Gaussian pdf in the middle, but has heavier tails than the normal one. This corresponds, for example, to the double exponential, or the Laplace pdf. Such choice corresponds further to the  $f(\cdot)$  function in (22) being equal to the Huber's M robust score function,  $\phi(\cdot)$ , in (1), with the first derivative,  $\psi(\cdot) = \phi'(\cdot)$ , given by

(4), [21–23]. Thus, in this case,  $g(\cdot) = f'(\cdot)$  reduces to the  $\psi$ -function in (4), for which the noise standard deviation,  $\sigma$ , is equal to one. In addition,  $\omega(k)$  in (25) is given by (8), with  $\mathbf{W}_0$  and  $s$  being equal to  $\hat{\mathbf{W}}(k-1)$  and  $s(k-1)$ , respectively. Furthermore, when the pdf,  $\bar{p}(n)$ , is the standard normal, the influence function  $g(\cdot)$  is linear and the weighting term  $\omega(k)$  in (25) becomes equal to one [13]. Finally, the application of the recursive algorithm (24) or (25) requires the initial guess  $s(0)$  to be given beforehand, as it is done in Eq. (37).

The net effect is to decrease the consequence of large errors, named outliers. The estimator is then called robust. In algorithm (24) or (25), this goal is achieved through the weighting term in (8), where  $\psi$  is the saturation type non-linearity in (4). Thus, the function  $\psi(\cdot)$  is linear for small and moderate arguments, but increases more slowly than the linear one for large arguments. Furthermore, in the normal case without outliers, one should want most of the arguments of the  $\psi(\cdot)$  function to satisfy the inequality  $|e(i, \mathbf{W}_0)/s| \leq \Delta s$ , because then  $\psi(e(i, \mathbf{W}_0)/s) = e(i, \mathbf{W}_0)/s$  and  $\omega$  in (8) is equal to unity. On the other hand, for large arguments satisfying  $|e(i, \mathbf{W}_0)| > \Delta s$ , the weighting term in (8) decreases monotonously with the argument absolute value and, as a consequence, reduces the influence of outliers.

For the scale estimation problem in question, the unknown noise variance is assumed to be constant. Therefore, after time increases, the derived recursive estimates (24) or (25) converge towards the constant value. Equations (24) and (25) represent the linear combination of the previous estimate and the robustly weighted current estimation error. The coefficients in the linear combination,  $1 - 1/k$  and  $1/k$ , depend on the time step,  $k$ . Thus, as  $k$  increases, these coefficients converge towards unity and zero, respectively. As a consequence, after a sufficiently large time step,  $k$ , the correcting term in (24) and (25) that is multiplied by the coefficient  $1/k$  is close to zero, so that the proposed algorithm eliminates the effect of possible outliers.

Moreover, in many practical problems, it is of interest to consider the situation in which the noise variance is time-varying. However, due to the described saturation effect, the proposed estimator cannot catch the changes. These situations can be covered by simple extension of Eqs. (24) and (25). A simple but efficient solution can be obtained by resetting. The forgetting or discounting factor,  $1/k$ , in (24) and (25) is then periodically reset to the unit value, for example each 100 steps, and the initial guess,  $s(0)$ , has to be set to the previous estimate.

### 3.2 Strategy for choosing adaptive robustifying variable forgetting factor

As mentioned before, the value of forgetting factor FF,  $\rho$ , belongs to the set of real numbers  $(0,1]$ , as it has to give

more heavily weights to the current samples, in order to provide for tracking of time-varying filter parameters. If a value of FF,  $\rho$ , is close to one, it needs rather long time to find the true coefficients. However, the parameter estimates should be with high quality in stationary situations. The speed of adaptation can be controlled by the asymptotic memory length, defined by [7, 25].

$$N = \frac{1}{1-\rho} \tag{26}$$

Thus, it follows from (26) that progressively smaller values of FF,  $\rho$ , provides an estimation procedure with smaller size of data window, what is useful in non-stationary applications.

If a signal is synthesised of sub-signals having different lengths of memory, changing between a minimum value,  $N_{\min}$ , and a maximum one,  $N_{\max}$ , the time-varying signal model coefficients can be estimated by using Eqs. (4), (8), (10–13), and (25), assigning to each sub-signal the corresponding FF,  $\rho$ , from (26), varying between  $\rho_{\min}$  and  $\rho_{\max}$ . However, in practice, the memory length and the starting points of sub-signals are unknown in advance. Thus, one has to find the degree of signal non-stationarity, in order to generate the value of FF,  $\rho$ , in the next step. Although many adaptation procedures have been analysed by changing the memory length, the method using the extended prediction error (EPE) criterion is emphasised, since it involves rather easy computation, and has good adaptability in non-stationary situations, and a low variance in the stationary one [10, 12, 13]. Particularly, the extended prediction error criterion, as a local measure of signal non-stationarity, is defined by [10]:

$$E(k) = \frac{1}{L} \sum_{i=k-L+1}^k e^2(i, \hat{\mathbf{W}}(i-1)) \tag{27}$$

Here,  $e(\cdot)$  is the prediction error, or residual, in (11), and the length of the sliding window  $L$  is a free parameter, which has to be set.

Thus, the quantity  $E(k)$  in (27) represents a measure of the local variance of prediction residuals at the given sliding data frame of size  $L$ , and it contains the information about the degree of data non-stationarity. In addition,  $L$  should be a small number compared to the minimum asymptotic memory length, so that averaging does not obscure the non-stationarity of signal. Thus, the value of  $L$  represents the trade-off between the estimation accuracy and tracking ability of time varying parameters. Unfortunately, the EPE statistics in (27), like the sample mean and sample variance, lacks robustness towards outliers [21–23]. Therefore, we suggest to derive a robust alternative to the EPE criterion in (27) using the M robust approach. Thus, if the prediction errors

$e(k)$  in (2) are assumed to be independent and identically distributed (i.i.d) random variables, a simple parameter estimation problem can be constructed. Define a random variable (r.v.),  $\zeta$ , on the sample space  $\Omega$ , from which the data  $e(k)$ ,  $k = 1, 2, \dots, N$ , are obtained. Based on empirical measurements, the mean,  $m_e$ , and variance,  $\sigma_e^2$ , of the unknown distribution of r.v.,  $\zeta$ , are to be estimated. As in (1), the robust M estimate  $\hat{m}_e(N)$  of  $m_e$  is defined by

$$\sum_{k=1}^N \psi \left( \frac{e(k) - \hat{m}_e(N)}{s} \right) = 0 \tag{28}$$

where  $\psi(\cdot)$  is the Huber's influence function in (4). Here,  $s$  is an estimate of the scale of the data  $\{e(k)\}$ . The estimating Eq. (28) is non-linear, and some form of WLS approximation, similar to (6), can be used for its solution. Moreover, a popular statistic  $s$  is the MAD estimation in (9). Although  $s$  in (9) is robust, it turns out to be less efficient than some other robust estimates of variance [21–23]. However,  $s$  is a nuisance parameter in the computation of  $m_e$ , and in this context, the efficient issue is not as crucial as in the estimation of the variance of the data, estimates of the latter being used in robustifying the EPE criterion (27) and setting the VFF. A more efficient estimator of the data variance should be based on the asymptotic variance formula for the location M estimate in (28). When  $s = \sigma_e$ , this formula is given by [21]

$$V = \lim_{N \rightarrow \infty} E \{ N [\hat{m}_e(N) - m_e]^2 \} = \frac{\sigma_e^2 E \left\{ \psi^2 \left( \frac{e(k) - m_e}{\sigma_e} \right) \right\}}{E^2 \left\{ \psi' \left( \frac{e(k) - m_e}{\sigma_e} \right) \right\}} \tag{29}$$

A natural estimate of  $V$  in (29) is

$$\hat{V}_N = s^2 \frac{\frac{1}{N} \sum_{k=1}^N \psi^2 \left( \frac{e(k) - \hat{m}_e(N)}{s} \right)}{\frac{1}{N} \sum_{k=1}^N \psi' \left( \frac{e(k) - \hat{m}_e(N)}{s} \right)} \tag{30}$$

where  $\hat{V}_N$  in (30) would appear to be a reasonable estimate of  $\sigma_e^2$ , with  $\hat{m}_e(N)$  being the M estimate of  $m_e$  in (28). Therefore, given  $m_e = 0$ , producing  $\hat{m}_e(N) = 0$  and the estimate  $s$  from the recursive M robust scale estimate,  $s(k)$ , in (25), a possible M robust alternative of the EPE criterion in (27) is given by

$$E_r(k) = s^2(k) \frac{\sum_{i=k-L+1}^k \psi^2\left(\frac{e(i, \hat{\mathbf{W}}(i-1))}{s(i)}\right)}{\sum_{i=k-L+1}^k \psi\left(\frac{e(i, \hat{\mathbf{W}}(i-1))}{s(i)}\right)} \quad (31)$$

where  $\psi(\cdot)$  is the Huber's influence function in (4), with  $\sigma = 1$  and  $\Delta = 1.5$ . If  $\psi(\cdot)$  is a linear function,  $\psi(x) = x$ , than the criterion (31) reduces to the standard EPE criterion in (27), under the assumption that the scale factor estimate,  $s(i)$ ,  $i = k - L + 1, \dots, k$ , on the sliding data frame of length  $L$  is close to the  $s(k)$  value.

On the other hand, the total noise variance robust estimate in (25) is rather insensitive to the local non-stationary effects. Therefore, in order to make the estimation procedure invariant to the noise level, one can define the normalised robust measure of non-stationarity or the so-called robust discrimination function

$$Q(k) = \frac{E_r(k)}{s^2(k)} \quad (32)$$

A strategy for choosing the VFF at current time instant,  $k$ , may now be defined by using the relations (25), (26), (31) and (32), that is

$$\rho(k) = 1 - \frac{1}{N(k)}, \quad N(k) = \frac{N_{\max}}{Q(k)} \quad (33)$$

Thus, the maximum asymptotic length,  $N_{\max}$ , will determine the adaptation speed. Furthermore, for a stationary signal with possible outliers, the quantity  $E_r(k)$  in (31) will converge to the noise variance, yielding  $Q(k) \approx 1$  and  $N(k) = N_{\max}$ . Finally, since (33) does not guarantee

that FF,  $\rho$ , does not become negative, a reasonable limit has to be placed on FF,  $\rho$ , yielding

$$\rho(k) = \max\left\{1 - \frac{1}{N(k)}, \rho_{\min}\right\} \quad (34)$$

where  $N(k)$  is given by the relations (25) and (31)–(33).

A brief description of the proposed algorithm with both scale and forgetting factors is given in Table 1. Since the proposed algorithm combines the three adaptive procedures (recursive robust parameter estimation, recursive robustified noise variance estimation and adaptive robustified variable forgetting factor calculation), the theoretical performance analysis is very difficult with the coupled algorithms. Therefore, the figure of merit of the proposed approach will be given by simulations in the next section. The following algorithms are tested:

1. Recursive robust weighted least squares type method, defined by (4), (8) and (10)–(13) together with the both recursive robust scale estimation in (25) and adaptive robustified VFF calculation in (31)–(34), denoted as RRWLSV (see Table 1).
2. Recursive least squares algorithm with exponentially weighted residuals defined by (11), (14) and (15) and VFF given by (27) and (32)–(34), denoted as RLSVF.
3. Recursive robustified least squares algorithm defined by (16)–(18) and MAD-based scale factor estimation in (9), denoted as RRLSS.

