INTERNATIONAL JOURNAL OF ROBUST AND NONLINEAR CONTROL
Int. J. Robust. Nonlinear Control 2012; **22**:1853–1872
Published online 13 September 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/rnc.1791

The filter design from data (FD2) problem: parametric-statistical approach

Carlo Novara[1,*,†], Mario Milanese[1], Eilyan Bitar[2] and Kameshwar Poolla[2]

[1]*Dipartimento di Automatica e Informatica, Politecnico di Torino, Torino, Italy*
[2]*Department of Mechanical Engineering, University of California at Berkeley, Berkeley, USA*

SUMMARY

A large body of literature exists on the filter design problem, assuming that the system to be filtered is known. However, in most practical situations, the system is not known, but a set of measured data is available. In such situations, a two-step procedure is typically adopted: a model is identified from this data set, and a filter is designed based on the identified model. In this paper, we consider an alternative approach, which uses the available data not for the identification of a model, but for the direct design of the filter. Such a direct design is investigated within a parametric-statistical framework for both the cases of linear time-invariant and nonlinear systems. The noise is assumed to be stochastic, and optimality refers to minimizing the estimation error variance. It is shown that the direct design has superior features with respect to the two-step design, especially in the presence of modeling errors. Another relevant advantage of the direct design over the two-step procedure is that minimum variance (Kalman) filters for nonlinear systems are, in general, difficult to derive and/or to implement. On the contrary, the direct approach allows for a very efficient filter design. To demonstrate the effectiveness of the proposed direct design, two examples are presented: the first is related to estimation of the Lorentz chaotic attractor; the second, involving real data, is related to estimation of vehicle yaw rate. Copyright © 2011 John Wiley & Sons, Ltd.

Received 19 November 2010; Revised 27 May 2011; Accepted 27 June 2011

KEY WORDS: direct filter design; nonlinear systems; parametric–statistical estimation

1. INTRODUCTION

Consider a nonlinear discrete-time system S, described in state-space form:

$$\begin{aligned} x^{t+1} &= F\left(x^t, \widetilde{u}^t\right) + w_x^t \\ \widetilde{y}^t &= H_y\left(x^t, \widetilde{u}^t\right) + w_y^t \\ \widetilde{z}^t &= H_z\left(x^t, \widetilde{u}^t\right) + w_z^t \end{aligned} \tag{1}$$

where $x^t \in X \subseteq \mathbb{R}^{n_x}$ is the state, $\widetilde{u}^t \in U \subseteq \mathbb{R}^{n_u}$ is the known input, $\widetilde{y}^t \in Y \subseteq \mathbb{R}^{n_y}$ is a measured output, $\widetilde{z}^t \in Z \subseteq \mathbb{R}^{n_z}$ is the variable to estimate, w_x^t is the process noise, w_y^t and w_z^t are output noises, the functions F, H_y and H_z are differentiable on $X \times U$.

The problem is to design a filter that, operating on $\widetilde{u}^\tau$ and $\widetilde{y}^\tau$, $\tau \leqslant t$, gives an (possibly optimal in some sense) estimate of the variable $\widetilde{z}^t$.

Optimal solutions to this problem have been derived in the case where the functions F, H_y and H_z are linear, under different assumptions on noise and optimality criteria. In the case of stochastic noises, the Kalman filter minimizes the estimation error variance [1–4]. In the case where the noise and the variable $\widetilde{z}^t$ belong to normed spaces, the H_∞ filter minimizes the induced norm from ℓ_2 to ℓ_2, the H_2 filter minimizes the induced norm from ℓ_2 to ℓ_∞, the ℓ_1 filter minimizes the induced

*Correspondence to: Carlo Novara, Dipartimento di Automatica e Informatica, Politecnico di Torino, Italy.
†E-mail: carlo.novara@polito.it

norm from ℓ_∞ to ℓ_∞ [5–8]. In the case of nonlinear functions F, H_y and H_z, optimal filters are, in general, hard to derive, and the usual approach is to obtain approximate solutions such as extended Kalman filters [1–3], unscented Kalman filters [9], ensemble filters [10], particle filters [11–13].

In all these works, the equations (1) describing the system to be filtered are assumed to be known. However, in most practical situations, the system equations are not known, and a two-step procedure is adopted: (1) a model of S is identified from data; and (2) a filter is designed from the identified model. Note that, except for particular cases where H_z is actually known, measurements not only of $\widetilde{y}^t$ but also of $\widetilde{z}^t$ are needed in step 1 to have a sufficiently informative data set for model identification. Then, not a problem of filter design from known system has to be solved, but a problem of *filter design from data* (FD2). Note that the latter problem is more general than the former one.

As previously noted, the usual approach to solve the FD2 problem is a two-step procedure, based on model identification from data and filter design from the identified model. However, this procedure is, in general, far from being optimal, as a result of the following reasons: (1) only an approximate model can be identified from measured data, and a filter which is optimal for the identified model may display a very large estimation error when applied to the real system; and (2) in the case of nonlinear system, designing a computationally tractable optimal filter is, in general, very difficult, and most of the times only approximate filters can be derived, whose stability is not even guaranteed. Evaluating how these two sources of approximation affect the filter estimation accuracy is a largely open problem. Note that robust filtering does not provide, at present, an efficient solution to the FD2 problem. Indeed, the design of a robust filter is based on the knowledge of an uncertainty model, for example a nominal model plus a description of the parametric uncertainty. However, identifying reliable uncertainty models from experimental data is an open problem, especially for nonlinear systems. Moreover, in the case of nonlinear systems, designing a computationally tractable robust filter is, in general, hard (see [14, 15]) and only approximate filters are commonly used [16, 17].

In this paper, an alternative approach is developed, which overcomes all these issues. This approach uses measured data not for the derivation of a model, but for the *direct design* of the filter. Indeed, the desired solution of the FD2 problem is a causal filter mapping $(\widetilde{u}^\tau, \widetilde{y}^\tau) \to \widehat{z}^t$, $\tau \leqslant t$, producing as output an estimate $\widehat{z}^t$ of $\widetilde{z}^t$, enjoying some optimality property of the estimation error $\widetilde{z}^t - \widehat{z}^t$. Thus, the idea is to directly design a filter from the available data, via identification of a filter that, using the inputs $(\widetilde{u}^t, \widetilde{y}^t)$ gives an output $\widehat{z}^t$, which minimizes the desired criterion for evaluating the estimation error $\widetilde{z}^t - \widehat{z}^t$. Such a filter is indicated as *Direct Virtual Sensor* (DVS). The direct approach thus represents a paradigm shift in filter design, which allows us to design optimal filters even for nonlinear systems, overcoming critical problems such as model uncertainty and nonlinear filter approximation.

The direct approach has been developed within a Set Membership framework for linear time-invariant (LTI) systems [18], linear parameter-varying (LPV) systems [19], and nonlinear systems [20]. Practical applications can be found in [21–25].

In the present paper, the direct approach is developed within a parametric-statistical framework for both the cases of LTI and nonlinear systems. The noises w_x^t and w_y^t, are assumed to be stochastic, a parametric filter structure and a Prediction Error (PE) method [26] are used for DVS design, and an optimality notion related to minimizing the estimation error variance is considered. In both the cases of LTI and nonlinear systems, an optimal DVS is derived and compared with the filter obtained by means of the standard two-step procedure, based on model identification and minimum variance filter design. It is shown that even in the most favorable situations for the two-step procedure, that is when the system S belongs to the model structure selected for identification and the minimum variance filter can be actually designed for the identified model, the DVS has estimation error variance no greater than the two-step filter. Moreover, in the presence of modeling errors, the DVS, although not absolutely optimal, is the minimum variance estimator among the selected filter class. A similar result is not assured for the two-step filter, whose performance deterioration caused by the modeling errors may be significantly larger.

---

†A preliminary version of this paper has been presented at the American Control Conference 2006.

DOI: 10.1002/rnc

It must be remarked that the contribution of this paper is not related to system identification but to filter design, and that the main result obtained is showing that the direct approach is more effective than the two-step approach, in both the linear and nonlinear cases. It must also be noted that, whereas here the direct approach is developed within a parametric-statistical framework, in [18], [19], and [20], the direct approach has been developed within a Set Membership framework, where the noises have been assumed unknown but bounded, and no parametric filter structures have been used. Moreover, in [18, 19], and [20], no comparison with the two-step approach have been carried out.

Two examples are presented to demonstrate the effectiveness of the proposed direct approach: the first, involving simulated data, is related to estimation of the Lorenz chaotic attractor; the second, involving real data, is related to estimation of vehicle yaw rate.

2. OBSERVABILITY NOTIONS AND PRELIMINARY RESULTS

In this section, some notations and observability notions used in the paper are introduced. Moreover, some preliminary results basic for the derivation of the optimal DVSs are presented.

First, the case where the noises w_x^t and w_y^t in Equation (1) are null is considered. The input and the output of the system (1) corresponding to null noises are denoted by u^t and y^t, respectively.

Several definitions of observability can be found in the literature, see for example [27–31]. In this paper, we use a notion of observability, similar to those considered in [29, 30], and to the notion of state reconstruction in [32, 33]. Let us introduce the following notation to indicate function composition:

$$F^k(x) \doteq \underbrace{F(F(\ldots F}_{k \text{ times}}(x)\ldots)), \quad H_y F(x) \doteq H_y(F(x)).$$

Let us define the map

$$\Omega\left(x^t, \mathbf{u}^t, r\right) \doteq \begin{bmatrix} H_y\left(x^t, u^t\right) \\ H_y F\left(x^t, u^t, u^{t-1}\right) \\ \vdots \\ H_y F^{r-1}\left(x^t, u^t, u^{t-1}, .., u^{t-r+1}\right) \end{bmatrix}.$$

Note that $\Omega(\cdot,\cdot,r) : \mathbb{R}^{n_x+rn_u} \to \mathbb{R}^{rn_y}$ and that, for given $\mathbf{u} \in \mathbb{R}^{rn_u}$, $\Omega(\cdot,\mathbf{u},r) : \mathbb{R}^n \to \mathbb{R}^{rn_y}$. The i-th component of Ω is denoted by Ω_i.

