Individualized Model Discovery: The Case of Anemia Patients

Elom Akabua[a,*], Tamer Inanc[a], Adam Gaweda[a,b], Michael E. Brier[b,d], Seongho Kim[c], Jacek M. Zurada[a]

[a] *Department of Electrical Engineering, University of Louisville. Louisville, KY. 40292*
[b] *Department of Medicine, University of Louisville. Louisville, KY. 40292*
[c] *Biostatistics Core Karmanos Cancer Institute, Wayne State University. Detroit, MI. 48201*
[d] *Department of Veteran Affairs, University of Louisville. Louisville, KY. 40292.*

Abstract

The focus of this work is to utilize robust identification techniques to extract the inherent dose-response relation for patients with anemia due to chronic renal failure. Exogenous administration of erythropoietin (EPO) is the treatment of choice for this kind of anemia; however, the high cost of the medication, the phenomenon of EPO resistance patients, and the risks of possible adverse effects all justify the need for an optimization based approach for individualized therapy. Using past patient measurement records, the robust identification procedure has the capabilities of developing and validating individualized models capable of predicting patient response to EPO administration. The resulting models can be used as the initial stage of dosage optimization efforts and of minimizing costs and adverse effects.

Keywords: Hemoglobin, Robust Identification, Erythropoietin, Anemia Management, Personalized Medicine

1. Introduction

The universal sequel to End-Stage Renal Disease (ESRD) is anemia. Over ninety percent of ESRD patients undergoing hemodialysis suffer from anemia due to insufficient endogenous production of human erythropoietin, a glycoprotein stimulating agent produced in the bone marrow. Untreated, the consequences of this pathogenesis include a decrease in patient quality of life (QoL), an increase in rate of hospitalization, and an increase in mortality. Until the advent of recombinant Human Erythropoietin (r-HuEPO) over 30 years ago [1], patients suffering from ESRD were treated mainly with multiple blood transfusions. Several complications abound in this treatment procedure including possible infections and transfusion reactions [2]. The discovery of r-HuEPO as an erythropoiesis stimulating agent has since changed ESRD treatment methods.

Noticeable characteristics common among patients of ESRD include short red blood cell life span and iron and vitamin deficiencies; however, it is clear that each ESRD patient lacks the ability to produce sufficient endogenous erythropoietin during homeostasis due to dysfunctional kidneys [3]. Exogenous administration of r-HuEPO is proven to [4, 5, 6] slow or avert the progression of anemia of ESRD. However, the use of such medication creates an added challenge to clinicians including a problem of determining a dosing strategy to achieve the desired effective hemoglobin concentration levels. The National Kidney Foundation-Kidney Disease Outcomes Quality Initiative (NKF-KDOQI) has created guidelines recommending such levels to be maintained between 11 and 12 g/dL [7]; however, the procedure of computing the dose quantity and/or frequency of the medication in order to achieve and maintain patient hemoglobin levels in the effective range remains a trial-and-error process [8]. Consequently, several anemia management facilities have developed their own Anemia Management Protocols (AMP) based on these guidelines which primarily involved trained personnel accessing patients' current hemoglobin levels and past few erythropoietin measures and adjusting the medication accordingly with the hope of hitting the targeted range. Obviously, this trial-and-error approach is not only labor intensive and time consuming but also non-optimal with respect to the cost of the EPO medication. Employing computation tools such as the highly advanced feedback control tools popular in the control engineering field to support medical personnel in this regard will save both time and cost in the dosage optimization efforts.

Several attempts to automate EPO delivery have already been reported in the literature [9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21]. Parametric identification in the Bayesian framework was performed with patient population data in [10]. This approach was subsequently enhanced by Fuzzy rule-based control strategy in [17]. In Gaweda et al. [18], two Artificial Neural Network based approaches, namely Multi-Layer Perceptron (MLP) and Redial Basis Function (RBF), were used and compared in anemia patient model development. It was shown there that the performance of MLP exhibits superiority compared to RBF. Similar to [10], the approach in that study was further enhanced with the Fuzzy rule-based to mainly distinguish responsive and non-responsive

*Corresponding author
Email address: eakabua@gmail.com (Elom Akabua)

patients to EPO medication [19, 20, 21]. In all the above approaches, patient population data were used during the model developments. Few attempts have been made to develop models tailored specifically to individual patients of ESRD [11, 21, 20, 22]. Support vector, specifically profile-dependent support vector, regression approach to the problem was attempted in [22]; however, the procedure was mainly to predict input EPO rather than output hemoglobin. Such objectives is less desirable than predicting hemoglobin output since predicting outputs provide a means of predicting inputs; however, the converse is not true. In [11, 21, 20] individualized anemia management was attempted; however, the approach was aimed at optimizing the EPO dosage rather than developing individualized patient models. Hence, the approach is more of model-free patients hemoglobin control which is different from the model-based approach provided in this study. In certain applications, it is important to have a visual mathematical equation linking output response to system inputs, such as in the case for patients of ESRD.