#### 4 Experimental analysis

To analyse the performances of previously discussed methods, a linear parameter estimation scenario in stationary and non-stationary contexts, related to the additive noise, is applied (see Fig. 1) [10]. The required

**Table 1** Summary of adaptive robust parameter estimation algorithm (10)–(13) with scale factor (25) and EPE-based VFF (27)–(30)

Step 1	Let at stage $k$ , $k \geq N$ , the parameter vector estimate $\hat{\mathbf{W}}(k-1)$ , the scale estimates $s(k-1), \dots, s(k-L+1)$ , the error signals $e(k-1), \dots, e(k-L+1)$ and the matrix $\mathbf{P}(k-1)$ from the $(L-1)$ previous stages are known.
Step 2	Take the current input, $x(k)$ , and form the regression vector in (3) $\mathbf{X}^T(k) = \{x(k), x(k-1), \dots, x(k-N+1)\}$ of length $N$ , assuming that the $(N-1)$ most recent inputs are given.
Step 3	Take the current output, $d(k)$ , and calculate the current error signal, $e(k)$ , from (11) using $\mathbf{X}(k)$ from step 2, and define the current data frame $E_L = \{e(k), e(k-1), \dots, e(k-L+1)\}$ of length $L < N$ , assuming that the $(L-1)$ most recent errors are previously stored.
Step 4	Calculate the normalised error $e(k)/s(k-1)$ and the winsorised error $\psi(e(k)/s(k-1))$ from (4) with $\sigma = 1$ ; then calculate the weight $\omega(k)$ in (25) by using (8) with $\mathbf{W}_0 = \hat{\mathbf{W}}(k-1)$ and $s = s(k-1)$ ; finally, calculate the scale factor $s(k)$ from (25).
Step 5	Define the current data frame of normalised residuals $E_{NL} = \{e(k)/s(k), e(k-1)/s(k-1), \dots, e(k-L+1)/s(k-L+1)\}$ from steps 1, 3 and 4; then calculate the robust discrimination function, $Q(k)$ , in (28), using the data set $E_{NL}$ ; finally, calculate the VFF, $\rho(k)$ , from (29) and (30).
Step 6	Calculate the winsorised error, $\psi(e(k)/s(k))$ , from (4), with $\sigma = 1$ and by using $e(k)$ from step 3 and $s(k)$ from step 4; then calculate the weight, $\omega(k)$ , in (8) by using (4) with $\mathbf{W}_0 = \hat{\mathbf{W}}(k-1)$ and $s = s(k)$ .
Step 7	Calculate the matrix, $\mathbf{M}(k)$ , in (12) with $\rho = \rho(k)$ from step 5; then calculate the matrix, $\mathbf{K}(k)$ , in (12) by using $\mathbf{X}(k)$ from step 2 and $\omega(k)$ from step 6.
Step 8	Calculate the parameter vector update, $\hat{\mathbf{W}}(k)$ , in (10), by using $d(k)$ and $e(k)$ from step 3, as well as $\mathbf{K}(k)$ from step 7.
Step 9	Calculate the weighting matrix, $\mathbf{P}(k)$ , in (13) by using $\mathbf{M}(k)$ and $\mathbf{K}(k)$ from step 7, together with $\mathbf{X}(k)$ from step 2.
Step 10	Tune the time counter, that is increase the time index, $k \leftarrow k + 1$ , and go back to step 2.

filter output,  $d(k)$ , is generated by passing the standard white Gaussian sequence,  $x(k)$ , of the zero-mean and unit variance, through the FIR system of the ninth order, with the true values of parameters

$$\mathbf{W} = [0.1, 0.2, 0.3, 0.4, 0.5, 0.4, 0.3, 0.2, 0.1]^T \quad (35)$$

In addition, a zero-mean white additive noise,  $n(k)$ , with corresponding variance is involved to its output. The value of variance is adopted so to give the desired signal-to-noise ratio (SNR) at the signal segment in question, before the impulsive noise component is introduced. Four situations regarding the additive noise are considered: stationary context with fixed variance and possible outliers and non-stationary context with changing variance and possible outliers. The variances are chosen so to give the different values of SNR equal to 15, 20 and 25 dB, respectively. The outliers, generated by the impulsive noise, are produced using the model  $n(k) = \alpha(k)A(k)$ , with  $\alpha(k)$  being an i.i.d binary sequence defined by the corresponding probabilities  $P(\alpha(k) = 0) = 0.99$  and  $P(\alpha(k) = 1) = 0.01$ , respectively, and  $A(k)$  is the zero-mean normal random variable with the variance  $\text{var}\{A(k)\} = 10^4/12$  that is independent of the random variable  $\alpha(k)$ . The random variable  $n(k)$  has zero-mean and variance proportional to  $\text{var}\{A(k)\}$  (see Appendix 3).

A priori information of the impulsive noise is used to choose the length,  $L = 5$ , of the sliding window that capture the non-stationarity. This means that no more than one outlier, in average, is in the sliding data frame of size  $L = 5$  during the robustified EPE calculation in (31) and MAD calculation in (9), respectively, when the fraction of outliers is 1 %.

The following algorithms, described in previous section, are tested: (1) the proposed adaptive robust filtering algorithm with the both robust adaptive scale and VFF, denoted as RRWLSV; (2) the conventional RLS with EPE based VFF, denoted as RLSVF; and (3) the standard M robust RLS with MAD-based scale factor, denoted as RRLS.

The analysed algorithms have been tested on both the time-varying parameter tracking ability and the log normalised estimation error norm

$$NEE(k) = 10 \log \frac{\|\hat{\mathbf{W}}(k) - \mathbf{W}\|^2}{\|\mathbf{W}\|^2} \quad (36)$$

where  $\|\cdot\|$  is the Euclidian norm, averaged on 30 independent runs. In each experiment, the following values are used to the initial conditions of analysed algorithms:

$$\hat{\mathbf{W}}(0) = 0, P(0) = 100 \times \mathbf{I}, s(0) = 1 \quad (37)$$

with  $\mathbf{I}$  being the identity matrix of corresponding order.

#### 4.1 Time-varying parameter tracking in a stationary zero-mean white normal noise with possible outliers

In this experiment, the first filter parameter  $w_1$  in (3) is changed using the trajectory depicted in Fig. 2. The variance of noise,  $n(k)$ , is taken so that the SNR of 25 dB is achieved before the impulsive noise component is added. Figure 3 gives a realisation of the zero-mean white Gaussian sequence without (Fig. 3a) and with impulsive component (Fig. 3b), respectively. In Figs. 4, 5 and 6, the true and the estimated trajectories of changing parameter, under the pure zero-mean white Gaussian noise, are shown. Moreover, the estimated scale and variable forgetting factors are also depicted in these figures. Figures 7, 8 and 9 give the simulation results in the case of zero-mean and white and Gaussian noise with outliers. Figures 10 and 11 depicted the normalised estimation error in (36) obtained in the two discussed stationary noise environments, for the three analysed algorithms.

The obtained results have shown that the algorithms RRWLSV and RLSVF provide good and comparable results, due to the application of VFF in the weighted matrix update equation (see Eqs. 10, 12, 14 and 15), while the algorithm RRLSS gives bad parameter tracking performance, since it uses the fixed unit value of FF in

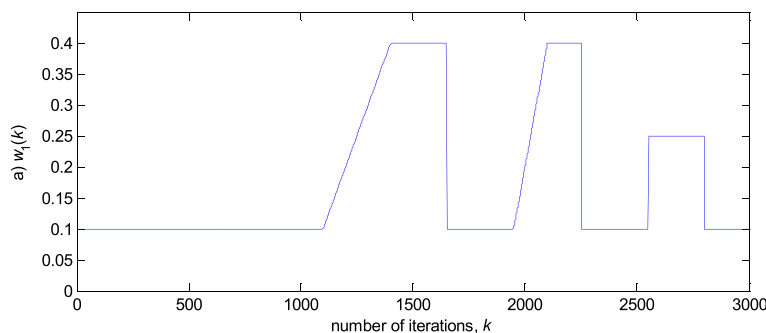
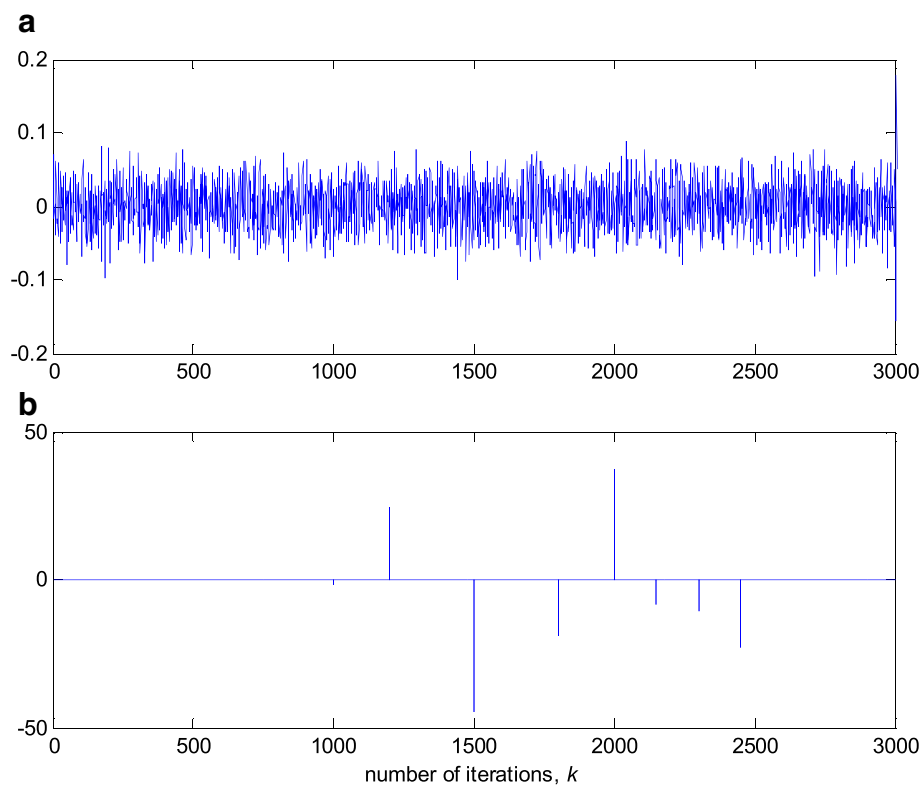


Fig. 2 Time varying parameter trajectory  $w_1(k)$  in (3)





**Fig. 3** Realisation of additive zero-mean white noise  $n(k)$ . **a** Pure zero-mean Gaussian samples (SNR = 25 dB). **b** Gaussian samples contaminated with outliers

(18), (see Figs. 4, 5, 6 and 10). Moreover, the results of simulations have indicated that the conventional RLS with VFF (RLSVF) is highly sensitive to outliers, since it does not use a robust influence function,  $\psi$ , in the parameter and weighted matrix update equations (see Eqs. 14 and 15). On the other hand, the RRLSS algorithm using scale factor estimation in (9), or combined the adaptive robustified scale and VFF factors (RRWLSV), are rather insensitive to outliers, due to the effect of robust influence function,  $\psi$ , in the parameter and weighted matrix update equations (Eqs. 17 and 18 or Eqs. 12 and 13). In addition, the algorithm RRWLSV has much better parameter tracking ability in the case of both time-varying parameters and impulsive noise environment than the algorithm RRLSS (see Figs. 7, 8, 9 and 11), due to the combined effects of scale factor and VFF (see Eqs. 12 and 13).