Definition 1

The couple (F, H_y) is observable if an integer $r \in [1, n_x]$ and a set of indices $M_{n_x} = \{i_1, i_2, \ldots, i_{n_x}\}$ exist such that, for any $\mathbf{u} \in U^r \subseteq \mathbb{R}^{rn_u}$, the map $\Omega(\cdot, \mathbf{u}, M_{n_x}) : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$, defined as

$$\Omega\left(x, \mathbf{u}, M_{n_x}\right) \doteq \begin{bmatrix} \Omega_{i_1}(x, \mathbf{u}, r) \\ \Omega_{i_2}(x, \mathbf{u}, r) \\ \vdots \\ \Omega_{i_{n_x}}(x, \mathbf{u}, r) \end{bmatrix}$$

is a diffeomorphism with respect to x.

Observability as defined earlier implies that the state can be uniquely determined from a finite number of input–output values, as stated by the following result. Suppose that an input sequence $u^0, u^1, \ldots, u^{r-1}$ is applied to system (1) starting from an initial condition x^0, and that the same input sequence is applied to system (1) starting from an initial condition $\widehat{x}^0$. Let the corresponding output sequences be $y^0, y^1, \ldots, y^{r-1}$ and $\widehat{y}^0, \widehat{y}^1, \ldots, \widehat{y}^{r-1}$, respectively.

DOI: 10.1002/rnc

Lemma 1
If (F, H_y) is observable then, for any initial conditions x^0 and $\widehat{x}^0 \neq x^0$ and any input sequence u^0, $u^1, \ldots, u^{r-1}$, it holds that $\mathbf{y}^{r-1} \neq \widehat{\mathbf{y}}^{r-1}$, where $\mathbf{y}^{r-1} = (y^{r-1}, \ldots, y^0)$ and $\widehat{\mathbf{y}}^{r-1} = (\widehat{y}^{r-1}, \ldots, \widehat{y}^0)$.

Proof
Consider that

$$\begin{gathered} y^0 = H_y\left(x^0, u^0\right) \\ y^1 = H_y\left(x^1, u^1\right) = H_y F\left(x^0, u^1, u^0\right) \\ \vdots \\ y^{r-1} = H_y F^{r-1}\left(x^0, u^{r-1}, .., u^0\right). \end{gathered}$$

This can be written as $\mathbf{y}^{r-1} = \Omega\left(x^0, \mathbf{u}^{r-1}, r\right)$, where $\mathbf{u}^{r-1} = (u^{r-1}, \ldots, u^0)$. Repeating this argument for the initial condition $\widehat{x}^0$, we obtain $\widehat{\mathbf{y}}^{r-1} = \Omega\left(\widehat{x}^0, \mathbf{u}^{r-1}, r\right)$. Because (F, H_y) is observable, then there exists a set of indices $M_n = \{i_1, i_2, \ldots, i_n\}$ such that the map $\Omega(x, \mathbf{u}, M_n)$ is a diffeomorphism. Because a diffeomorphism is a one-to-one map, if $\widehat{x}^0 \neq x^0$ then $\Omega\left(\widehat{x}^0, \mathbf{u}, M_n\right) \neq \Omega\left(x^0, \mathbf{u}, M_n\right)$. This implies that $\Omega\left(\widehat{x}^0, \mathbf{u}, r\right) \neq \Omega\left(x^0, \mathbf{u}, r\right)$ and thus $\widehat{\mathbf{y}}^{r-1} \neq \mathbf{y}^{r-1}$. □

Now, suppose that the noises w_x^t and w_y^t in Equation (1) are not null. A result is presented, which is essential in the next sections to derive the optimal DVSs.

Lemma 2
If (F, H_y) is observable, then a function f_z exists such that

$$\widetilde{z}^t = f_z\left(\widetilde{y}^t, \ldots, \widetilde{y}^{t-n_x+1}, \widetilde{u}^t, \ldots, \widetilde{u}^{t-n_x+1}, w^t, \ldots, w^{t-n_x+1}\right) \tag{2}$$

where $w^t = (w_x^t, w_y^t)$.

Proof
The proof can be obtained by minor modifications of the proof of Lemma 1 in [20]. □

Note that if the system (1) is linear, then the function f_z in Equation (2) is linear.

Remark 1
Verifying the observability of a nonlinear system is an open problem in the literature. However, Lemma 2, together with the validation analysis of [34], may give indications on observability verification from a finite set of measured data $\widetilde{u}^t, \widetilde{y}^t, \widetilde{z}^t$, $t = 1, 2, \ldots, T$. Indeed, according to Lemma 2, the existence of the function f_z implies the system observability. On the other hand, the validation analysis of [34] provides necessary and sufficient conditions for the existence of a function consistent with the measured data and prior assumptions. These results may be used together for the verification of the system observability. It must be noted that this verification method is still at a preliminary stage and is the subject of current and future research.

3. THE FILTER DESIGN FROM DATA PROBLEM: DIRECT VERSUS TWO-STEP APPROACH

In this section, the two-step approach to the FD2 problem is first described. Next, the direct filter design approach is proposed for the considered parametric-statistical framework. Then, a theoretical comparison is performed between the two approaches for both the cases of LTI and nonlinear system S.

Basic assumptions

- *The functions F, H_y and H_z in Equation (1), defining the system S to be filtered, are not known.*
- *(F, H_y) is observable.*

- *A set of data* $\{\widetilde{u}^t, \widetilde{y}^t, \widetilde{z}^t, t = 1, 2, \ldots, T\}$ *is available.*
- *The noises* w_x^t, w_y^t *and* w_z^t *are unmeasured stochastic variables.*
- *Let* $\bar{E}v^t \doteq \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} Ev^t$, *where* E *is the mean value (or expectation), and it is assumed that the limit exists whenever the symbol* $\bar{E}$ *is used.*

Under these assumptions, the filter design problem can be formulated as follows:

Filter design problem: *Design a causal filter that, operating on* $(\widetilde{u}^\tau, \widetilde{y}^\tau)$, $\tau \leqslant t$, *gives an estimate* $\widehat{z}^t$ *of* $\widetilde{z}^t$, *having minimum estimation error variance* $\bar{E}\left\|\widetilde{z}^t - \widehat{z}^t\right\|^2$ *for any* t.

The two-step design consists in model identification from data and filter design from the identified model. In the model identification step, a parametric model structure

$$M(\theta_M) : \theta_M \in \Theta_M$$

is selected, where Θ_M is a compact subset of $\mathbb{R}^{n_{\theta M}}$ and $n_{\theta M}$ is the number of parameters of the model structure. This model structure defines the following model set:

$$\mathcal{M} \doteq \{M(\theta_M) : \theta_M \in \Theta_M\}.$$

Then, a model $\widehat{M}$ of the system S is identified from the data set

$$D_M \doteq \{\widetilde{u}^t, \left(\widetilde{y}^t, \widetilde{z}^t\right), t = 1, 2, \ldots, T\}.$$

Note that, in the data set D_M, $\widetilde{u}^t$ is considered as the input of the model $M(\theta_M)$ and $\left(\widetilde{y}^t, \widetilde{z}^t\right)$ as its outputs. The PE method [26] is used for the identification of the model $\widehat{M}$, obtained as

$$\widehat{M} = M(\widehat{\theta}_M)$$
$$\widehat{\theta}_M = \arg\min_{\theta_M \in \Theta_M} J_T(\theta_M)$$
$$J_T(\theta_M) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{2}\left\|e^t(\theta_M)\right\|^2$$

where $e^t(\theta_M) = \left(\widetilde{y}^t, \widetilde{z}^t\right) - \left(y_M^t, \widehat{z}_M^t\right)$ is the PE of the model $M(\theta_M)$, being $\left(y_M^t, \widehat{z}_M^t\right)$ the prediction given by $M(\theta_M)$, and $\|\cdot\|$ is the ℓ_2 norm.

In the filter design step, a (steady-state) minimum variance filter

$$\widehat{K} \equiv K(\widehat{\theta}_M)$$

is designed to estimate $\widetilde{z}^t$ on the basis of the identified model $\widehat{M} = M(\widehat{\theta}_M)$. The filter $\widehat{K}$ gives as output an estimate $\widehat{z}_K^t$ of $\widetilde{z}^t$, using measurements $(\widetilde{u}^\tau, \widetilde{y}^\tau)$, $\tau \leqslant t$, thus providing a *Model-based Virtual Sensors*. Note that the filter structure has not been chosen in the two-step procedure. It just depends on the structure of the identified model.

In this paper, we propose an alternative approach to the FD2 problem, based on the direct identification of the filter from data. In such a direct approach, instead of selecting a parametric structure for the model $M(\theta_M)$ as done in the two-step procedure, a parametric structure

$$V(\theta_V) : \theta_V \in \Theta_V$$

is selected for the filter to be designed, where, Θ_V is a compact subset of $\mathbb{R}^{n_{\theta V}}$ and $n_{\theta V}$ is the number of parameters of the filter structure. This filter structure defines the following filter set:

$$\mathcal{V} \doteq \{V(\theta_V) : \theta_V \in \Theta_V\}.$$

A filter $\widehat{V}$ is then identified by means of the PE method from the data set

$$D_V \doteq \{\left(\widetilde{u}^t, \widetilde{y}^t\right), \widetilde{z}^t, t = 1, 2, \ldots, T\}.$$

DOI: 10.1002/rnc

Note that the set D_V is different from the set D_M, although both use the same data. In fact, in the data set D_V, $\left(\widetilde{u}^t, \widetilde{y}^t\right)$ are considered as the inputs of the filter $V(\theta_V)$ and $\widetilde{z}^t$ as its output. Thus, $\widehat{V}$ is obtained by means of the PE method as

$$\begin{aligned} \widehat{V} &= V(\widehat{\theta}_V) \\ \widehat{\theta}_V &= \arg \min_{\theta_V \in \Theta_V} J_T(\theta_V) \\ J_T(\theta_V) &= \frac{1}{T} \sum_{t=1}^{T} \frac{1}{2} \left\| e^t(\theta_V) \right\|^2 \end{aligned} \tag{3}$$

where $e^t(\theta_V) = \widetilde{z}^t - \widehat{z}_V^t$ is the estimation error of the filter $V(\theta_V)$, which has input $(\widetilde{u}^t, \widetilde{y}^t)$ and output $\widehat{z}_V^t$. The filter $\widehat{V} = V(\widehat{\theta}_V)$ can be used as a virtual sensor to generate an estimate $\widehat{z}_V^t$ of $\widetilde{z}^t$ from measurements $(\widetilde{u}^\tau, \widetilde{y}^\tau)$, $\tau \leqslant t$. Thus, $\widehat{V}$ is a DVS, designed directly from data without identifying a model of the system S.