Many therapeutic solutions are closely related to system control problem and feedback measurements [23, 24, 25]. In the clinical environment, physicians wish to achieve and stabilize the response to certain medication over some desired therapeutic range with minimal effective dosage while avoiding toxicity. This objective is related to feedback control problems. Translating the drug optimization problem into feedback control theory requires three main ingredients: the patient (control process or *plant*), EPO medication (*input*), and the hemoglobin (*output* measure). From control perspectives, developing an individualized patient model capable of predicting effects of EPO dosage on hemoglobin level is indispensable. A mathematical equation describing the relationship between input (EPO) and output (Hb) is called a *model*. The aim is to use robust identification techniques to develop such models using patient specific input-output measurement data, consequently, allowing the design of AMPs using various feedback control methodologies.

The robust identification method employed in this study is different from the classical approach of modeling using input-output data. Major distinctions between the two approaches are that, in the classical settings (i) there is an assumption that the noise affecting measurements is known to posses certain statistical characteristics i.e Gaussian, furthermore, (ii) there is an assumption that the system is of a fixed model structure with unknown parameters [26]. Issues with this approach are that: (i) it assumes a *known* statistical error is affecting the measurements, and (ii) it assumes a *fixed* model structure can explain the system dynamics regardless of its complexity. Robust identification techniques on the other hand make no such assumptions. It respects the fact that the system of interest may be complex and that a single fixed model structure may not be able to define such a complex system. Hence, it assume the system belongs to a certain model *class* (details on this is in the later sections 3) and the noise affecting measurements is assumed unknown but bounded by some value which is usually available and provided by measuring instrument manufacturers. The result is a model set consisting of models able to explain the *a posteriori* measurement data with the given *a priori* assumptions. The center of this model set can now be considered as the nominal model.

Finally, we previously introduced robust identification procedures to individualized anemia patient model discovery in [27]. In that paper, there was no consideration for the fact that the system (patient) was already in progress prior to identification and that the initial conditions are not zero at t_0. Ignoring such information could lead to artificially high model prediction error. The aim of this paper is to incorporate such information into the model discovery process by encapsulating the effect of basal patient erythropoiesis into a single variable to minimize model prediction error. We refer to this approach as *semi-blind* robust identification [28].

Organization of the paper is as follows. In section 2, we provide an overview of anemia management problems while pointing out a few of its challenges especially regarding individualized model discovery. Section 3 provides brief overview on robust identification, especially in ℓ_1 identification framework, subsequently introducing *semi-blind* robust identification procedures in the same section. We combine the robust identification techniques developed in section 3 with anemia management problems in section 2 to produce nominal individualized model in section 4. Section 5 concludes the paper.

2. Anemia Management Problem

2.1. Introduction

At steady state, about 10^{10} new red blood cells (RBC) are synthesized per hour in healthy individuals during erythropoiesis process to maintain hemoglobin levels within a specified range [29]. RBC is an oxygen carrying agent to tissues and other vital organs in the body including the brain and the heart. Besides sensing oxygen availability to tissues, the kidney is also responsible for releasing erythropoietin into circulation [4]. For patients suffering from anemia of ESRD, such benefits the kidney offers are lacking due to dysfunctional kidneys. Prior the 1990s, treatment of such disease relied on iron supplements, repeated blood transfusions, and occasional androgen therapy. While blood transfusion was popular among the treatment methods, it comes with several demerits such as infections and possible iron overload. The discovery of recombinant human EPO (rHuEPO) has since shifted the treatment of End-Stage Renal Disease (ESRD) patients from blood transfusions to rHuEPO therapy. A major challenge facing rHuEPO therapy is the means of determining optimal dosing strategy to maintain patient hemoglobin concentration within the required target range. Interpatient variabilities culminated with intraindividual variabilities

makes the problem even more challenging. Further, the high cost of the medication makes clinicians take judicious approach to the dosage therapy.