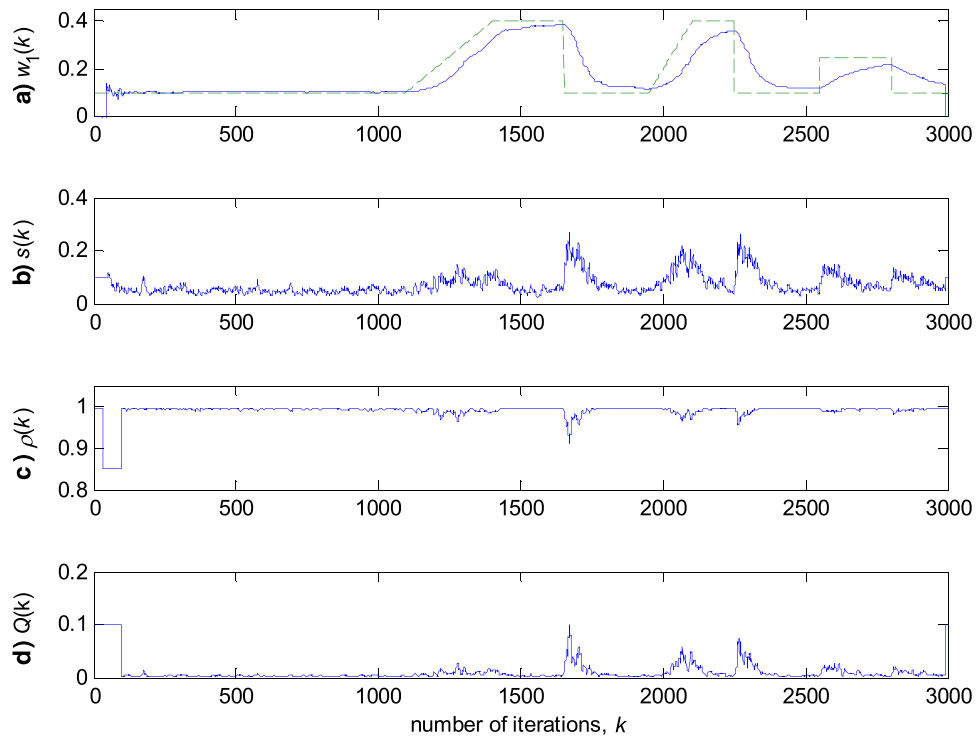
#### 4.2 Time-invariant parameter tracking in a non-stationary zero-mean normal noise with possible outliers

In this experiment, the filter parameters are taken to be time invariant, but the noise variance is changed during the simulation, producing a non-stationary signal. The variance of additive noise,  $n(k)$ , is taken so that for the three sequential signal segments, the SNRs of 25, 15 and

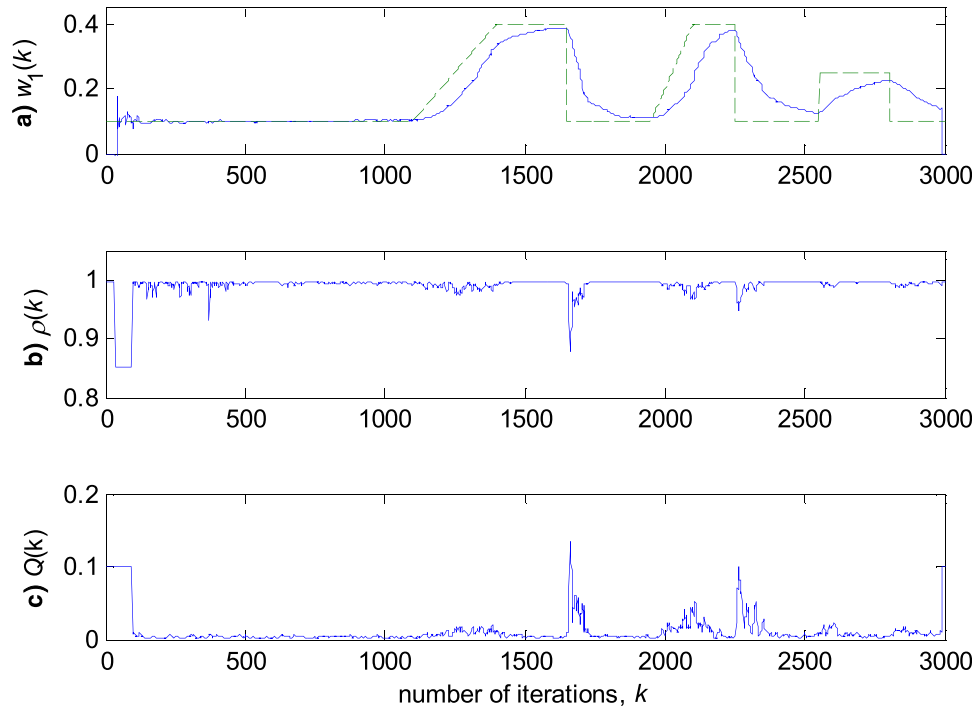
20 dB, respectively, are produced, before the impulsive noise component is added. In order to obtain better capture of long-term non-stationarity, in the sense of achieving good estimates of noise level changes, the scale factor,  $s$ , is averaged on the data frame of 400 samples. On the other hand, this will not affect the ability of VFF algorithm to track the filter parameter changes. Figure 12 shows a realisation of noise, without (Fig. 12a) and with (Fig. 12b) impulsive noise component.

Figures 13 and 14 depict the obtained values of normalised estimation error (NEE) criterion in (36), without (Fig. 13) and with (Fig. 14) the presence of impulsive noise component.

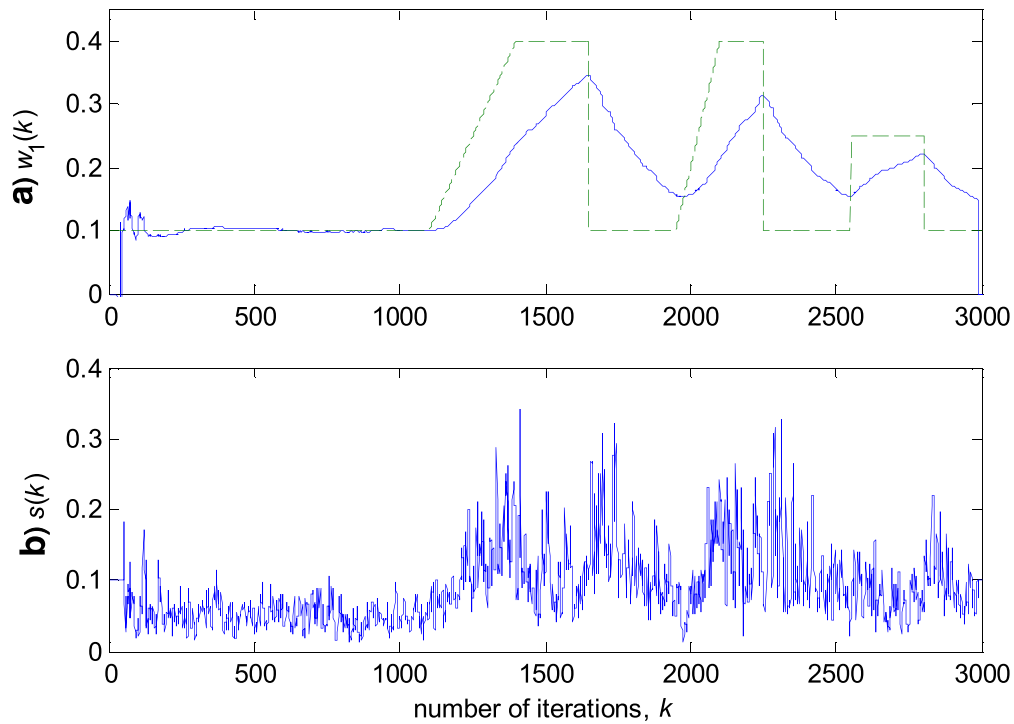
In a pure zero-mean normal white noise environment (see, Fig. 12a), at the beginning of parameter estimation trajectory, all three discussed methods have similar behaviour (Fig. 13). Moreover, all three algorithms are rather insensitive to the changes of noise variance, due to the effects of scale factor estimations in (9) or (25), or VFF in (27)–(34) or (31)–(34), respectively, and give similar results at the whole parameter estimation trajectory, since there are no outliers (Fig. 13). The presented results in Fig. 14 indicated that the conventional RLS method with VFF (RLSVF) is highly sensitive to outliers, due to the lack of non-



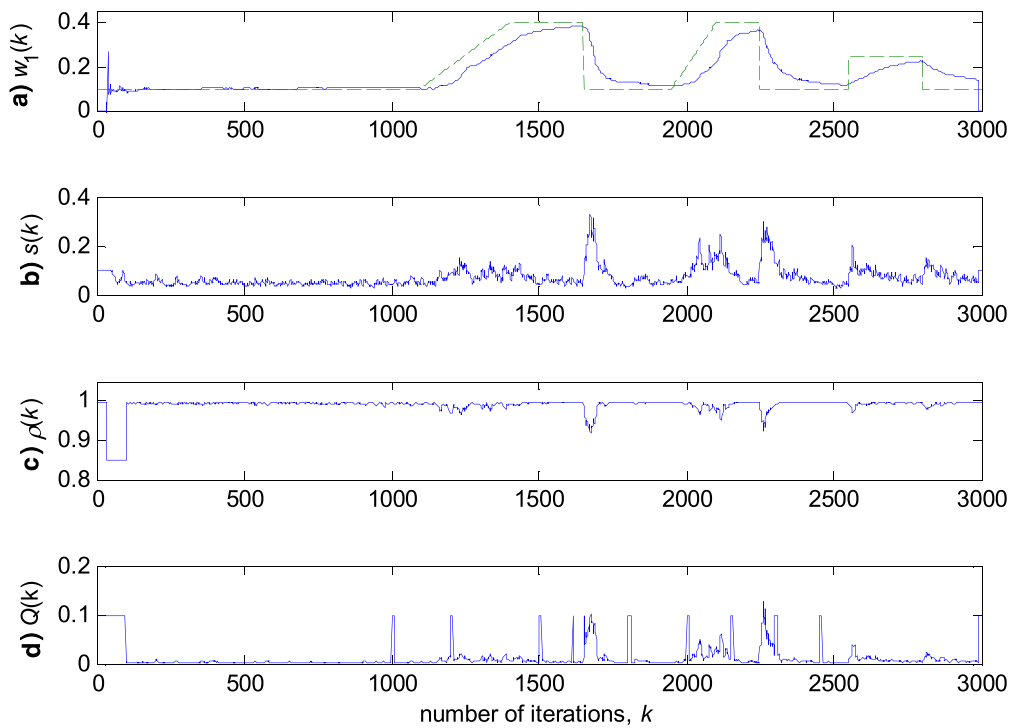
**Fig. 4** Experimental results for RRWLSV algorithm in zero-mean white Gaussian noise (see Fig. 3a). **a** Estimated (*solid line*) and true parameter (*dashed line*) trajectories (see Fig. 2). **b** Scale factor estimation,  $s(k)$ . **c** VFF calculation,  $\rho(k)$ . **d** Discrimination function calculation,  $Q(k)$