The direct and two-step filter design approaches are now compared for both the cases of linear and nonlinear system S.

3.1. Linear case

In this section, the case of linear system (1) is investigated. The considered assumptions are first described, the main result is then presented, and this result is finally discussed.

Linear case assumptions:

- *The system S described by Equation (1) is linear.*
- *The signals $\widetilde{u}^t$ and $\widetilde{y}^t$ are bounded.*
- *The noises w_x^t, w_y^t and w_z^t are i.i.d. stochastic variables with zero mean and bounded moments of order $4+\delta$, for some $\delta > 0$.*
- *In the two-step approach, a uniformly stable linear model structure $M(\theta_M)$ is selected in the identification phase. Assuming that the identified model $\widehat{M}$ is observable from the output y^t, the filter $\widehat{K}$ is the (linear steady-state) Kalman filter designed to estimate $\widetilde{z}^t$ on the basis of the model $\widehat{M}$.*
- *In the direct approach, a uniformly stable linear filter structure $V(\theta_V)$ is selected.*

Note that, for the given linear observable model $M(\theta_M)$ of order n_M, the corresponding Kalman filter $K(\theta_M)$ is a linear stable filter of order n_M. Thus, if a filter structure $V(\theta_V)$ of order n_M is selected, it results that $K(\theta_M) \in \mathcal{V}$. The following theorem can now be presented:

Theorem 1

The following results hold with probability (w.p.) 1 as $T \to \infty$:

(i) $\widehat{V} = \arg\min_{V(\theta_V)} \bar{E} \left\| \widetilde{z}^t - \widehat{z}_V^t \right\|^2$.

(ii) If $\widehat{K} \in \mathcal{V}$, then $\bar{E} \left\| \widetilde{z}^t - \widehat{z}_V^t \right\|^2 \leqslant \bar{E} \left\| \widetilde{z}^t - \widehat{z}_K^t \right\|^2$.

(iii) If $S = M(\theta_M^o) \in \mathcal{M}$ and $K(\theta_M^o) \in \mathcal{V}$, then $\widehat{V}$ is a minimum variance filter among all linear causal filters mapping $(\widetilde{u}^\tau, \widetilde{y}^\tau) \to \widetilde{z}^t$, $\tau \leqslant t$.

(iv) If $S = M(\theta_M^o) \in \mathcal{M}$, $K(\theta_M^o) \in \mathcal{V}$, $M(\theta_M)$ is globally identifiable, S is stable, and the data are informative enough, then $\bar{E} \left\| \widetilde{z}^t - \widehat{z}_V^t \right\|^2 = \bar{E} \left\| \widetilde{z}^t - \widehat{z}_K^t \right\|^2$.

Proof

The filter $\widehat{V}$ is obtained from a PE method using the data set D_V, the uniformly stable filter structure $V(\theta_V)$, and a quadratic criterion. Lemma 8.2 of [26] can be applied, provided that the data set D_V is subject to condition D1 (see the Appendix). This condition requires that the transfer functions from $\widetilde{u}^t, \widetilde{y}^t, w_x^t, w_y^t$ and w_z^t to $\widetilde{z}^t$ are stable (i.e. have impulse responses with bounded ℓ_1 norm), and that the signals $\widetilde{u}^t$, $\widetilde{y}^t$, $\widetilde{z}^t$ are jointly quasi-stationary. From Lemma 2, where in the present case f_z is linear, it follows that the preceding transfer functions have finite impulse responses and consequently are stable. Joint quasi-stationarity of $\widetilde{u}^t$, $\widetilde{y}^t$, $\widetilde{z}^t$ follows from this stability result and Theorem 2.2 in [26].

All assumptions of Lemma 8.2 of [26] are thus verified (see the Appendix), giving

$$\sup_{\theta_V \in \Theta_V} |J_T(\theta_V) - \bar{E}\frac{1}{2}\left\|e^t(\theta_V)\right\|^2| \to 0, \quad w.p.1 \quad as \quad T \to \infty.$$

The cost function $J_T(\theta_V)$ thus converges to the variance of the estimation error (divided by 2). Because the convergence is uniform on Θ_V, the minimizer $\widehat{\theta}_V$ of $J_T(\theta_V)$ also converges to a minimizer (not necessarily unique) of the estimation error variance. This implies that

$$\widehat{V} = \arg\min_{V(\theta_V)} \bar{E}\left\|\widetilde{z}^t - \widehat{z}_V^t\right\|^2, \quad w.p.1 \quad as \quad T \to \infty, \tag{4}$$

proving (i).

Claim (ii) is also proven, because $\widehat{K}$ is not ensured to have the same property.

Under the assumptions of (iii), the Kalman filter $K(\theta_M^o)$ is the minimum variance estimator of $\widetilde{z}^t$, among all linear estimators making use of the measurements $\widetilde{y}^t$ and $\widetilde{u}^t$. Because $K(\theta_M^o) \in \mathcal{V}$, then $\exists\theta_V^o$ such that $K(\theta_M^o) = V(\theta_V^o)$. Thus, in the filter set $\mathcal{V}$ there exists at least one element giving the minimum variance estimate of $\widetilde{z}^t$, among all linear estimators. This fact, together with Equation (4), proves (iii).

Under the assumptions of (iv), from Theorems 8.2 and 8.3 (see [26] and the Appendix), it follows that $\widehat{\theta}_M \to \theta_M^o$, $w.p.1$ as $T \to \infty$, that is $\hat{K} \to K(\theta_M^o)$, $w.p.1$ as $T \to \infty$. Because the Kalman filter $K(\theta_M^o)$ is the minimum variance linear causal estimator of $\widetilde{z}^t$, claim (iv) follows from (iii). □

This result shows that the solution of the FD2 problem provided by the direct procedure presents better features than the one provided by the two-step procedure, even in the present linear case where the optimal minimum variance (Kalman) filter design required by the two-step procedure can be performed. Indeed, at best (e.g. under the assumption $S \in \mathcal{M}$, i.e. no undermodeling), the filter $\widehat{K}$ is proven to be asymptotically optimal provided that the system S is stable, whereas the DVS $\widehat{V}$ gives minimum variance estimation error, even in the case that the system S is unstable.

Even more favorable features of the direct approach over the two-step procedure are obtained in the more realistic situation that $S \notin \mathcal{M}$, because, in general, only approximate model structures are used. For example, consider that the system S is of order n_x (not known), and a model structure of order $n_M < n_x$ is selected. Then, it is not ensured that the corresponding Kalman filter $\widehat{K}$ gives the minimal variance estimate of $\widetilde{z}^t$ among all causal filters of the same order n_M. On the contrary, such an optimality feature holds for the DVS $\widehat{V}$ designed by selecting a filter structure of order $n_V = n_M < n_x$. Indeed, the accuracy deterioration of the Model-based Virtual Sensors $\widehat{K}$ with respect to the DVS $\widehat{V}$ of the same order may be significant, see for example [24], [23].

3.2. *Nonlinear case*

In this section, the case of nonlinear system (1) is investigated. The considered assumptions are first described, the main result is then presented, and this result is finally discussed.

Nonlinear case assumptions:

- *The system S described by Equation (1) is nonlinear.*
- *The signals $\widetilde{u}^t$ and $\widetilde{y}^t$ are bounded.*
- *The noises w_x^t, w_y^t and w_z^t are i.i.d. stochastic variables with zero mean.*
- *In the two-step approach, a nonlinear parametric model structure $M(\theta_M)$ is selected, satisfying condition M1 of* [35] *(see the Appendix). The corresponding filter $\widehat{K}$ is the minimum variance (nonlinear) filter for the identified model $\widehat{M}$.*
- *In the direct approach, a parametric nonlinear regression form for the filter is considered:*

DOI: 10.1002/rnc

$$\widehat{z}_V^t = f_V(\theta_V, \widehat{z}_V^{t-1}, \ldots, \widehat{z}_V^{t-n_V}, \widetilde{y}^t, \ldots, \widetilde{y}^{t-n_V}, \widetilde{u}^t, \ldots, \widetilde{u}^{t-n_V}). \quad (5)$$

The related filter structure $V(\theta_V)$ is assumed to satisfy condition M1 of [35] *(see the Appendix).*

Condition M1 essentially means that the filter structure has uniform exponential fading memory, that is, the remote past inputs are forgotten at an exponential rate. Note that the minimum variance filter $\widehat{K}$ can be described as well by a fading memory regression equation $\widehat{z}_K^t = f_K(\widehat{z}_K^{t-1}, \ldots, \widehat{z}_K^{t-n_V}, \widetilde{y}^t, \ldots, \widetilde{y}^{t-n_V}, \widetilde{u}^t, \ldots, \widetilde{u}^{t-n_V})$ because, under quite general assumptions, fading memory is necessary for a filter to give a bounded estimation error. The DVS $\widehat{V}$ is described by

$$\widehat{z}_V^t = f_V(\widehat{\theta}_V, \widehat{z}_V^{t-1}, \ldots, \widehat{z}_V^{t-n_V}, \widetilde{y}^t, \ldots, \widetilde{y}^{t-n_V}, \widetilde{u}^t, \ldots, \widetilde{u}^{t-n_V}).$$

where $\widehat{\theta}_V$ is obtained from the optimization problem (3). Now, the following extension of Theorem 1 to the case where the system S is nonlinear is given.

Theorem 2

The following results hold w.p. 1 as $T \to \infty$:

(i) $\widehat{V} = \arg\min_{V(\theta_V)} \bar{E} \left\| \widetilde{z}^t - \widehat{z}_V^t \right\|^2$.

(ii) If $\widehat{K} \in \mathcal{V}$, then $\bar{E} \left\| \widetilde{z}^t - \widehat{z}_V^t \right\|^2 \leqslant \bar{E} \left\| \widetilde{z}^t - \widehat{z}_K^t \right\|^2$.

(iii) If $S = M(\theta_M^o) \in \mathcal{M}$ and $K(\theta_M^o) \in \mathcal{V}$, then $\widehat{V}$ is a minimum variance filter.