2.2. Definition

Ultimately, the desire of anemia management protocol (AMP) is to stabilize the hemoglobin levels of patients of ESRD in the prescribed range of 11 to 12 g/dL. This narrow therapeutic range was established and recommended by the National Kidney Foundation-Kidney Disease Outcomes Quality Initiative (NKF-KDOQI) [7], a body designated for such tasks. To achieve this goal, several dialysis

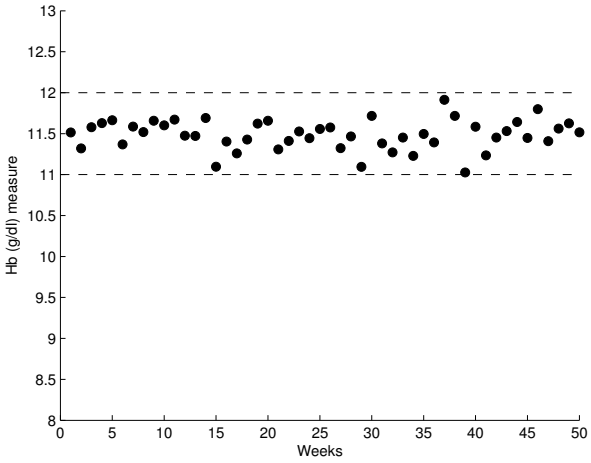

Figure 1: Ideal Hb behavior after EPO therapy where patient hemoglobin measurements fall within the targeted range of 11-12 g/dL.

facilities developed their own protocols for EPO dosage adjustments [30]. However, most of these protocols are based on patient population responses and clinicians' past experiences. Consequently, such approach often leads to unstable hemoglobin levels in certain patients due to the fact that these patients do not fit very well such a generalized structure. Figure 1 shows the desired ESRD patient response to EPO, however, what is often observed in typical patients is shown in Figure 2 where hemoglobin levels fluctuates around the target therapeutic range of 11-12g/dL.

A challenge to EPO therapy in anemia of ESRD is maintaining patient hemoglobin response within the targeted therapeutic range. A phenomenon known as hemoglobin cycling occurs where patients overshoot and undershoot the targeted hemoglobin range even with average administered dosage. Over 90% of ESRD patients suffer from this phenomenon during EPO therapy [31]; perhaps since most hemodialysis facilities apply a general population model to individual patients. Developing patient specific models using past input-output measurement data and using the benefits offered in advanced feedback control fields could minimize such phenomenon consequently stabilizing patient hemoglobin levels within the desired range.

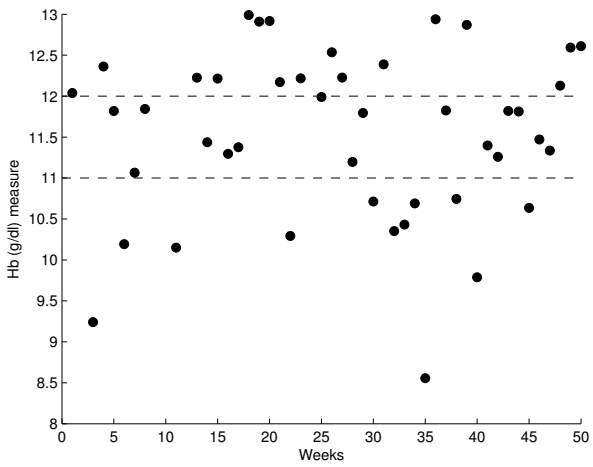

Figure 2: Realistic Hb behavior after EPO therapy where patient's hemoglobin measurements widely overshoots and undershoots the targeted range of 11-12 g/dL.

2.3. Control Problem

Figure 3 shows a typical interaction between physician and ESRD patient where physician attempts to stabilize patient hemoglobin level within some desired hemoglobin (Hb_d) range. However, measurement noise and other known/unknown factors make this objective a challenging affair. In this figure, Hb_d is the desired hemoglobin level a Physician wishes to achieve by administering erythropoietin (EPO) to the Patient. The amount of EPO to administer is determined based on some Control output, either through the Physician's past experience or some optimization based algorithm. Meal intakes, exercises, and other medications individual patients used can all be classified under *disturbance* while errors that occur during hemoglobin measurements are considered as *noise*. Hence Hb and Hb_m are defined as the actual and the measured hemoglobin levels, respectively. Clearly, this process can be viewed as a feedback control problem in a closed-loop where the goal is to stabilize dosage outcome within a certain therapeutic range. Using feedback system in drug optimization therapy has long been recognized in the literature [23]. Basic requirements for such an approach include:

(i) a clear design objective (stabilize hemoglobin level within the target range of 11 to 12g/dL)
(ii) controller (algorithm determining the optimal input EPO to the patient)
(iii) plant (patient)
(iv) plant model– a mathematical equation relating input-output (EPO to Hb)
(v) output (Hb response) measurements.

Clearly, the only missing piece of information and arguably the most important one to completely translate the anemia management problem to the feedback control framework is the mathematical model of the patient relating

EPO to Hb response. Developing a methodology capable of producing individualized patient models will complete the needed requirements to fully translate AMP to feedback control problem.