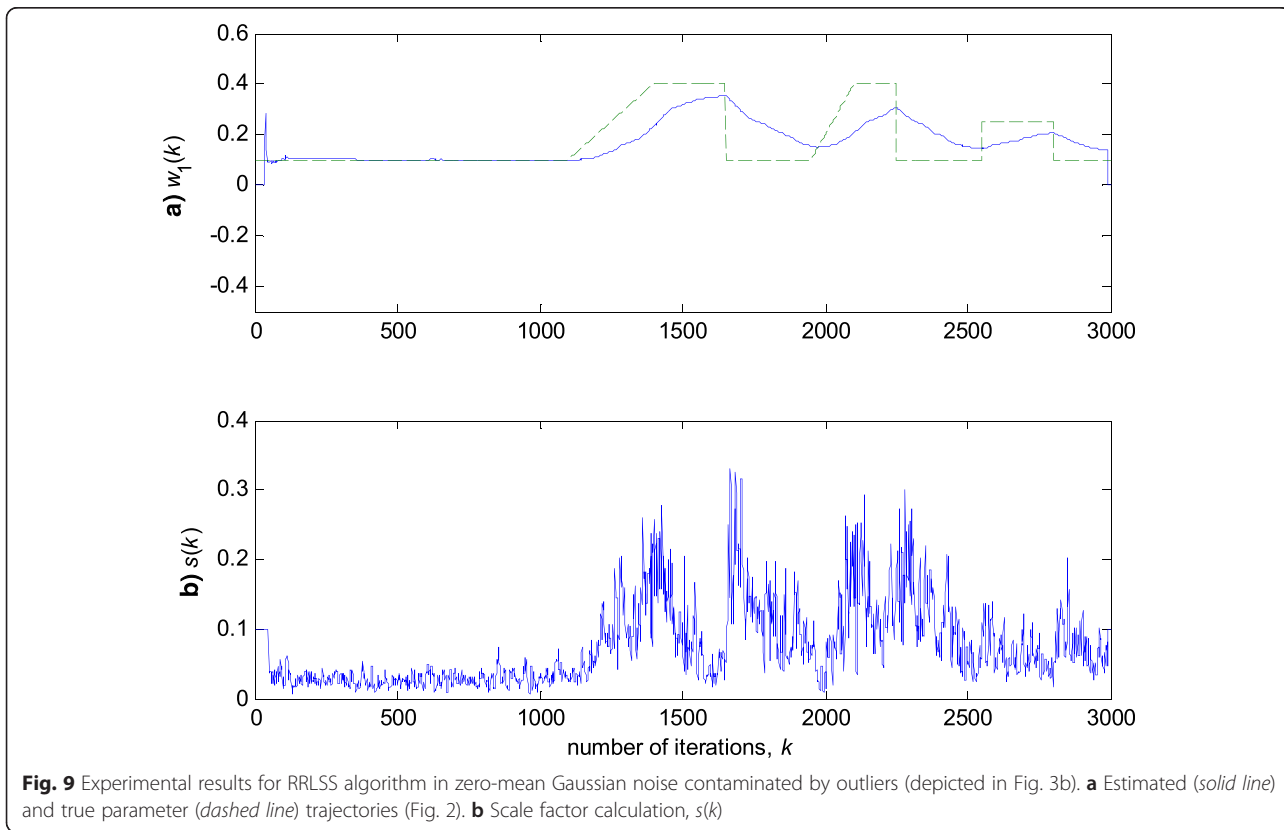
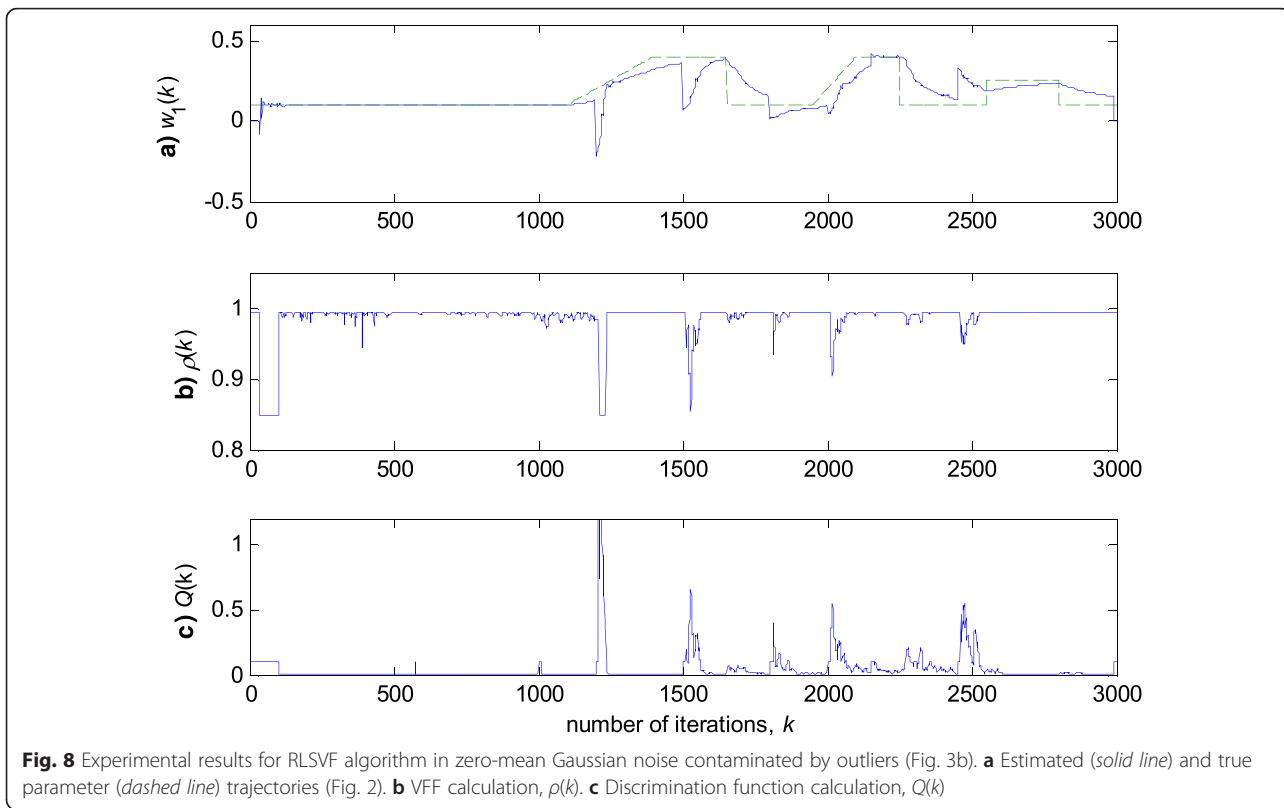
**Fig. 5** Experimental results for RLSVF algorithm in zero-mean Gaussian noise (Fig. 3a). **a** Estimated (*solid line*) and true parameter (*dashed line*) trajectories (see Fig. 2). **b** VFF calculation,  $\rho(k)$ . **c** Discrimination function calculation,  $Q(k)$

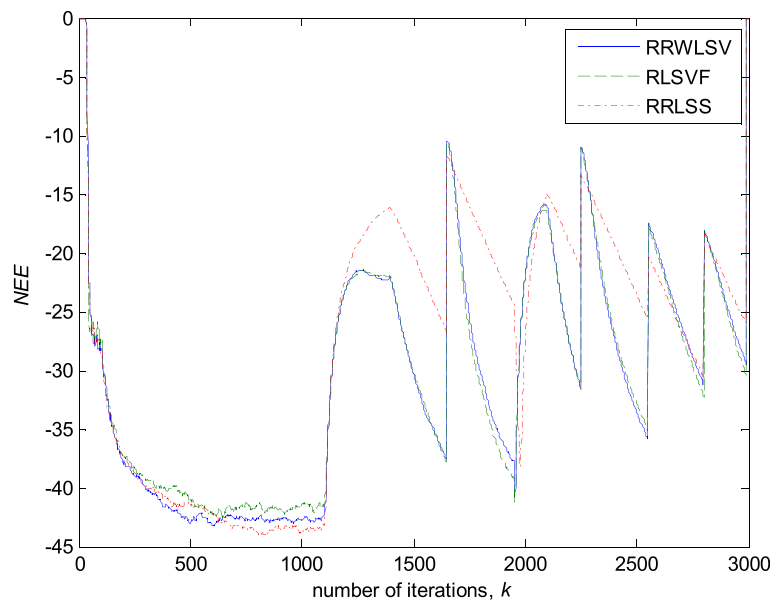


**Fig. 6** Experimental results for RRLSS algorithm in zero-mean Gaussian noise (Fig. 3a). **a** Estimated (solid line) and true parameter (dashed line) trajectories (Fig. 2). **b** Scale factor estimation,  $s(k)$



**Fig. 7** Experimental results for RRWLSV algorithm in zero-mean Gaussian noise contaminated by outliers (Fig. 3b). **a** Estimated (solid line) and true parameter (dashed line) trajectories (Fig. 2). **b** Scale factor estimation,  $s(k)$ . **c** VFF calculation,  $\rho(k)$ . **d** Discrimination function calculation,  $Q(k)$



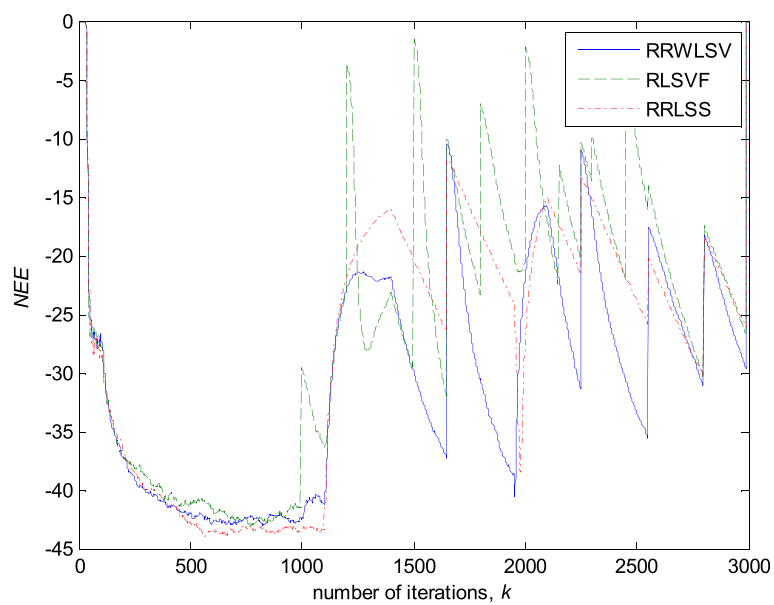


**Fig. 10** Normalised estimation errors  $NEE$  in (36), for different algorithms, in stationary zero-mean white normal Gaussian noise ( $SNR = 25\text{dB}$ ); Experimental conditions are given in Figs. 2 and Fig 3a

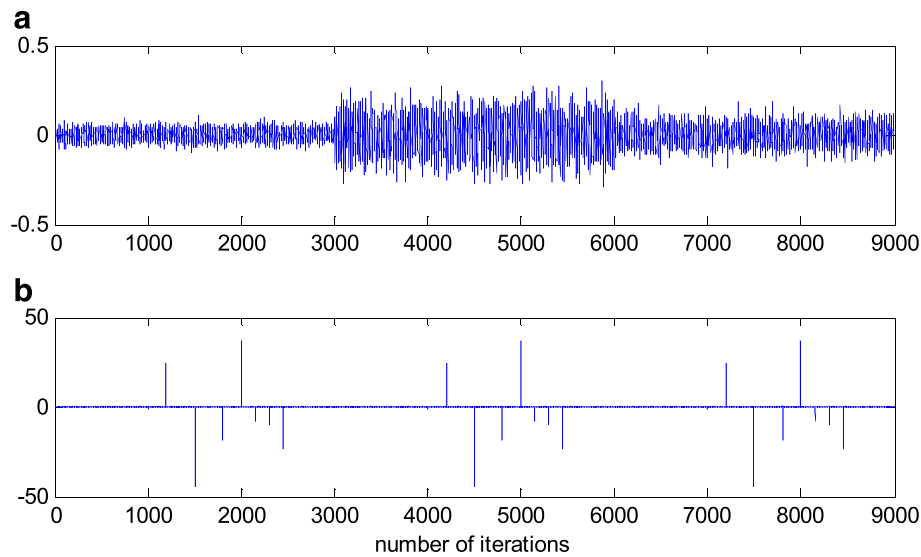
linear residual transformation,  $\psi$  (see Eqs. 14 and 15). On the other hand, the robust algorithms RRLSS and RRWSLV are rather insensitive to impulsive noise component, and give also comparable results, due to the effect of robust influence function,  $\psi$ , in the parameter and matrix update equations (see Eqs. 17 and 18 or Eqs. 12 and 13).