Proof

The filter $\widehat{V}$ is obtained from a PE method using the data set D_V, the filter structure $V(\theta_V)$, and a quadratic criterion. Lemma 3.1 of [35] can be applied, provided that the data set D_V is subject to condition S3 of [35] (see the Appendix). This condition requires that the mapping from $\widetilde{u}^t$, $\widetilde{y}^t$, w_x^t and w_y^t to $\widetilde{z}^t$ has exponential fading memory. In the present case, condition S3 holds, because from Lemma 2, it follows that at any time t, $\widetilde{z}^t$ depends only on a finite number of past values of $\widetilde{u}^t$, $\widetilde{y}^t$, w_x^t and w_y^t.

Then, Lemma 3.1 of [35] (see the Appendix), giving

$$\sup_{\theta_V \in \Theta_V} |J_T(\theta_V) - \bar{E}\frac{1}{2} \left\| e^t(\theta_V) \right\|^2 | \to 0, \quad w.p.1 \quad as \quad T \to \infty.$$

The cost function $J_T(\theta_V)$ thus converges to the variance of the estimation error (divided by 2). Because the convergence is uniform on Θ_V, the minimizer $\widehat{\theta}_V$ of $J_T(\theta_V)$ also converges to a minimizer (not necessarily unique) of the estimation error variance, thus proving (i).

Claim (ii) immediately follows from (i).

In order to prove (iii), note that if $S = M(\theta_M^o)$, then the filter $K(\theta_M^o)$ is the minimum variance estimator of $\widetilde{z}^t$, among all estimators making use of measurements $\widetilde{u}^t$ and $\widetilde{y}^t$. Because $K(\theta_M^o) \in V(\theta_V)$, then $\exists \theta_V^o$ such that $K(\theta_M^o) = V(\theta_V^o)$. Thus, in the filter set $\mathcal{V}$, there exists at least one element giving the minimum variance estimate of $\widetilde{z}^t$. This fact, together with (ii), proves (iii). □

In view of Theorem 2, the comments on the advantages of the direct approach in solving the FD2 problem reported after Theorem 1, extend also to the case of nonlinear systems. In addition, it must be remarked that, in the nonlinear case, the minimum variance filter $\widehat{K}$, in general, cannot be actually computed, and only approximations of $\widehat{K}$ can be derived. These approximations may often lead to large deteriorations in estimation accuracy with respect to the theoretical minimum variance. Thus, the performance improvement of the direct approach over the two-step procedure for nonlinear systems may be even more significant than for linear systems, because the direct approach can take advantage of the recent advances in nonlinear system identification, leading to quite efficient filter design, as shown by the examples presented in Section 4 and in [21–25].

3.3. *Summary of the Direct Virtual Sensor design*

The DVS design procedure in the present parametric-statistical framework can be summarized as follows:

Direct Virtual Sensor design procedure

- Collect a set of data $D_V \doteq \{(\widetilde{u}^t, \widetilde{y}^t), \widetilde{z}^t, t = 1, 2, \ldots, T\}$ from the system S.
- If the system S described by Equation (1) is linear, use a linear filter structure (e.g. ARX, OE, ARMAX, state-space). Otherwise, use a nonlinear filter structure (e.g. NARX, NOE, NARMAX), where the function f_V in (5) is chosen according to one of the standard parametrizations, which can be found in the literature (e.g. neural networks, polynomials, radial basis functions, etc.).
- Using the data set D_V and the chosen filter structure, design the DVS solving the optimization problem (3).

If the system S is linear, the designed DVS is given by

$$\widehat{z}_V^t = \widehat{\theta}_V \cdot (\widehat{z}_V^{t-1}, \ldots, \widehat{z}_V^{t-n_V}, \widetilde{y}^t, \ldots, \widetilde{y}^{t-n_V}, \widetilde{u}^t, \ldots, \widetilde{u}^{t-n_V}) \tag{6}$$

where $\widehat{\theta}_V \in \mathbb{R}^{n_z \times n_V(n_z+n_y+n_u)}$, and $\cdot$ denotes the dot product. If the system S is nonlinear, the designed DVS is given by

$$\widehat{z}_V^t = f_V(\widehat{\theta}_V, \widehat{z}_V^{t-1}, \ldots, \widehat{z}_V^{t-n_V}, \widetilde{y}^t, \ldots, \widetilde{y}^{t-n_V}, \widetilde{u}^t, \ldots, \widetilde{u}^{t-n_V}). \tag{7}$$

4. EXAMPLES

In this section, examples related to the FD2 problem for nonlinear systems are presented. Examples for linear systems can be found in [22–25].

4.1. *Example 1: estimation of Lorenz attractor*

The Lorenz system is a nonlinear three-dimensional dynamical system derived from the simplified equations of convection rolls arising in atmospheric dynamics. For certain parameter values, the system exhibits chaotic behavior and displays what is called a strange attractor.

An explicit Euler discretization of the continuous-time Lorenz system has been considered:

$$\begin{aligned}
x_1^{t+1} &= (1-\tau\sigma)\, x_1^t + \tau\sigma x_2^t \\
x_2^{t+1} &= (1-\tau)\, x_2^t - \tau x_1^t x_3^t + \tau\rho x_1^t \\
x_3^{t+1} &= x_3^t + \tau x_1^t x_2^t - \tau\beta x_3^t \\
y^t &= x_1^t \\
z^t &= x_2^t x_3^t
\end{aligned} \tag{8}$$

where $\tau = 0.01$, $\sigma = 10$, $\rho = 28$, $\beta = 2.6667$. Note that, for such parameter values, the system is characterized by a chaotic behavior, and thus filtering is a quite difficult task.

A set of 8000 data have been generated from Equation (8). The values of y^t and z^t have been corrupted by i.i.d. Gaussian noises of zero means and standard deviations 0.3 and 10, respectively. Note that the standard deviation of y^t is about 8.2, giving a signal to noise ratio of 27. The standard deviation of z^t is about 250, giving a signal to noise ratio of 25.

The first 6000 data have been used for filter design:

Design data set:
$D_V \doteq \{\widetilde{y}^t, \widetilde{z}^t, t = 1, \ldots, 6000\}$.

The data from 6001 to 8000, not previously used for design, have been used for filter test:

Testing data set:
$D_T = \{\widetilde{y}^t, z^t, t = 6001, \ldots, 8000\}$.

DOI: 10.1002/rnc

The following filters have been derived:

Extended Kalman filter (EKF)

An extended Kalman filter (EKF) has been derived using as model the true system (8), and assuming exact knowledge of the probability density functions of initial conditions and noise. Note that this is the most favorable situation for the two-step approach, where in the first step, an exact identification is performed both of system dynamics and noise statistics. Obviously, this situation rarely occurs in practical applications. The EKF filter has been obtained by linearization of Equation (8) along the estimated trajectory.

Particle filter (PF)

A particle filter (PF) using Sampling Importance Resampling (SIR) (see e.g. [36], [12]) has been implemented to sequentially filter the noisy observation set $\{\tilde{y}^t : t = 1, \dots, 8000\}$ corresponding to the chaotic Lorenz system (8). The SIR filter has been implemented using exact knowledge of the deterministic Lorenz system equations (8), measurement noise statistics, and statistics on the initial condition. Five hundred particles have been used to approximate the prior and posterior distributions sequentially. The distribution on the initial condition has been assumed Gaussian with identity covariance and mean equal to the true initial condition.

This filter showed quite low filtering accuracy. Indeed, a SIR filter implemented in this way performs poorly when trying to filter noisy observations coming from a deterministic chaotic system, because the support for the posterior distribution on $\widetilde{z}^t$ almost surely collapses to a single particle. Particle filters use discrete sample-based distributions to approximate the true posterior distributions being sought after. Because the system being studied is chaotic in nature, the particles making up the sample-based distributions naturally become disperse with time. Consequently, particles drifting far from the value of the true state receive low probability weights during the importance sampling stage. As a result of their low weights, these particles are less likely to survive during the resampling stage. This effect causes the support of the sample-based approximation of the posterior distribution to almost surely collapse onto a single particle.

In order to prevent the support of the discrete approximation of the posterior distribution from collapsing onto a single particle, we have implemented an ad hoc modification of the SIR filter, which we denote as the PF filter. In this PF filter, at time t, before the importance sampling stage, each particle p^t is updated to $p^{t\prime}$ by perturbing it slightly by a zero mean Gaussian random vector w^t with diagonal covariance $\sigma^2\mathbf{I}$:

$$p^{t\prime} = p^t + w^t.$$

Note that this method is akin to modeling the dynamical system as having additive Gaussian noise w^t. The purpose of these slight perturbations is to introduce variability in the particle population in order to avoid collapse of the support onto a single particle. The key is to choose σ^2 large enough to introduce reasonable separation between identical particles and small enough to avoid a significant increase in the variance of the particle distribution.

Linear direct virtual sensor (LDVS)

A linear DVS has been designed from the data set D_V following the procedure of subsection 3.3. This DVS is given by

$$\widehat{z}_L^t = \widehat{\theta}_L \cdot (\widehat{z}_L^{t-1}, \dots, \widehat{z}_L^{t-3}, \widetilde{y}^t, \dots, \widetilde{y}^{t-3}) \tag{9}$$

where $\widehat{\theta}_L = (0.3973, 0.1372, -0.0553, 0.0664, 0.0001, -0.0369, -0.0165)$ and $\cdot$ denotes the dot product.

Nonlinear direct virtual sensor (DVS)

A nonlinear DVS has been designed from the data set D_V following the procedure of Subsection 3.3. This DVS is given by

$$\widehat{z}_V^t = f_V(\widehat{\theta}_V, \widehat{z}_V^{t-1}, \dots, \widehat{z}_V^{t-3}, \widetilde{y}^t, \dots, \widetilde{y}^{t-3}).$$

The function f_V is a hidden layer neural network (see e.g. [37]) composed by r neurons:

$$f_V(\theta_V, \varphi) = \sum_{i=1}^{r} \alpha_i \sigma(\beta_i \cdot \varphi - \lambda_i) + \zeta \tag{10}$$

DOI: 10.1002/rnc

where $\theta_V = \{\alpha_i, \lambda_i, \zeta \in \mathbb{R}, \beta_i \in \mathbb{R}^{n_\varphi}, i = 1, \dots, r\}$ and $\sigma(h) = 2/(1 + e^{-2h}) - 1$ is a sigmoidal function. Several neural networks of the form (10) with different number of neurons (from $r = 1$ to $r = 15$) have been identified on the set D_V by means of the NNOE function of the NNSYD Matlab Toolbox [38]. A neural network composed by $r = 6$ neurons has been chosen.