**4.3 Time-varying parameter tracking in a non-stationary zero-mean normal noise with possible outliers**

In this experiment, the simulation scenario is the combination of the previous two examples, with the exception that the first filter parameter,  $w_1$ , is adopted to be time-varying, using the parameters trajectory presented in Figs. 15, 16 and 17 depict the obtained  $NEE$  criterion



**Fig. 11** Normalised estimation errors  $NEE$  in (36) for different algorithms in stationary zero-mean white normal noise ( $SNR = 25\text{ dB}$ ) contaminated by outliers; experimental conditions are given in Figs. 2 and Fig 3b

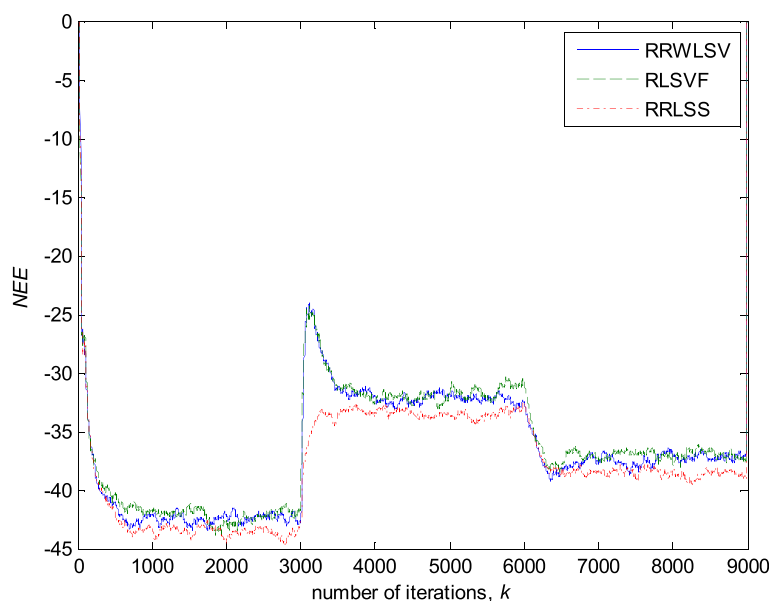


**Fig. 12** Realisation of additive noise,  $n(k)$ . **a** Non-stationary zero-mean white normal noise samples with changing variances (SNR = 25, 15 and 20 dB). **b** Non-stationary zero-mean white normal noise samples, with changing variances, contaminated by outliers

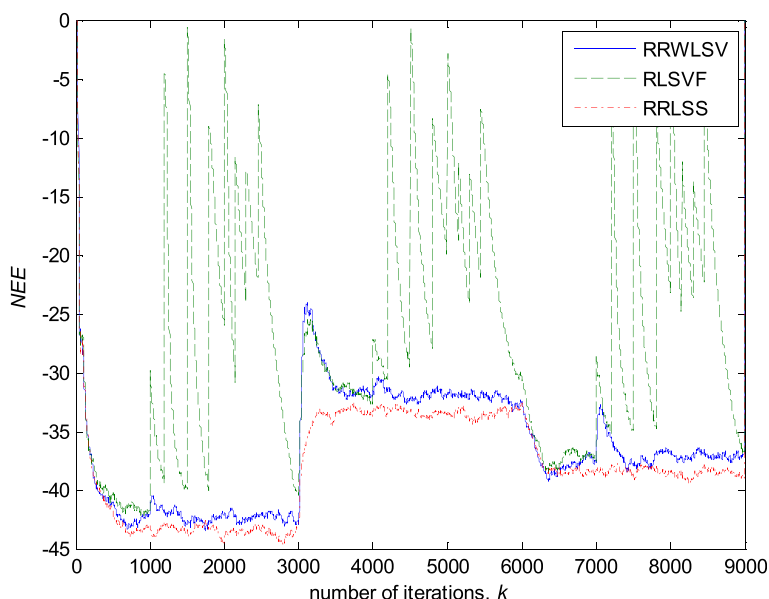
(36), for different algorithms, without (Fig. 16) and with (Fig. 17) outliers, added to the non-stationary zero-mean normal noise with changing variance (see Fig. 12). The obtained results are in accordance to the conclusions derived from the previous two experiments.

In summary, the obtained results from the all three experiments have shown that, in order to properly estimate the time-varying parameters under the non-stationary and impulsive noise environment, both robustified VFF and adaptive M robust estimator with simultaneous estimation

of parameters and scale factor (RRWLSV) are required. Moreover, an adequate non-linear residual transformation in the parameter update equation (Eq. 17 in RRLSS algorithm) is not sufficient for protecting well against the influence of outliers. Additionally, an important problem is also related to the way of recursive generation of the weighted matrix,  $\mathbf{P}$ . It is found that the introduction of the Huber's saturation type non-linearity,  $\psi$ , coupled with the proper decrease of the weighted matrix,  $\mathbf{P}$ , depending on the non-linearly transformed residuals and VFF (Eqs. 12



**Fig. 13** Normalised estimation error norm ( $NEE$ ), for different algorithms in non-stationary Gaussian noise environment (Fig. 12a), and fixed parameters in (35)



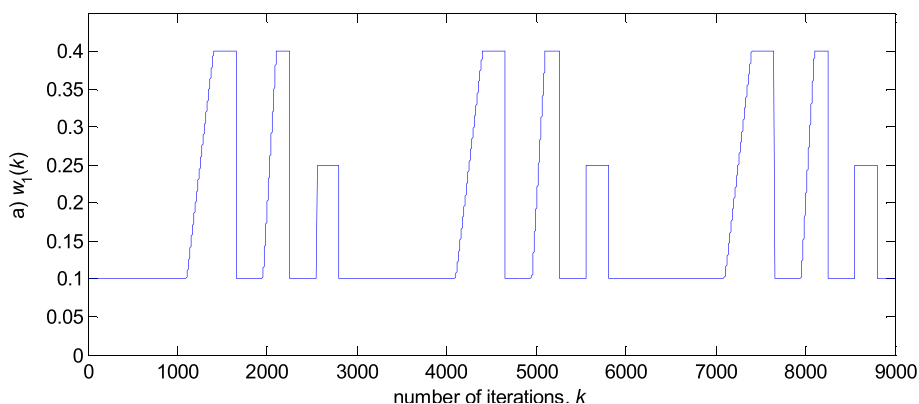
**Fig. 14** Normalised estimation error norm  $NEE$ , for different algorithms in non-stationary and impulsive noise environment (Fig. 12b), and fixed parameters in (35)

and 13) provides low sensitivity to outliers and good parameter tracking performance (Figs. 4 and 7). The conventional M robust estimator (RRLSS) may converge very slowly, due to the introduction of the non-linearity first derivate,  $\psi'$ , and the unit FF in the recursive generation of the weighted matrix (Eq. 18). Namely, large residual realisations,  $e$ , make the decrease of the weighted matrix,  $\mathbf{P}$ , very slow, since  $\psi'(e) = 0$  in the saturation range (Eq. 18). This, in turn, results in an effect producing a slow convergence of parameter estimates (Figs. 6 and 9). The conventional RLS is also sensitive to outliers, but in a different way. In this algorithm, the weighted matrix,  $\mathbf{P}$ , is not influenced by the residuals (Eq. 15). As a consequence, the fast decrease of  $\mathbf{P}$ , independent of residuals, leads to the biased parameter estimates in the presence of outliers (see Fig. 8).

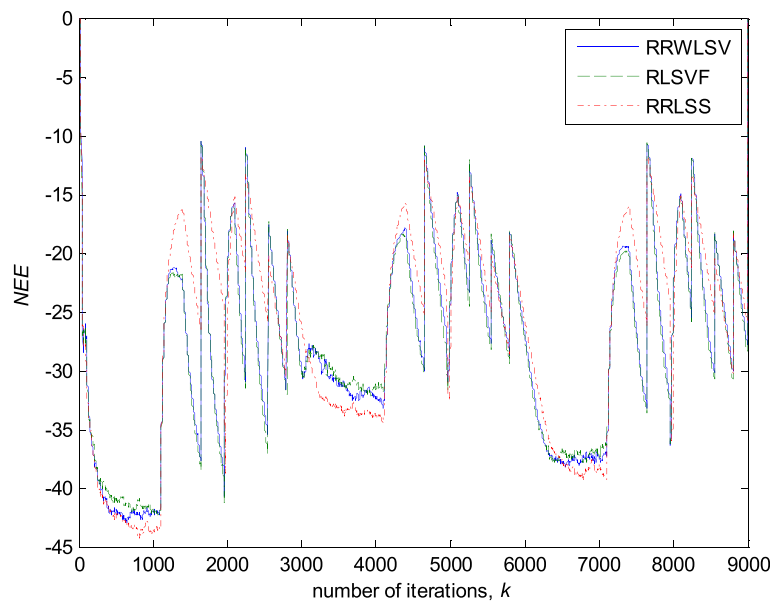
In addition, the estimation of unknown noise variance (Eqs. 9 or 25) is essential to residual normalisation, in order to achieve a low sensitivity to the parameters defining the non-linear transformation of the prediction residuals in (4).