Filtering results

The root mean square estimation error (RMSE) and the maximum estimation error in absolute value (MAXE) provided by the filters on the testing data set are reported in Table I. The true signal and the ones estimated by the designed filters are shown in Figure 1 for a portion of the testing data set.

It can be noted that the estimation accuracy of the EKF filter is quite poor, even if this filter has been designed using exact knowledge of system equations, noise statistics, and initial conditions statistics. The estimation accuracy of the LDVS filter, although not extremely good, is still decent, indicating that a very simple linear DVS can be significantly better than a quite complicated nonlinear filters, even in filtering of complex nonlinear systems. This fact suggests that, in any filter design

Table I. Absolute (relative) RMSE and MAXE estimation errors.

Filter	RMSE	MAXE
EKF	270 (110%)	1760 (230%)
PF	11 (4%)	106 (10%)
LDVS	55 (22%)	278 (38%)
DVS	6.3 (2.5%)	27 (3.5%)

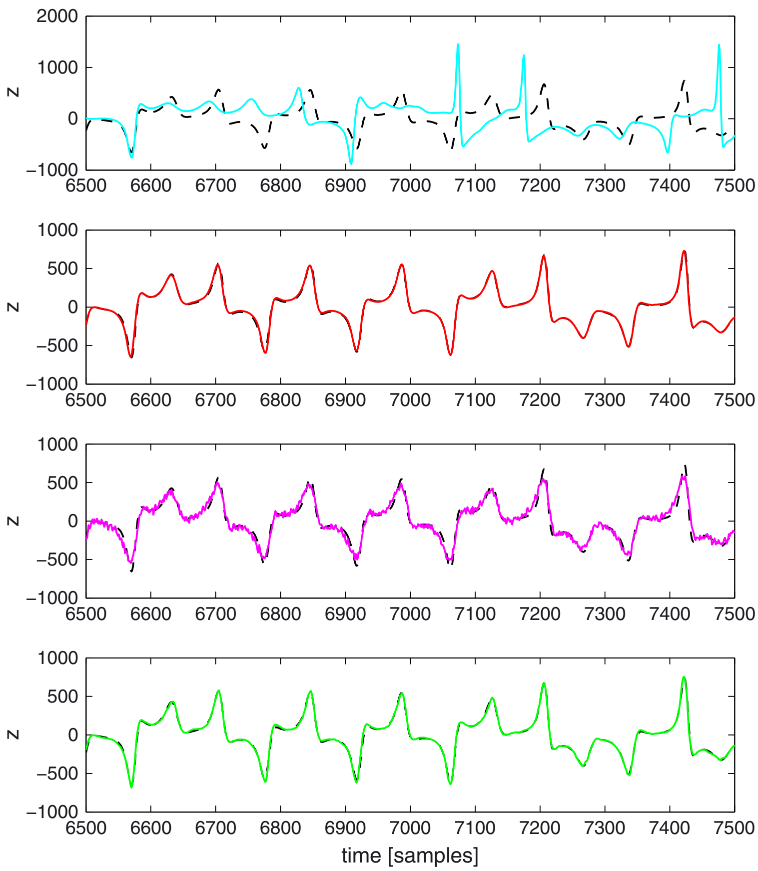

Figure 1. From top to bottom: EKF, PF, LDVS and DVS estimates. Dashed (black): true signal. Continuous (colored): filter estimate. [The colored version of the figures is available online.]

DOI: 10.1002/rnc

problem, it may be convenient to first try with a simple linear DVS, then move to more complex nonlinear filters if the linear DVS does not give a satisfactory estimation accuracy. The PF filter provides quite good estimation accuracy. The DVS filter provides an excellent estimation accuracy. It must be remarked that the PF filter has been implemented using exact knowledge of system equations, noise statistics, and statistics on initial conditions, whereas the DVS filter has been designed without using such a strong information.

4.2. *Example 2: estimation of vehicle yaw rate*

In this example, filter design for estimating the vehicle yaw rate has been considered. The knowledge of such a variable is used by Vehicle Dynamics Control (VDC), a highly studied active safety system aimed at enhancing the vehicle stability, see, for example [39]. The VDC generally provides a control action, which prevents the vehicle from under-steering or over-steering in a handling maneuver (e.g. lane change, steering angle step, etc.). In order to generate the required control actions, the commercially available VDC systems use the values of yaw rate, lateral acceleration, and vehicle longitudinal velocity, measured by appropriate sensors. However, the cost of the yaw rate sensors alone is quite high compared with the overall cost of the VDC system. Thus, the availability of an accurate yaw rate virtual sensor, eliminating the need of the yaw rate sensor, could allow a significant reduction in the VDC systems production costs and, consequently, a larger diffusion of active safety systems on commercial cars, even in the segments B and C.

Filter design has been performed using the experimental data measured on a passenger car provided by Fiat Auto S.p.A. The data have been obtained from different maneuvers, including steering angle steps of different amplitude (from 30 to 80 degrees), double lane changes, and frequency sweeps performed on a dry road. The data set is composed of the measurements of the following variables: $\dot{\psi}^t$: yaw rate, α_S^t : steering angle, a_y^t : lateral acceleration, and v_{lon}^t : longitudinal velocity. A set of 2641 data has been recorded over a time interval of 211 s. with a sample time of 0.08 s. The data are shown in Figure 2.

Considering the case in which the VDC control actuation is realized by steering, the vehicle dynamics relating these variables may be described by the set of equations (1), where the steering angle and the longitudinal velocity are the measured inputs: $\widetilde{u}^t = (\alpha_S^t, v_{lon}^t)$, the lateral acceleration is the measured output: $\widetilde{y}^t = a_y^t$, and the yaw rate is the variable to be estimated: $\widetilde{z}^t = \dot{\psi}^t$.

The data set corresponding to the first 132 s of the complete data set has been used for filter design:

$$\begin{aligned} &\text{Design data set:} \\ &D_V = \{(\widetilde{u}^t, \widetilde{y}^t), \widetilde{z}^t, t = 1, \ldots, 1656\} \\ &\quad\;\; = \{(\alpha_S^t, v_{lon}^t, a_y^t), \dot{\psi}^t, t = 1, \ldots, 1656\}. \end{aligned}$$

The designed filters have been tested on the set composed of the data recorded from second 132 to second 211 of the complete data set, not previously used for design:

$$\begin{aligned} &\text{Testing data set:} \\ &D_T = \{(\widetilde{u}^t, \widetilde{y}^t), \widetilde{v}^t, t = 1657, \ldots, 2641\} \\ &\quad\;\; = \{(\alpha_S^t, v_{lon}^t, a_y^t), \dot{\psi}^t, t = 1657, \ldots, 2641\}. \end{aligned}$$

The following filters have been designed:

Kalman filter designed from an identified linear parameter-varying physical model (KF)

The filter KF has been designed from a physical model of vehicle lateral dynamics. A large literature exists on deriving physical models describing this dynamics (see e.g. [40], [41]). It must be remarked that, in the working conditions of the VDC systems, this dynamics is quite complex, and thus obtaining accurate models is, in general, a difficult task.

DOI: 10.1002/rnc

The following LPV model, called single-track model (see e.g. [40], [42]), has been considered:

$$
\begin{aligned}
m v_{lon}^{t}(t)\dot{\beta}(t) + m v_{lon}^{t}(t)\dot{\psi}(t) &= F_{yf}(t) + F_{yr}(t) \\
J\ddot{\psi}(t) &= aF_{yf}(t) - bF_{yr}(t) + w_x(t) \\
F_{yf}(t) + l_f/v_{lon}^{t}(t)\dot{F}_{yf}(t) &= -c_f(\beta(t) + a\dot{\psi}(t)/v_{lon}^{t}(t) - \alpha_S(t)) \\
F_{yr}(t) + l_r/v_{lon}^{t}(t)\dot{F}_{yr}(t) &= -c_r(\beta(t) - b\dot{\psi}(t)/v_{lon}^{t}(t)) \\
a_y(t) &= (F_{yf}(t) + F_{yr}(t))/m + w_y(t)
\end{aligned}
\tag{11}
$$

where β is the side-slip angle, $\dot{\psi}$ is the yaw rate, F_{yf} and F_{yr} are the front and rear axle lateral forces, v_{lon}^{t} is the longitudinal vehicle speed, α_S is the steering angle, w_x is the process noise modeled as a disturbing torque, w_y is measurement noise on the lateral acceleration output. The model parameters are: m is the vehicle mass, J is the moment of inertia around the vertical axis, a and b are the distances between the center of gravity and the front and rear axles respectively, l_f and l_r are the front and rear tire relaxation lengths, and c_f and c_r are the front and rear axle cornering stiffnesses. The inputs of the model are α_S and v_{lon}^{t}, the outputs are a_y and $\dot{\psi}$.

The values of these parameters have been identified from the following data set:

Identification data set:

$$
\begin{aligned}
D_M &= \{\widetilde{u}^t, (\widetilde{y}^t, \widetilde{z}^t), t = 1, \ldots, 1656\} \\
&= \{(\alpha_S^t, v_{lon}^t), (a_y^t, \dot{\psi}^t), t = 1, \ldots, 1656\}.
\end{aligned}
$$

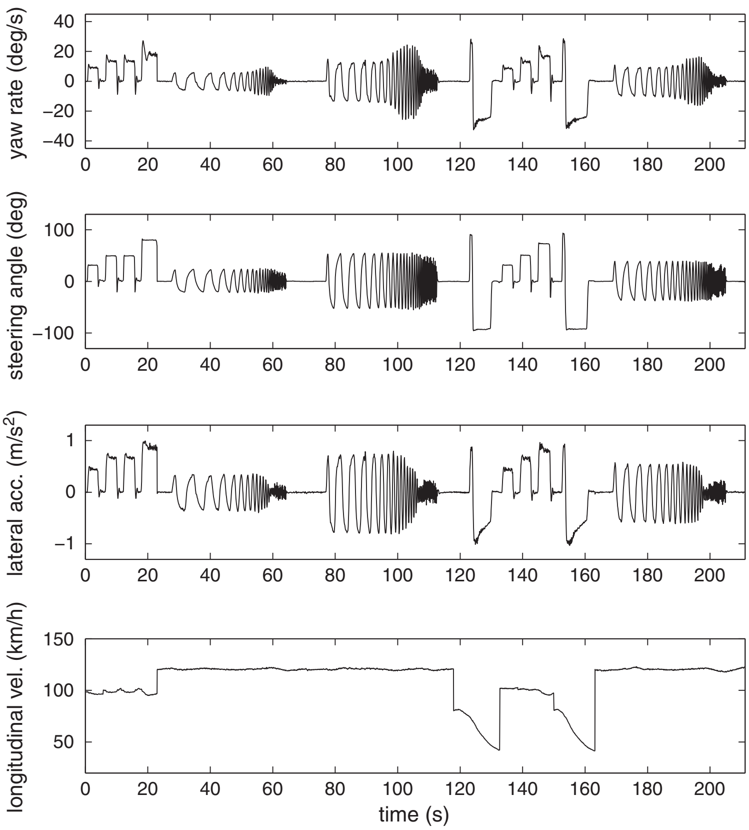

Figure 2. Recorded data. Sequence of manoeuvres: four steering steps, two frequency sweeps, one double-lane change, four steering steps, one double-lane change, one frequency sweep.