**4.4 Influence of outlier statistics**

A real outlier statistics is not exactly known in practice, so a low sensitivity to outliers is very important for achieving the practical robustness. The effect of desensitising the parameter estimates related to the influence of outliers is illustrated in Tables 2 and 3, depicting the evaluation of sample mean and sample variance of the  $NEE$  statistics (36), for different outlier probability and intensity. It can be observed that the proposed algorithm



**Fig. 15** Time varying parameter trajectory,  $w_1(k)$



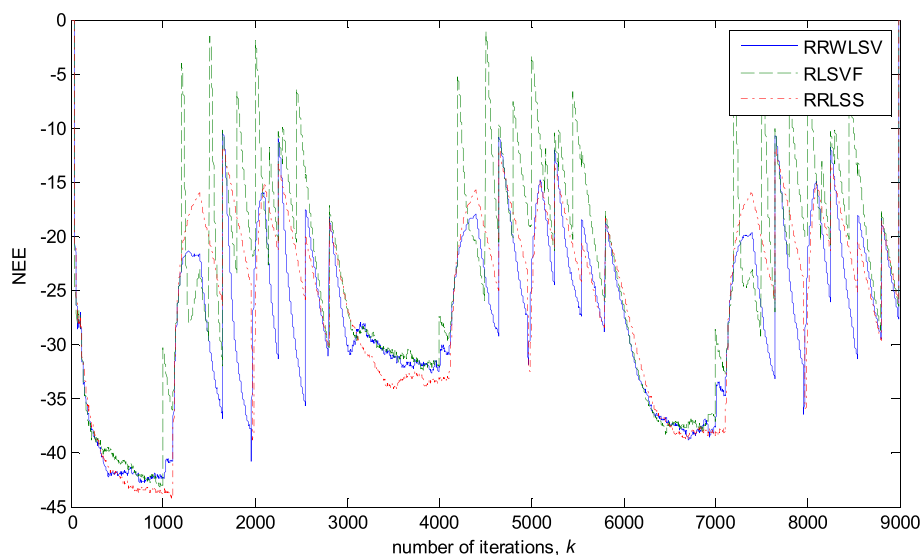
**Fig. 16** Normalised estimation error norm ( $NEE$ ), for different algorithms, in non-stationary zero-mean Gaussian noise environment (Fig. 12a) and the time-varying parameter trajectory (Fig. 15)

automatically damp out the effects of outliers, and it is rather insensitive to various fractions of outliers (see Table 2), up to 15 %, as well as to high level outliers (see Table 3). For higher percentage of outliers, more than 20 %, the normal noise model contaminated by outliers is not adequate. Furthermore,  $s(k)$  is a nuisance parameter in the estimation of filter coefficients, as well as in the VFF computation. Its proper estimate is crucial for good performance of overall estimator, consisting of three coupled adaptive schemes. Since

the complete algorithm performs quite well, this means that the estimation of scale factor also performs properly.

#### 4.5 Influence of initial conditions

The proposed RRWLSV algorithm, exposed in Table 1, is non-linear and, consequently, may be highly influenced by the initial conditions,  $\mathbf{W}(0)$  and  $\mathbf{P}(0)$  in (33), respectively. However, from the practical point of view, a low sensitivity to the initial conditions represents the



**Fig. 17** Normalised estimation error norm ( $NEE$ ), for different algorithms, in non-stationary noise contaminated by outliers (Fig. 12b), and the time-varying parameter trajectory (Fig. 15)



**Table 2** The evaluation of the mean value and variance of *NEE*, for different outlier density,  $\varepsilon = P(a(k) = 1)$

$\varepsilon$	0.01	0.05	0.10	0.15	0.20	0.25	0.3
NEE							
Mean value	-30.2529	-30.6593	-30.1427	-30.3464	-24.4299	-18.6086	-14.2801
Variance	92.9838	95.1626	111.2679	112.4948	226.8080	322.398	410.0782

desirable performance measure. Table 4 illustrates the influence of the initial condition  $\mathbf{P}(0)$  to the estimation error.

In general, large residual realisations in the initial steps, caused by large  $\|\mathbf{P}(0)\|$ , can make the decrease of  $\|\mathbf{P}(i)\|$  very slow. This, in turn, may result in an undesirable effect producing a slow convergence of filter parameter estimates. However, the proposed RRWLSV algorithm is found to be relatively insensitive to the initial conditions, due to the proper way of generating of the weighting matrix,  $\mathbf{P}(i)$ , in (12) and (13), where the weighted term,  $\omega(i)$  in (8), is used. This factor keeps the norm  $\|\mathbf{P}(i)\|$  at values enough high for obtaining good convergence and, at the same time, enough small for preventing the described undesirable convergence effect. Similarly as in the experiment 4.1, it is observed that for higher values of  $\|\mathbf{P}(0)\|$ , the RRLSS algorithm converges very slowly. The reason lies in the introduction of the first derivative,  $\psi'$ , instead of the weighted term  $\omega$  in (8), in the weighted matrix update equation,  $\mathbf{P}(i)$ , in (18), which is equal to zero in the saturation range. Thus, large residuals in the initial steps, caused by large  $\|\mathbf{P}(0)\|$ , make the decrease of  $\|\mathbf{P}(i)\|$  very slow. This, in turn, results in the described cumulative effect, producing a slow convergence of parameter estimates (see Table 4).

**4.6 Influence of model order**

The selection of model structure depends on the intended model application [31]. With the adopted FIR model structure, one has to select the order,  $n$ , of the parameter vector,  $\mathbf{W}$  in (3). The model order should not be selected too low, since then, all system dynamics cannot be described properly. However, it should not be selected too high either, since the higher the model order, the most parameters need to be estimated and the higher the variance of the parameter estimates is. Thus, increasing the model order beyond the true order of the system will not add to the quality of the model. Hence,

**Table 3** The evaluation of the mean value and variance of *NEE*, for different outlier intensity,  $\sigma_e^2 = \text{var}A(k)$ , where the nominal value is  $\sigma_0^2 = 10^4/12$

$\sigma_e^2$	$10^{-2}\sigma_0^2$	$10^{-1}\sigma_0^2$	$\sigma_0^2$	$10\sigma_0^2$	$10^2\sigma_0^2$
NEE					
mean value	-30.8406	-30.3820	-30.4177	-30.4417	-30.3870
variance	94.2645	92.8128	94.9259	95.1020	94.6418

the model order is some sort of compromise between fitting the data and model complexity [32, 33]. Figure 18 illustrates the discussion through the obtained *NEE* criterion in (36), for the experimental conditions described in 4.1, in the cases of lower ( $n = 7$ ), higher ( $n = 12$ ) and exact ( $n = 9$ ) model order of the FIR system.

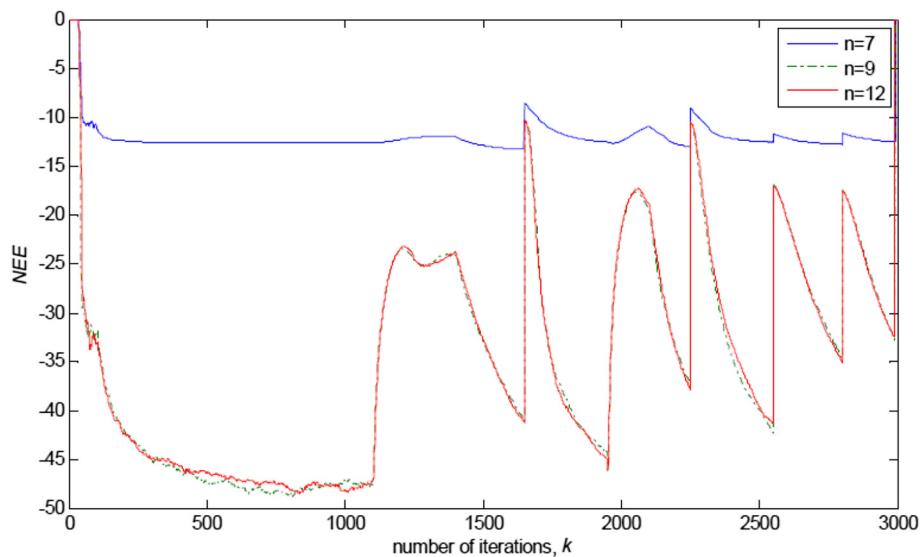
**4.7 Influence of additive noise correlations**

The FIR model structure does not take into account the coloured additive noise. However, in general, additive noise,  $n(k)$ , may represent any kind of noise, with arbitrary colour. Such noise can be simulated by filtering a zero-mean white noise through a linear time-invariant (LTI) system [34, 35]. In this way, one can shape the fit of a time-invariant model under a stationary noise environment to the frequency domain. It can be shown that, in general case, the resulting estimate is a compromise between fitting the estimated constant model parameters to the true one and fitting the noise model spectrum to the prediction error spectrum [35]. In addition, the estimates can be improved by using prefilters. The analysis is based on the limit parameter value to which the parameter estimates converge asymptotically and corresponding limit criterion [35]. However, such an analysis is not possible on the short data sequences, whether the situation is stationary or not. Therefore, the answer to the question concerning the insensitivity to noise statistics, as well as the other questions related to the practical robustness, can be obtained only by simulations. Figure 19 depicts the obtained *NEE* criterion in (36), with the experimental conditions exposed in 4.1, but using the coloured noise with the given autocorrelation function in Fig. 20, [36]. The experimental results indicate that the proposed RRWLSV filtering algorithm copes satisfactorily with the coloured noise.

It should be noted that in a FIR model structure, the noise model is fixed and equal to the unit constant.

**Table 4** The evaluation of the mean square error norm,  $\|\mathbf{P}(i)\|$ , for different initial conditions (experimental conditions are the same as in the experiment 4.1)

$\mathbf{P}(0)$	$10^2\mathbf{I}$	$10\mathbf{I}$	$\mathbf{I}$	$0.1\mathbf{I}$
Iterations				
50	366.9852	36.6985	0.3670	0.3530
100	0.9492	0.4582	1.1870	1.3091
1000	0.0100	0.0092	0.0085	0.0119

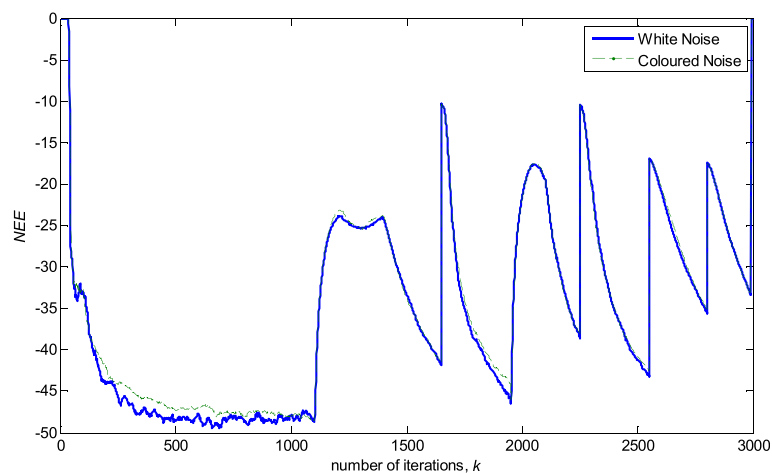


**Fig. 18** Normalised estimation error norm ( $NEE$ ) in (36), for the experimental conditions described in 4.1 (Figs. 2 and 3b) and the algorithm RRWLSV, in the cases of lower ( $n=7$ ), higher ( $n=12$ ) and exact ( $n=9$ ) model order

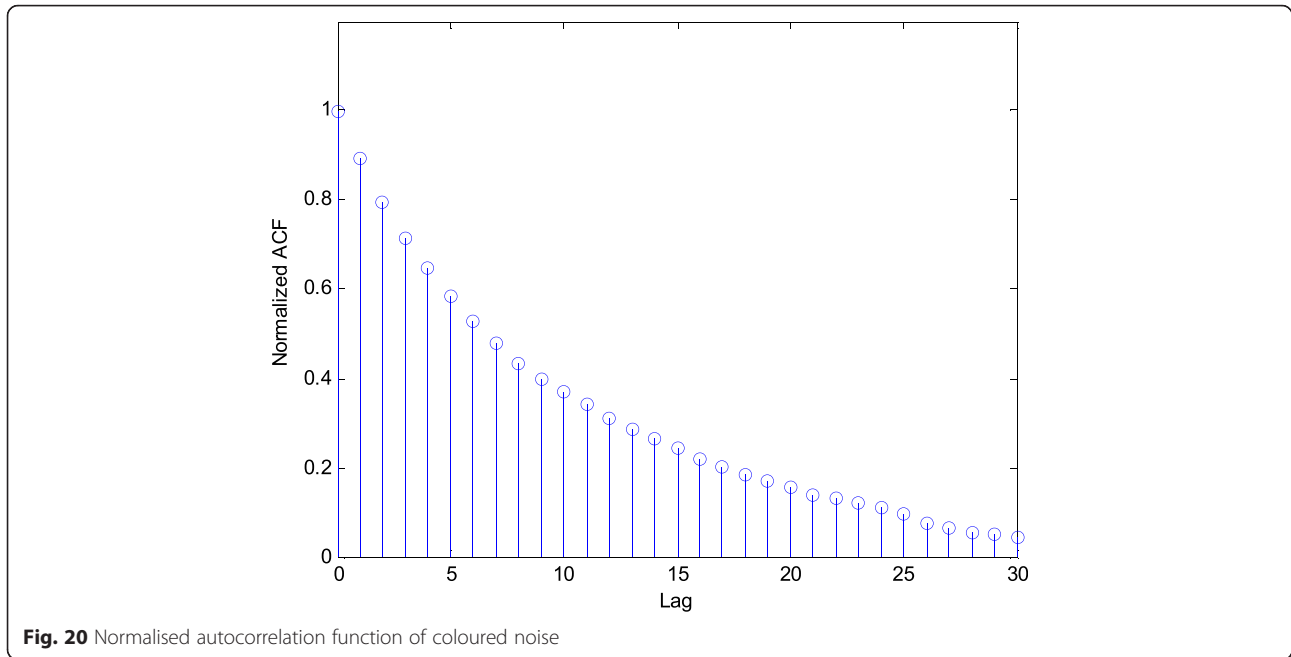
## 5 Conclusions

The estimation problem of time-varying adaptive FIR filter parameters in the situations characterised by non-stationary and impulsive noise environments has been discussed in the article. The posed problem is solved efficiently by application of a new adaptive robust algorithm, including a combination of the M robust concept, extended to the estimation of both filter parameters and unknown noise variance simultaneously, and adaptive robustified variable forgetting factor. The variable forgetting factor is determined by linear mapping of a suitably defined robust discrimination function, representing the ratio of robustified extended prediction error criterion, using M robust approach, and M robust type recursive estimate of noise variance. Since the robustified version

of extended prediction error criterion is calculated on sliding data frame of proper length, it represents a robust measure of local data non-stationarity. On the other hand, the total noise variance robust recursive estimate is rather insensitive to the local non-stationarity effects, so that the adopted robust discrimination function represents a suitable normalised robust measure of the degree of signal non-stationarity. In addition, since the total noise variance, or the so-called scale factor, and variable forgetting factor are adaptively calculated with respect to the prediction errors, the proposed algorithm works properly in stationary and non-stationary situations with possible outliers. Simulation results have shown that the new method gives higher accuracies of parameter estimates, and ensures better parameter tracking ability, in



**Fig. 19** Normalised estimation error norm ( $NEE$ ) in (36), for the experimental conditions in the case of white and coloured noise (see Fig. 20)



**Fig. 20** Normalised autocorrelation function of coloured noise

comparison to the conventional least-squares algorithm with variable forgetting factor, and the standard M robust algorithm with scale factor estimation. Moreover, the standard least-squares algorithm with variable forgetting factor is very sensitive to outliers, while the new adaptive robust method with combined scale and variable forgetting factor, and the conventional M robust-based method, with scale factor, are rather insensitive to such a disturbance. However, the proposed adaptive M robust algorithm with combined scale and forgetting factors has much better parameter tracking performance than the conventional M robust algorithm with only scale factor. The experimental analysis has shown that the real practical robustness and good tracking performances are connected with both the non-linear transformation of prediction-residuals and an adequate recursive generation of the weighted matrix, depending on the non-linearity form and variable forgetting factor. Moreover, a recursive estimation of the unknown noise variance is essential for defining properly the non-linearity form. In summary, in order to properly estimate the time-varying parameters under the non-stationary and impulsive noise, both robustified variable forgetting factor and simultaneous adaptive M robust estimation of system parameters and unknown noise variance are required.

The proposed adaptive M robust estimators for generating scale factor and variable forgetting factor are general, while the adaptive M robust parameter estimation procedure depends on assumed signal, or system, model structure. Moreover, it can be easily applied to the other commonly used signal, or system models, including AR, ARX, ARMA and ARMAX models, respectively, or even

non-linear models. Furthermore, the proposed adaptive robustified algorithm is the one that, with proper adaptation, can be used as a robust alternative to the conventional recursive least squares or Kalman filter, respectively. These applications arise in various fields, including speech processing, biomedical signal processing, image analysis, and failure detection in measurement and control.

## 6 Appendix 1

### 6.1 Derivation of relations (10)–(13).

The solution of (6) is given by

$$\hat{\mathbf{W}}(k) = \mathbf{R}^{-1}(k) \sum_{i=1}^k \beta(k, i) \mathbf{X}(i) d(i) \tag{38}$$

where

$$\beta(k, i) = \rho^{k-i} \omega(i) \tag{39}$$

with  $\omega(i)$  being defined by (8) when  $\mathbf{W}_0$  is replaced by  $\hat{\mathbf{W}}(k-1)$ . Moreover,  $e(i, \mathbf{W})$  in (2) is defined by (11) with  $\mathbf{W} = \hat{\mathbf{W}}(k-1)$ , while the matrix  $\mathbf{R}$  in (38) is given by

$$\begin{aligned} \mathbf{R}(k) &= \sum_{i=1}^k \beta(k, i) \mathbf{X}(i) \mathbf{X}^T(i) \\ &= \rho \sum_{i=1}^{k-1} \beta(k-1, i) \mathbf{X}(i) \mathbf{X}^T(i) \\ &\quad + \omega(k) \mathbf{X}(k) \mathbf{X}^T(k) \end{aligned} \tag{40}$$

from which it follows

$$\mathbf{R}(k) = \rho\mathbf{R}(k-1) + \omega(k)\mathbf{X}(k)\mathbf{X}^T(k) \quad (41)$$

Furthermore, one can also write

$$\sum_{i=1}^k \beta(k, i)\mathbf{X}(i)d(i) = \rho \sum_{i=1}^{k-1} \beta(k-1, i)\mathbf{X}(i)d(i) + \omega(k)\mathbf{X}(k)d(k) \quad (42)$$

Taking into account (38), one concludes from (42)

$$\sum_{i=1}^k \beta(k, i)\mathbf{X}(i)d(i) = \rho\mathbf{R}(k-1)\hat{\mathbf{W}}(k-1) + \omega(k)\mathbf{X}(k)d(k) \quad (43)$$

Furthermore, by calculating  $\rho\mathbf{R}(k-1)$  from (41), and replacing it into (43), one obtains

$$\sum_{i=1}^k \beta(k, i)\mathbf{X}(i)d(i) = [\mathbf{R}(k) - \omega(k)\mathbf{X}(k)\mathbf{X}^T(k)]\hat{\mathbf{W}}(k-1) + \omega(k)\mathbf{X}(k)d(k) \quad (44)$$

Finally, by replacing (44) into (38), one obtains the relations (10) and (11) with the gain matrix

$$\mathbf{K}(k) = \mathbf{R}^{-1}(k)\omega(k)\mathbf{X}(k) \quad (45)$$

In addition, by introducing the weighting matrix  $\mathbf{P}(k) = \mathbf{R}^{-1}(k)$ , it follows from (41)

$$\mathbf{P}(k) = [\rho\mathbf{P}^{-1}(k-1) + \omega(k)\mathbf{X}(k)\mathbf{X}^T(k)]^{-1} \quad (46)$$

After applying the well-known matrix inversion lemma [25, 26]

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1} \quad (47)$$

on (46) with  $\mathbf{A} = \rho\mathbf{P}^{-1}(k-1)$ ,  $\mathbf{B} = \mathbf{X}(k)$ ,  $\mathbf{C} = \omega(k)$  and  $\mathbf{D} = \mathbf{X}^T(k)$ , one concludes from (47)

$$\mathbf{P}(k) = \frac{1}{\rho} \left\{ \mathbf{P}(k-1) - \frac{\mathbf{P}(k-1)\mathbf{X}(k)\mathbf{X}^T(k)\mathbf{P}(k-1)}{\rho\omega^{-1}(k)\mathbf{X}^T(k)\mathbf{P}(k-1)\mathbf{X}(k)} \right\} \quad (48)$$

By substituting  $\mathbf{R}^{-1}(k) = \mathbf{P}(k)$  from (48) into (45), one can write further

$$\mathbf{K}(k) = \frac{\mathbf{P}(k-1)\mathbf{X}(k)\omega(k)}{\rho + \mathbf{X}^T(k)\mathbf{P}(k-1)\mathbf{X}(k)\omega(k)} \quad (49)$$

Finally, by introducing  $\mathbf{M}(k)$  from (12), and substituting (49) into (48), one obtains the relations (12) and (13), which completes the proof.

## 7 Appendix 2

### 7.1 Derivation of relations (24)–(25).

Starting from (21)–(23), one can write

$$kJ_k(s/\hat{\mathbf{W}}) = \ln(s) + (k-1)J_{k-1}(s/\hat{\mathbf{W}}) + f\left(\frac{e(k, \hat{\mathbf{W}})}{s}\right) \approx \ln(s) + f\left(\frac{e(k, \hat{\mathbf{W}})}{s}\right) \quad (50)$$

and

$$k \frac{\partial J_k(s/\hat{\mathbf{W}})}{\partial s} = \frac{1}{s} - \frac{e(k, \hat{\mathbf{W}})}{s^2} g\left(\frac{e(k, \hat{\mathbf{W}})}{s}\right) \quad (51)$$

where  $g(\cdot) = f'(\cdot)$ . Furthermore, one can also write from (19)

$$\frac{\partial J(\sigma/\hat{\mathbf{W}})}{\partial \sigma} = \frac{1}{\sigma} - E \left\{ \frac{e(k, \hat{\mathbf{W}})}{\sigma^2} g\left(\frac{e(k, \hat{\mathbf{W}})}{\sigma}\right) \right\} \quad (52)$$

and

$$\begin{aligned} \frac{\partial^2 J(\sigma/\hat{\mathbf{W}})}{\partial \sigma^2} &= -\frac{1}{\sigma^2} \\ &+ E \left\{ g' \left( \frac{e(k, \hat{\mathbf{W}})}{\sigma} \right) \frac{e^2(k, \hat{\mathbf{W}})}{\sigma^4} \right\} \\ &+ \frac{2}{\sigma} E \left\{ \frac{e(k, \hat{\mathbf{W}})}{\sigma^2} g \left( \frac{e(k, \hat{\mathbf{W}})}{\sigma} \right) \right\} \end{aligned} \quad (53)$$

In the vicinity of the optimal solution, one concludes  $\partial J(\sigma/\hat{\mathbf{W}})/\partial \sigma = 0$  in (52), from which it follows

$$E \left\{ \frac{e(k, \hat{\mathbf{W}})}{\sigma^2} g \left( \frac{e(k, \hat{\mathbf{W}})}{\sigma} \right) \right\} = \frac{1}{\sigma} \quad (54)$$

Furthermore, by substituting (54) in (53), one obtains

$$\frac{\partial^2 J_k(s/\hat{\mathbf{W}})}{\partial s^2} \approx \frac{\partial^2 J(s/\hat{\mathbf{W}})}{\partial \sigma^2} = \frac{a}{s^2} \quad (55)$$

where the constant

$$a = 1 + E \left\{ g' \left( \frac{e(k, \hat{\mathbf{W}})}{s} \right) \frac{e^2(k, \hat{\mathbf{W}})}{s^2} \right\} \approx 1 \quad (56)$$

Finally, if one substitutes (51), (55) and (56) into (20), one obtains an approximate optimal solution for the scale factor estimation in the recursive form (24). In addition, if we assume for the moment that  $s(k-1)$  is close to  $s(k)$ , and take into account the relation (8), the expression (24) can be rewritten in the alternative form (25), which completes the proof.