Identification has been performed using a nonlinear optimization routine, minimizing the cost function

$$J(\theta) = a_1 \sum_{t=1}^{T} \left(\widetilde{y}^t - \widehat{y}^t(\theta)\right)^2 + a_2 \sum_{t=1}^{T} \left(\widetilde{z}^t - \widehat{z}^t(\theta)\right)^2 + \sum_{i=1}^{8} b_i \left(\theta_i^{nom} - \theta_i\right)^2$$

where $\widehat{y}^t(\theta)$ is the lateral acceleration simulated by a discretization of the model (11), $\widehat{z}^t(\theta)$ is the simulated yaw rate, and $\theta = [m, J, a, b, l_f, l_r, c_f, c_r]$ is the vector of the parameters to estimate. a_1, a_2 and b_i's are weighting coefficients aimed at balancing the contributions of the various terms in the cost function. The term $\sum_{i=1}^{8} b_i \left(\theta_i^{nom} - \theta_i\right)^2$ has been added in order to guarantee that the estimated parameters are not too different from the nominal ones, given by $\theta^{nom} = [1715, 2697, 1.07, 1.47, 0.1, 0.1, 89733, 114100]$, so that the identified model maintains a reasonable physical meaning. Identification has also been performed without this term: a model has been obtained, with accuracy very similar to the one presented here, but with parameters having poor physical meaning. For example, the value estimated for the mass m was 867 Kg (smaller than a half of the nominal value), i.e. a value which has nothing to do with the real car from which the data have been collected. Note that this drawback is quite typical when the prediction error method is used to identify the parameters of a physical model: using only the criterion of prediction error minimization may yield a model whose parameters have been forced to give a good fitting of the data, but have poor physical reliability.

Several values of the weighting coefficients have been considered and, correspondingly, several discrete-time models have been identified. A model representing an acceptable compromise between physical meaning and accuracy has been selected, obtained using the following weights: $a_1 = 1.6e5$, $a_2 = 336$, $b_1 = b_2 = 2$, $b_3 = \ldots = b_8 = 1$. The parameter vector of this model is $\widehat{\theta} = [1681, 2280, 0.95, 1.5, 0.375, 0.53, 98351, 134570]$. It can be noted that these parameters are not very different from the starting ones in θ^{nom}, and still maintain a reasonable physical meaning. In Figure 3, the data simulated by the identified single-track model are compared with the measured data on a portion of the testing data set, for both the outputs a_y^t and $\dot{\psi}^t$. It can be observed that the model turns out to be fairly accurate.

The KF has been designed from this model, assuming the following variances: $\mathcal{Q} = 10^5$ (process noise variance) and $\mathcal{R} = 0.04$ (measurement error variance).

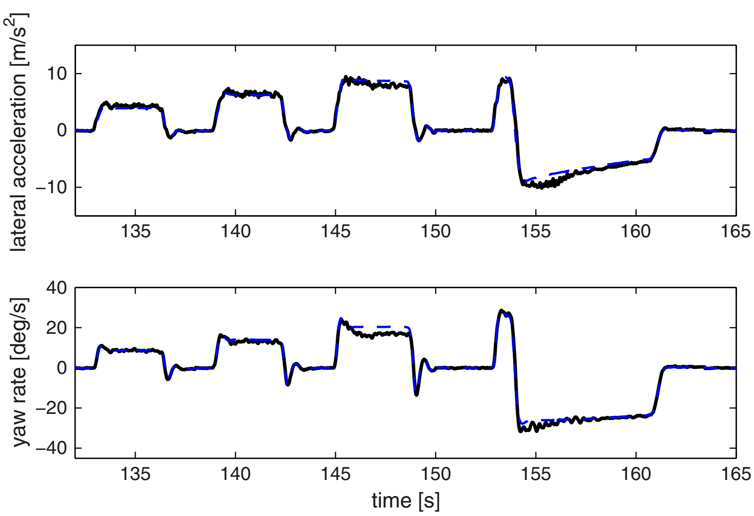

Figure 3. Comparison between experimental data and simulation of the identified single-track model. Bold (black) line: experimental data. Dashed (blue) line: single-track model. [The colored version of the figures is available online.]

DOI: 10.1002/rnc

Particle filter (PF)

A PF using SIR has been implemented to sequentially filter the noisy measured observations of the vehicle lateral acceleration a_y in order to estimate the vehicle yaw rate. The PF filter has been implemented assuming the vehicle to behave according to the identified single-track vehicle dynamics (11). One hundred particles have been used to approximate the prior and posterior distributions sequentially. The distribution on the initial condition has been assumed Gaussian with identity covariance and mean equal to zero.

Extended Kalman filter designed from an identified nonlinear model (EKF)

The following Nonlinear Auto-Regressive with eXternal input model structure has been considered:

$$v_M^t = f_M(\theta_M, v_M^{t-1}, \ldots, v_M^{t-n_M}, \widetilde{u}^{t-1}, \ldots, \widetilde{u}^{t-n_M}) \tag{12}$$

where $v_M^t = (y_M^t, z_M^t)$ is the model output and f_M is a neural network of the form (10), composed of r neurons.

Several orders n_M and numbers of neurons r have been considered. Correspondingly, several models of the form (12) have been identified from the data set D_M, using the PE method by means of the NNOE function of the NNSYD Matlab Toolbox [38]. A model with $n_M = 3$, and $r = 4$, showing good simulation accuracy, has been selected.

In order to design an Extended Kalman filter, the model has been written in the following state-space form:

$$\begin{aligned} x_M^{t+1} &= \widehat{F}\left(x_M^t, \widetilde{U}^t\right) + w_x^t \\ y_M^t &= \widehat{H}_y\left(x_M^t, \widetilde{U}^t\right) + w_y^t \\ z_M^t &= \widehat{H}_z\left(x_M^t, \widetilde{U}^t\right) \end{aligned} \tag{13}$$

where $x^t = (v_M^t, \ldots, v_M^{t-2})$, $\widetilde{U}^t = (\widetilde{u}^t, \ldots, \widetilde{u}^{t-2})$, and the functions $\widehat{F}$, $\widehat{H}_y$, and $\widehat{H}_z$ can be directly obtained from f_M.

An Extended Kalman filter has been designed from this state-space model, by linearization of the equations (13) along the estimated trajectory. The following variances for the noises w_x^t and w_y^t have been assumed: $\mathcal{Q} = 10^3 I$ (noise w_x^t), where I is the identity matrix, and $\mathcal{R} = 0.04$ (noise w_y^t).

Direct virtual sensor (DVS)

A nonlinear DVS has been designed from the data set D_V following the procedure of Subsection 3.3. This DVS is given by

$$\widehat{z}_V^t = f_V(\theta_V, \widehat{z}_V^{t-1}, \ldots, \widehat{z}_V^{t-3}, \widetilde{y}^t, \ldots, \widetilde{y}^{t-3}, \widetilde{u}^t, \ldots, \widetilde{u}^{t-3}).$$

The function f_V is a neural network of the form (10), composed of r neurons. Several neural networks with different number of neurons (from $r = 1$ to $r = 15$) have been identified on the set D_V by means of the NNOE function of the NNSYD Matlab Toolbox [38]. A neural network composed by $r = 4$ neurons has been chosen.

Filtering results

The root mean square estimation error and the maximum estimation error in absolute value provided by the filters are reported in Tables II and III for the different maneuvers of the testing data set. In Figures 4 and 5, the measured yaw rate and the one estimated by DVS are compared in three maneuvers of the Testing data set.

It can be noted that the DVS provides significantly better estimation accuracy compared with the KF and PF filters. Another relevant advantage of the DVS with respect to the KF and PF filters is the greater design simplicity. Indeed, DVS design only requires to identify a single-output filter. On the other hand, the design of the KF or PF filter requires the following steps: (1) identification of a two-output model (which is quite more complicated than identification of a single-output model); and (2) filter design from the identified model.

DOI: 10.1002/rnc

Table II. Absolute (relative) RMSE estimation errors (deg). Steering angle step (SAS), Double-lane change (DLC), Frequency sweep (FS).

Maneuver	KF	PF	EKF	DVS
SAS 30^o	0.34 (6%)	0.34 (6%)	0.36 (7%)	0.3 (4%)
SAS 50^o	0.77 (8%)	0.64 (7%)	0.53 (7%)	0.44 (4%)
SAS 75^o	2.57 (18%)	1.85 (13%)	1.05 (11%)	0.58 (4%)
DLC	1.43 (8%)	1.42 (8%)	1.19 (7%)	0.6 (3%)
FS	0.58 (8%)	0.60 (8%)	0.55 (8%)	0.3 (4%)

Table III. Absolute (relative) MAXE estimation errors (deg). Steering angle step (SAS), Double-lane change (DLC), Frequency sweep (FS).

Maneuver	KF	PF	EKF	DVS
SAS 30^o	1.06 (10%)	0.97 (9%)	0.93 (9%)	1.05 (10%)
SAS 50^o	2.27 (14%)	2.24 (14%)	1.81 (11%)	1.29 (8%)
SAS 75^o	5.49 (23%)	5.35 (22%)	3.05 (13%)	1.61 (7%)
DLC	5.58 (18%)	7.04 (22%)	4.73 (15%)	2.24 (7%)
FS	1.96 (12%)	2.16 (13%)	2.19 (13%)	1.22 (7%)

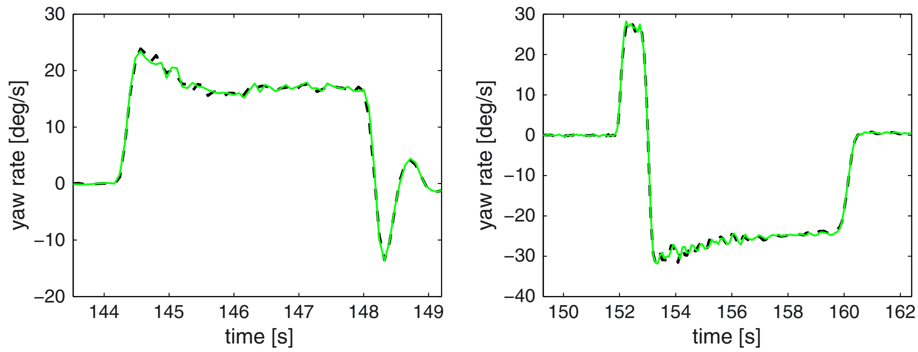

Figure 4. Left: SAS 75^o. Right: DLC. Dashed (black): measured yaw rate. Continuous (green): DVS estimate. [The colored version of the figures is available online.]

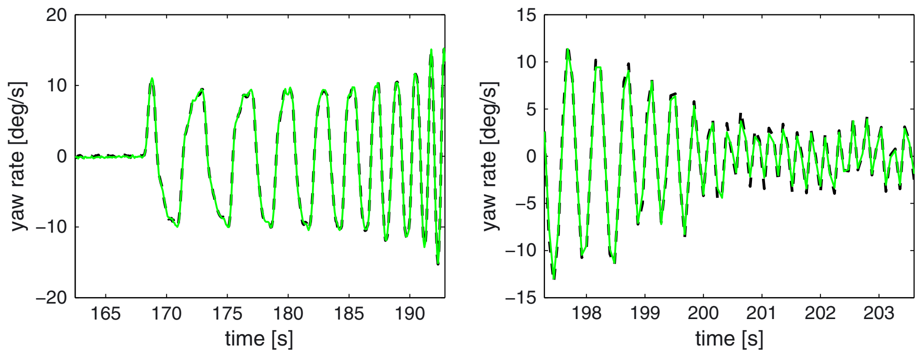

Figure 5. Left: FS, low frequencies. Right: FS, high frequencies. Dashed (black): measured yaw rate. Continuous (green): DVS estimate. [The colored version of the figures is available online.]

5. CONCLUSIONS

In this paper, the problem of filter design for LTI and nonlinear systems is investigated within a parametric-statistical framework. The paper presents advantages over the existing literature on several aspects.

DOI: 10.1002/rnc

First, the existing literature investigates problems that can be denoted as "filter design from known systems", indicating that the filter is designed assuming the knowledge of the equations describing the system to be filtered. In this paper, a more general filtering problem is considered, denoted as "filter design from data (FD2)", where the system is not known, but a set of data is available. Clearly, a solution to this problem can be obtained by identifying a model from data, whose equations are then used by any of the available filter design methods for known systems. However, this two-step procedure is, in general, not optimal. Indeed, finding optimal solutions to the filter design from data problem appears to be not an easy task. In this paper, a methodology able to derive directly from measured data an optimal filter is presented .

Second, even in the most favorable situations for the two-step design, that is, the system S belongs to the model structure selected for identification and the minimum variance filter for the identified model can be actually designed, the direct design leads to filters with estimation error lower or equal to the one obtained from the two-step design. Moreover, in the presence of undermodeling, the direct design leads to the minimum error estimator among the selected filter class. A similar result is not ensured for filters obtained from the two-step design, whose performance deterioration caused by undermodeling may be significantly large, see for example [24], [23], and the examples presented in this paper.

Third, the advantages of the direct approach may be particularly relevant in the case of nonlinear system. Indeed, optimal filters for nonlinear systems are, in general, difficult to derive and/or to implement, and approximate solutions often exhibit poor performance and stability problems. On the contrary, the direct approach can take advantage of the recent advances in nonlinear identification, leading to very efficient filter design.

Fourth, the proposed direct approach is significantly simpler than the standard two-step approach. Indeed, the DVS design only requires the identification of a single-output filter (with output z). The design of a two-step filter requires the identification of a multi-output model (with outputs z and y), and, furthermore, the design of a filter from this model.

It can be concluded that the direct filter design is an efficient solution to the FD2 problem, having features far superior to the two-step approach.

APPENDIX A: BASIC CONDITIONS AND RESULTS

In this appendix, some basic conditions and results of [35] and [26] are summarized, which are used in Section 3 to prove our theorems. In the present context, these conditions and results can be formulated as follows:

Consider a system in regression form

$$\widetilde{z}^t = g(\widetilde{z}^{t-1}, \dots, \widetilde{z}^{t-n_x}, \widetilde{q}^t, \dots, \widetilde{q}^{t-n_x}) + w^t \tag{14}$$

where $\widetilde{q}^t \in Q \subseteq \mathbb{R}^{n_q}$ is a deterministic input, $\widetilde{z}^t \in Z \subseteq \mathbb{R}^{n_z}$ is the output, and $w^t \in \mathbb{R}^{n_z}$ is an i.i.d. stochastic noise with zero mean. Let $\mathfrak{W}$ be the σ-algebra generated by w^t.

Condition S3 [35]

The system (14) is exponentially stable. That is, for any input sequence $\widetilde{q}^t$, and for any $t, s : t \geqslant s$, there exist a random variable $z_s^t \in \mathfrak{W}$ independent of $\mathfrak{W}$ such that

$$\bar{E} \left\| \widetilde{z}^t - z_s^t \right\|^4 < C \lambda^{t-s}$$

for some $0 \leqslant C < \infty$ and $0 \leqslant \lambda < 1$.

Condition S3 essentially means that $\widetilde{z}^t$ can be approximated by a random variable z_s^t that is independent of the remote past. In other words, this condition requires that the mapping from $\widetilde{q}^t$ and w^t to $\widetilde{z}^t$ has exponential fading memory. It is immediate to verify that the system (2) satisfies Condition S3 because, at any time t, $\widetilde{z}^t$ depends only on a finite number of past values of $\widetilde{u}^t$, $\widetilde{y}^t$, and $w^t = (w_x^t, w_y^t)$.

Consider now a parametric model of the system (14):

$$\widehat{z}^t = f(\theta, \widehat{z}^{t-1}, \dots, \widehat{z}^{t-n_x}, \widetilde{q}^t, \dots, \widetilde{q}^{t-n_x}) \tag{15}$$

DOI: 10.1002/rnc

where $\theta \in \Theta \subseteq \mathbb{R}^{n_\theta}$ is a parameter vector. Note that, for fixed $\theta \in \Theta$, $f(\theta, \varphi^t)$ is a single model, whereas for variable $\theta \in \Theta$, $f(\theta, \varphi^t)$ defines a *model structure* [26].

For any fixed $\theta \in \Theta$, the output $\widehat{z}^t$ at time t is computed by iteration of the difference Equation (15):

$$\begin{aligned}
\widehat{z}^{n_x+1} &= f(\theta, \widehat{z}^{n_x}, \ldots, \widehat{z}^1, \widetilde{q}^{n_x+1}, \ldots, \widetilde{q}^1) \\
\widehat{z}^{n_x+2} &= f(\theta, \widehat{z}^{n_x+1}, \ldots, \widehat{z}^2, \widetilde{q}^{n_x+2}, \ldots, \widetilde{q}^2) \\
&= f(\theta, f(\theta, \widehat{z}^{n_x}, \ldots, \widehat{z}^1, \widetilde{q}^{n_x+1}, \ldots, \widetilde{q}^1), \ldots, \widehat{z}^2, \widetilde{q}^{n_x+2}, \ldots, \widetilde{q}^2) \\
&\vdots \\
\widehat{z}^t &= f^t\left(\theta, \mathbf{z}^{n_x}, \mathbf{q}^t\right)
\end{aligned}$$

where $\mathbf{z}^{n_x} = \left(\widehat{z}^{n_x}, \ldots, \widehat{z}^1\right)$ is the system initial condition, $\mathbf{q}^t = \left(\widetilde{q}^1, \ldots, \widetilde{q}^t\right)$ is the input sequence, and f^t is the function obtained by $t - n_x$ compositions of the function f.

Condition M1 [35]

(i) The function f^t is differentiable with respect to θ for $\forall\theta \in \Theta$, $\forall\mathbf{z}^{n_x} \in Z^{n_x}$, $\forall\mathbf{q}^t \in Q^t$, $\forall t \in \mathbb{N}$.
(ii) Θ is compact.
(iii) There exist $0 \leqslant C < \infty$ and $0 \leqslant \lambda < 1$ such that

$$1)\ \left\| f^t\left(\theta, \mathbf{z}^{n_x}, \mathbf{q}^t\right) - f^t\left(\theta, \bar{\mathbf{z}}^{n_x}, \bar{\mathbf{q}}^t\right)\right\| \leqslant C \sum_{k=0}^{t-1} \lambda^k \left\|\widetilde{q}_{t-k} - \bar{\widetilde{q}}_{t-k}\right\| + C\lambda^t \left\|\mathbf{z}^{n_x} - \bar{\mathbf{z}}^{n_x}\right\|$$

$$2)\ \left\| f^t(\theta, 0, 0)\right\| \leqslant C$$

for $\forall t \in \mathbb{N}$, $\forall\mathbf{z}^{n_x}, \bar{\mathbf{z}}^{n_x} \in Z^{n_x}$, $\forall\mathbf{q}^t, \bar{\mathbf{q}}^t \in Q^t$, $\forall\theta \in \overline{\Theta}$, where $\overline{\Theta}$ is an open neighborhood of Θ.
(iv) $df^t\left(\theta, \mathbf{z}^{n_x}, \mathbf{q}^t\right)/d\theta$ is subject to (iii).

Condition M1 essentially means that the model structure $f(\theta, \varphi^t)$ is uniformly exponentially stable. It also means that the model structure has uniform exponential fading memory, that is, the remote past inputs are forgotten at an exponential rate.