## 8 Appendix 3

### 8.1 The derivation of mean value and variance of the random variable generating the outliers

The outliers  $n(k)$  for each time step  $k$  are generated by the random variable (r.v.)  $n = \alpha A$ , representing the function of two variables, where  $\alpha$  is the discrete binary r.v. taking the values  $\alpha_1 = 0$  with the probability  $p_1 = P\{\alpha = \alpha_1\}$ , and  $\alpha_2 = 1$  with the probability  $p_2 = P\{\alpha = \alpha_2\}$ ,  $p_1 + p_2 = 1$ . Moreover,  $A$  is the continuous zero-mean r.v. with variance  $\sigma_A^2$ , and it is supposed that  $\alpha$  and  $A$  are independent. The mean  $E\{n\}$  can be determined directly in terms of the joint p.d.f.  $f_{\alpha A}(\alpha, A)$  without the need for evaluating p.d.f. of  $n$ , that is [35]

$$\begin{aligned} E\{n\} &= E\{\alpha A\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \alpha A f_{\alpha A}(\alpha, A) d\alpha dA \\ &= E\{\alpha\}E\{A\} = m_{\alpha}m_A \end{aligned} \quad (57)$$

since  $\alpha$  and  $A$  are independent, yielding  $f_{\alpha A}(\alpha, A) = f_{\alpha}(\alpha) f_A(A)$ . However, due to the fact that  $m_A = E\{A\} = 0$ , one concludes  $m_n = E\{n\} = 0$ . The variance of r.v.  $n$  is given by

$$\sigma_n^2 = E\{n^2\} - E^2\{n\} = m_{2,n} - m_n^2 \quad (58)$$

where  $m_{2,n}$  and  $m_n$  are non-centralized moments of the second and the first order, respectively. Similarly as before, the variance can be determined as

$$\sigma_n^2 = \sigma_{\alpha A}^2 = E\{(\alpha A)^2\} - E^2\{\alpha A\} \quad (59)$$

that is, since  $E\{\alpha A\} = 0$

$$\begin{aligned} \sigma_{\alpha A}^2 &= E\{\alpha^2 A^2\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \alpha^2 A^2 f_{\alpha A}(\alpha, A) d\alpha dA \\ &= m_{2,\alpha} m_{2,A} \end{aligned} \quad (60)$$

Here,  $m_{2,\alpha} = \sigma_{\alpha}^2$ , since  $m_A = 0$ , while

$$m_{2,\alpha} = \alpha_1^2 p_1 + \alpha_2^2 p_2 = 0 p_1 + 1 p_2 = p_2 \quad (61)$$

from which it follows

$$\sigma_n^2 = p_2 \sigma_A^2; p_2 = P\{\alpha = 1\} = \varepsilon \quad (62)$$

where  $\varepsilon$  is the density or fraction of outliers.

#### Abbreviations

EPE: extended prediction error; FIR: finite impulse response; FF: forgetting factor; FFF: fixed value of forgetting factor; LS: least squares; M: approximate maximum-likelihood; MAD: median of absolute median deviations; ML: maximum-likelihood;  $n$ : parameter vector order;  $N$ : memory length; NEE: normalised estimation error; PA: parallel adaptation; Q: discrimination function; RLS: recursive least squares; RLSVF: recursive least squares with variable forgetting factor; RRLS: robustified recursive least squares; RRLSS: robustified recursive least squares with scale factor; RRWLSV: robustified recursive weighted least squares with both scale and variable forgetting factors; SNR: signal-to-noise ratio; VFF: variable forgetting factor; WLS: weighted least-squares.

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgments

This research was supported by Serbian Ministry of Education and Science (Project TR 32038 and III 42007)

#### Author details

<sup>1</sup>School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, Belgrade, Serbia. <sup>2</sup>School of Electrical and Computer Engineering, 283 Vojvode Stepe St., Belgrade, Serbia. <sup>3</sup>Faculty of Informatics and Computing, Singidunum University, 32 Danijelova St., Belgrade, Serbia.

Received: 6 November 2015 Accepted: 23 March 2016

Published online: 31 March 2016

#### References

1. JV Candy, *Model-based signal processing* (John Wiley, New York, 2006)
2. SW Smith, *Digital signal processing: a practical guide for engineers and scientist* (Elsevier, Amsterdam, 2003)
3. SX Ding, *Model based fault diagnosis techniques, design schemes, algorithms, and tools* (Springer-Verlag, Berlin Heidelberg, 2008)
4. M Barkat, *Signal detection and estimation* (Artech House, Norwood, 2005)
5. M Verhaegen, V Verdult, *Filtering and system identification, a least squares approach* (Cambridge University Press, New York, 2012)
6. S Haykin, *Adaptive filter theory* (Prentice Hall, Pearson, 2013)
7. B Kovačević, Z Banjac, M Milosavljević, *Adaptive digital filters* (Springer-Verlag, Berlin, Heidelberg, 2013)
8. M Basseville, A Benveniste, *Detection of abrupt changes in signals and dynamical systems* (Springer-Verlag, Berlin, Heidelberg, 1986)
9. S Peters, DA Antoniou, A Parallel, Adaptation algorithm for recursive least squares adaptive filters in nonstationary environments. *IEEE Trans Signal Process* **43**(11), 2484–2494 (1995). doi:10.1109/78.482100
10. YS Cho, SB Kim, EJ Powers, Time-varying spectral estimation using AR models with variable forgetting factors. *IEEE Trans Signal Process* **39**(6), 1422–1426 (1991). doi:10.1109/78.136549
11. Z. Banjac, B. Kovačević, I. Kostić Kovačević, *Variable forgetting factor estimation in impulsive noise environment*, *Telecommunications Forum (TELFOR) 22nd, Belgrade, 25–27 November 2014*, pp. 449–452, doi:10.1109/TELFOR.2014.7034443
12. B. Kovačević, M. Milosavljević, M. Veinović, *Time-varying AR speech analysis using robust RLS algorithm with variable forgetting factor*, *Proceedings of the 12<sup>th</sup> IAPR International Conference on Pattern Recognition, (Jerusalem, 09–13 October 1994)*, vol 3, pp.211–213, doi:10.1109/ICPR.1994.577162
13. G Kvaščev, Ž Đurović, B Kovačević, I Kostić Kovačević, *Adaptive estimation of time-varying parameters in AR models with variable forgetting factor*, *Mediterranean Electrotechnical Conference MELECON, (Beirut, 13–16 April 2014)*, pp. 68–73, doi:10.1109/MELCON.2014.6820509
14. J Chambers, A Alvonitis, A Robust, Mixed norm adaptive filter algorithm. *IEEE Signal Process Lett* **4**(2), 46–48 (1997)
15. Y Zou, SC Chan, TS Ng, Robust M-estimate adaptive filtering. *IEE Proceedings Vis Image Signal Process* **148**(4), 289–294 (2001)
16. Y Zou, SC Chan, TS Ng, A recursive least M-estimate (RLM) adaptive filter for robust filtering in impulse noise. *Signal Process Lett IEEE* **7**(11), 975–991 (2000)
17. HS Kim, JS Lim, SJ Baek, KM Sung, Robust Kalman filtering with variable forgetting factor against impulsive noise. *IEICE Trans Fundam Electron Commun Comput Sci* **E84-A**(1), 363–366 (2001)
18. T Yang, JH Lee, KY Lee, KM Sung, On robust Kalman filtering with forgetting factor for sequential speech analysis. *Signal Process* **63**(2), 151–156 (1997)
19. Y.Zhou, S. C. Chan, K. L. Ho, A new variable forgetting factor QR based recursive least M-estimate algorithm for robust adaptive filtering in impulsive noise environment, *14<sup>th</sup> European Signal Processing Conference, (Florence, 4–8 September 2006)*, pp. 1–5 IEEE <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7071126&isnumber=7065146> ISSN: 2219-5491
20. ZG Zhang, SC Chan, Recursive parametric frequency/spectrum estimation for nonstationary signals with impulsive components using variable forgetting factor. *IEEE Trans Instrum Meas* **62**(12), 3251–3264 (2013)
21. PJ Huber, EM Ronchetti, *Robust statistics* (John Wiley, New Jersey, 2009)
22. RR Wilcox, *Introduction to robust estimation and hypothesis testing* (Academic Press, Boston, 2012)

23. WN Venables, BD Ripley, *Modern applied statistics with S* (Springer, Berlin Heidelberg New York, 2002)
24. Z Banjac, B Kovačević, *Robust parameter and scale factor estimation in nonstationary and impulsive noise environments*, in *Proceedings of IEEE Conf. EUROCON, (Belgrade, 21–24. November, 2005)*, pp. 1546–1549. doi:10.1109/EURCON.2005.1630261
25. L Ljung, T Soderstrom, *Theory and practice of recursive identification* (MIT Press, Cambridge, 1983)
26. YZ Tsyipkin, *Foundations of informational theory of identification* (Nauka, Moscow, 1984)
27. M Veinović, B Kovačević, M Milosavljević, Robust nonrecursive AR speech analysis. *Signal Process* **37**(2), 189–201 (1994)
28. B Kovačević, M Veinović, M Milosavljević, M Marković, *Robust speech processing* (Springer Verlag, Berlin, Heidelberg, 2016). in press
29. AK Kohli, A Rai, Numeric variable forgetting factor RLS algorithm for second-order Volterra filtering. *Circuits Syst Signal Process* **32**(1), 223–232 (2013)
30. K Ak, A Rai, M Patel, Variable forgetting factor LS algorithm for polynomial channel model. *ISRN Signal Process* **2011**, Article ID 915259 (2011)
31. PPJ van den Bosch, AC van der Klauw, *Modeling, identification and simulation of dynamical systems* (CRC Press, Boca Raton, 1994)
32. R Stoica, R Eykhoff, R Janssen, T Soderstrom, Model structure selection by cross validation. *Internatt J Control* **43**, 1841–1878 (1986)
33. H Akaike, Modern development of statistical methods, in *Trends and progress in system identification*, ed. by Eykhoff (Pergamon Press, New York, 1981), pp. 169–184
34. L Ljung, *System identification—theory for the user*, 2nd edn. (PTR Prentice Hall, Upper Saddle River, 1999)
35. B Kovačević, Ž Djurović, *Fundamentals of stochastic signals, systems and estimation theory: with worked examples* (Springer, Verlag, Berlin, 2008)
36. M Viswanathan, *Simulation of digital communication systems using Matlab*, 2013. <http://www.gaussianwaves.com/simulation-of-digital-communication-systems-using-matlab-ebook>. Accessed 01 march 2016

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---