Suppose now that a set of measurements collected from the system (14) is available:

$$D \doteq \{\widetilde{q}^t, \widetilde{z}^t, t = 1, 2, \ldots, T\}$$

and the PE method [26] is used for the identification of a model of the system (14). This model, denoted by $\widehat{M} = M(\widehat{\theta})$, is of the form (15) and is identified solving the following optimization problem:

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} J_T(\theta)$$

$$J_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{2}\left\|e^t(\theta)\right\|^2$$

where $e^t(\theta) = \widetilde{z}^t - \widehat{z}_M^t$ is the PE of the model $M(\theta)$, being $\widehat{z}_M^t$ the prediction given by $M(\theta)$, and $\|\cdot\|$ is the ℓ_2 norm. Lemma 3.1 of [35] can now be reported.

Lemma 3.1 [35]

Let the system (14) be subject to Condition S3. Let the model structure $f(\theta, \varphi^t)$ be subject to Condition M1. Then,

$$\sup_{\theta\in\Theta} |J_T(\theta) - \bar{E}\frac{1}{2}\left\|e^t(\theta)\right\|^2| \to 0, \quad w.p.1 \quad as \quad T \to \infty$$

DOI: 10.1002/rnc

Lemma 3.1 holds, in general, for nonlinear systems, whereas a simplified version of it for linear systems is given by Lemma 8.2 in [26]. This version is based on Condition D1 in [26] which, in the present context, can be formulated as follows:

Condition D1 [26]

(i) $\widetilde{q}^t$ is a bounded input signal.
(ii) w^t is an i.i.d. random variable with zero mean and bounded moments of order $4+\delta$, for some $\delta > 0$.
(iii) The system (14) is *stable*, that is

$$\sum_{t=1}^{\infty} h^t < \infty$$

where h^t, $t = 1, 2, \ldots$, $h^t \in \mathbb{R}^{n_z}$ is the system impulse response.
(iv) $\widetilde{q}^t$ and $\widetilde{z}^t$ are jointly quasi-stationary (see [26]).

Lemma 8.2 of [26] also uses the notion of *uniformly stable* model structure, which is a simplified version of Condition M1 for linear systems (see [26] for details).

Lemma 8.2 [26]
Let the system (14) be linear and subject to Condition D1. Let the model structure $f(\theta, \varphi^t) : \theta \in \Theta$ be uniformly stable. Then,

$$\sup_{\theta \in \Theta} |J_T(\theta) - \bar{E}\frac{1}{2}\left\|e^t(\theta)\right\|^2| \to 0, \quad w.p.1 \quad as \quad T \to \infty.$$

An important corollary of Lemma 8.2 is the following.

Theorem 8.2 [26]
Let the system (14) be linear and subject to Condition D1. Let the model structure $f(\theta, \varphi^t)$ be uniformly stable. Then,

$$\widehat{\theta} \to MS, \quad w.p.1 \quad as \quad T \to \infty$$

where MS is the set of minimizers of $\bar{E}\frac{1}{2}\left\|e^t(\theta)\right\|^2$.

Under the additional conditions of *globally identifiable* model structure and *informative enough* data set (see [26] for the related definitions), this lemma leads to the following theorem.

Theorem 8.3 [26]
Let the system (14) be linear, subject to Condition D1, and $g(\cdot) = f(\theta^o, \cdot)$, for some $\theta^o \in \Theta$. Let the model structure $f(\theta, \varphi^t)$ be subject to Condition M1 and globally identifiable. Let the data set D be informative enough. Then, $MS = \{\theta^o\}$.

REFERENCES

1. Jazwinski A. *Stochastic Processes and Filtering Theory*. Academic: New York, 1970.
2. Gelb A. *Applied Optimal Estimation*. The MIT Press: Cambridge, Mass., 1974.
3. Maybeck PS. *Stochastic Models, Estimation, and Control*. Mathematics in Science and Engineering, Academic Press: New York, USA, 1979.
4. Anderson B, Moore J. *Optimal Filtering*. Prentice-Hall: Englewood Cliffs, NJ, 1979.
5. Nagpal K, Khargonekar P. Filtering and smoothing in an $\mathcal{H}_\infty$ setting. *IEEE Transactions on Automatic Control* 1991; **36**:152–166.
6. Shaked U, Theodor Y. $\mathcal{H}_\infty$-optimal estimation: a tutorial. *Proceeedings of the IEEE Conference on Decision and Control* 1992; **2**:2278–2286.
7. Nagpal K, Abedor J, Poolla K. A linear matrix inequality approach to peak-to-peak gain minimization. *IEEE Transactions on Automatic Control* 1996; **41**(1):43–48.
8. Voulgaris PG. On optimal ℓ_∞ to ℓ_∞ filtering. *Automatica* 1995; **31**(3):489–495.
9. Julier S, Uhlmann J, Durrant-Whyte H. A new method for the nonlinear transformation of mean and covariances in filters and estimators. *IEEE Transactions on Automatic Control* 2002; **45**:477–482.
10. Evensen G. *Data Assimilation: The Ensemble Kalman Filter*. Springer: New York, 2007.

11. Ristic B, Arulampalam S, Gordon N. *Beyond Kalman Filter: Particle Filters and Tracking Applications*. Artech House: Boston, MA, 2004.
12. Doucet A, de Freitas N, Gordon N. *Sequential Monte Carlo Methods in Practice*. SpringerVerlag: New York, 2001.
13. Kitagawa G. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics* 1996; **5**(1):1–25.
14. Nguang S, Fu M. Robust nonlinear H_∞ filtering. *Automatica* 1996; **32**(8):1195–1199.
15. McEneaney W. Robust/H_∞ filtering for nonlinear systems. *Systems and Control Letters* 1998; **33**(5):315–325.
16. Seo J, Yu M-J, Park CG, Lee JG. An extended robust h_∞ filter for nonlinear constrained uncertain systems. *IEEE Transactions on Signal Processing* nov. 2006; **54**(11):4471–4475.
17. Chen B-S, Chen W-H, Wu H-L. Robust h_2/h_{infty} global linearization filter design for nonlinear stochastic systems. *IEEE Transactions on Circuits and Systems I: Regular Papers* july 2009; **56**(7):1441–1454.
18. Novara C, Ruiz F, Milanese M. Direct design of optimal filters from data. *17th IFAC World Congress*, Seoul, Korea, 2008.
19. Ruiz F, Novara C, Milanese M. Direct design from data of optimal filters for LPV systems. *Systems and Control Letters* 2010; **59/1**:1–8.
20. Milanese M, Novara C, Hsu K, Poolla K. The filter design from data (FD2) problem: Nonlinear Set Membership approach. *Automatica* 2009; **45**(10):2350-2357.
21. Milanese M, Regruto D, Fortina A. Direct virtual sensor (DVS) design in vehicle sideslip angle estimation. *American Control Conference*, New York City, USA, 2007.
22. Borodani P. Virtual sensors: an original approach for dynamic variables estimation in automotive control systems. *Proceedings of the 9th International Symposium on Advanced Vehicle Control (AVEC)*, Kobe, Japan, 2008.
23. Milanese M, Ruiz F, Taragna M. Virtual sensors for linear dynamic systems: structure and identification. *3rd International IEEE Scientific Conference on Physics and Control (PhysCon 2007)*, Potsdam, Germany, 2007.
24. Milanese M, Ruiz F, Taragna M. Linear virtual sensors for vertical dynamics of vehicles with controlled suspensions. *European Control Conference ECC'07*, Kos, Greece, 2007.
25. Canale M, Fagiano L, Ruiz F, Signorile M. A study on the use of virtual sensors in vehicle control. *47th IEEE Conference on Decision and Control and European Control Conference*, Cancun, Mexico, 2008.
26. Ljung L. *System Identification: Theory for the User*. Prentice Hall: Upper Saddle River, N.J., 1999.
27. Isidori A. *Nonlinear Control Systems*. Springer-Verlag: Berlin, 1989.
28. Hermann R, Krener AJ. Nonlinear controllability and observability. *IEEE Transactions on Automatic Control* 1977; **22**(5):728–740.
29. Vivalda JC. On the genericity of the observability of uncontrolled discrete nonlinear systems. *SIAM Journal on Control and Optimization* 2003; **42**(4):1509–1522.
30. Germani A, Manes C. State observers for nonlinear systems with slowly varying inputs. *Proceedings of the 36th IEEE Conference on Decision and Control*, 1997; 5054–5059.
31. Levin A, Narendra K. Control of nonlinear dynamical systems using neural networks. II. observability, identification, and control. *IEEE Transaction on Neural Networks* 1996; **7**(1):30–42.
32. Takens F. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence*, Rand DA, Young L-S (eds). Springer-Verlag: New York, USA, 1981.
33. Stark J, Broomhead D, Davies ME, Huke J. Delay embeddings for forced systems. II. stochastic forcing. *Journal of Nonlinear Science* 2003; **13**(6):519–577.
34. Milanese M, Novara C. Set membership identification of nonlinear systems. *Automatica* 2004; **40/6**:957–975.
35. Ljung L. Convergence analysis of parametric identification methods. *IEEE Transactions on Automatic Control* 1978; **23**(5):770–783.
36. Gordon N, Salmond D, Smith A. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F on Radar and Signal Processing* 1993; **140**(2):107–113.
37. Hertz J, Krogh A, Palmer G. *Introduction to the Theory of Neural Computation*. Addison-Wesley, Santa Fe Institute Studies in the Sciences of Complexity: Reading (Mass.), 1991.
38. Norgaard M, Ravn O, Poulsen N, Hansen LK. *Neural Networks for Modeling and Control of Dynamic Systems*. Springer: London, GB, 2000.
39. Zanten A, Erhardt R, Pfaff G. VDC, the vehicle dynamics control system of Bosch. *SAE Technical Paper 950759*, 1995.
40. Rajamani R. *Vehicle Dynamics and Control*. Springer: New York, USA, 2006.
41. Bakker E, Lidner L, Pacejka H. A new tyre model with an application in vehicle dynamics studies. In *SAE technical paper 890087*, 1989; 83–93.
42. Canale M, Fagiano L, Milanese M, Borodani P. Robust vehicle yaw control using an active differential and IMC techniques. *Control Engineering Practice* 2007; **15/8**:923–941.

DOI: 10.1002/rnc