In personalized therapy, it is often necessary to be able to develop individualized model from small measurement samples. This is a major challenge in the anemia management problem since very often only few weeks of individualized patient data are available for analysis. Such small sample set is insufficient [18] for individualized model derivation using the existing approaches of model discovery primarily due to the statistical assumptions made on measurement data [32]. Consequently, patient population data are used. The robust identification algorithm employed here makes no assumption on the nature of noise affecting measurements nor assumptions on the order of the system; hence, it renders the capabilities of developing individualized models with small dataset (as demonstrated in section 4). It should be mentioned that model-free based control algorithms are available and can be used [11] to achieve similar objectives of stabilizing hemoglobin levels. However, such approach lacks an important benefit of providing a clear mathematical relations linking EPO to Hb response as provided in model based approaches.

3. Robust Identification

3.1. Introduction

We provide a brief introduction to robust identification while emphasizing the distinctions between the approach and that of classical methods of system identification.

3.2. System identification

The conventional system identification approach to model discovery such as the AutoRegressive with Exogenous (ARX), Prediction Error Estimate (PEM), Output Error (OE), etc. (see [26] for details) involve

(i) assumption on fixed mathematical model structure with known order and unknown parameters
(ii) definition of performance measure
(iii) use of measured data to estimate the unknown parameters
(iv) validation of the model with unseen data.

Usually, the least squares techniques are used to determine these unknown parameters. However, least squares methods are known to be sensitive to outliers, hence the use of such procedures is intrinsically non-robust. A troubling issue with this approach is the assumption that the error affecting measurements is *known* to be of certain statistical characteristics, usually of white Gaussian noise nature. Even more troubling is the assumption that a fixed mathematical model structure can explain the system regardless of its complexity. When such statistical information on the system is questionable, and/or when a single prescribed model structure is unrealistic, robust deterministic identification approach is a sound alternative.

3.3. Robust Identification

The robust identification technique accounts for the fact that information available on the system of interest is finite, partial, and corrupt hence restraining its model to a fixed structure may be erroneous or short lived. The technique relies on both functional analysis and interpolatory theory as well as the theory and concepts of information-based complexity (IBC), thoroughly studied by Traub (see [33] and references therein) where the goal is to determine the minimal computation cost to a given problem. Information is *partial* because there is no complete information on the problem at hand, *finite* since measurement data is finite, and *corrupt* to reflect the fact that the available measurement is corrupted in a way (e.g. by noise). This was initially realized and utilized by Helmicki [34] followed by many others [35, 36, 37] when the conventional approaches to model discovery using input-output measurements failed to provide the *hard* bound requirement necessary for modern robust controller synthesis.

An attractive feature of the robust identification method is the minimal *a priori* requirements on the system. There is no assumption on the model order nor statistical assumptions on the noise affecting output measurements. The only assumption is that the system belongs to a particular *class* and its measurement noise is *unknown but bounded* by a known value, usually provided by measuring device manufacturer. The aim is to obtain a suitable nominal model concurrent with a *hard* uncertainty bound using the *a posteriori* information (experimental input-output data) and the given *a priori* information. Results of the proposed approach can be visualized in Figure 4, where the center of the circle represents the nominal model while its radius represents the uncertainty in the model. The derived nominal model can be readily used in any advanced controller design synthesis such as adaptive control, predictive control, intelligence control, and more importantly, robust control which requires a nominal model as well as *hard* bound on model uncertainty. It should be emphasized that the uncertainty bound obtained is a culminated uncertainty due to measurement noise as well as uncertainty in the model itself.

To continue the discussion on the subject, let us establish a few needed notations. Of interest are the subset of functions $\mathcal{H}_\infty$ in $\mathcal{L}_\infty$ which are analytically continuous inside the unit disk equipped with the norm $\|G\|_\infty \doteq ess\ sup_{|z|<1}\ \overline{\sigma}(G(z))$ where $\mathcal{L}_\infty$ are Lebesgue functions and $\overline{\sigma}(\mathbf{G}(z))$ is the maximum singular value of system $G(z)$. For $G(z) \in \mathcal{H}_\infty$, define $G(z) \in \mathcal{BRH}_\infty$ as bounded real rational transfer functions in $\mathcal{H}_\infty$. Also of interest is space of transfer matrices analytic in the $|z| \le \rho$ equipped with the norm $\|G\|_{\infty,\rho} \doteq ess\ sup_{|z|<\rho}\ \overline{\sigma}(G(z))$ defined as $\mathcal{H}_{\infty,\rho}$. The right-sided Z-transform of real sequence $\{x\}$ is defined as $\mathcal{X}(z) = \sum_{i=0}^{\infty} x_i z^i$.

Given an input-output data sequence, we define an operator G as a function mapping the input sequence to the

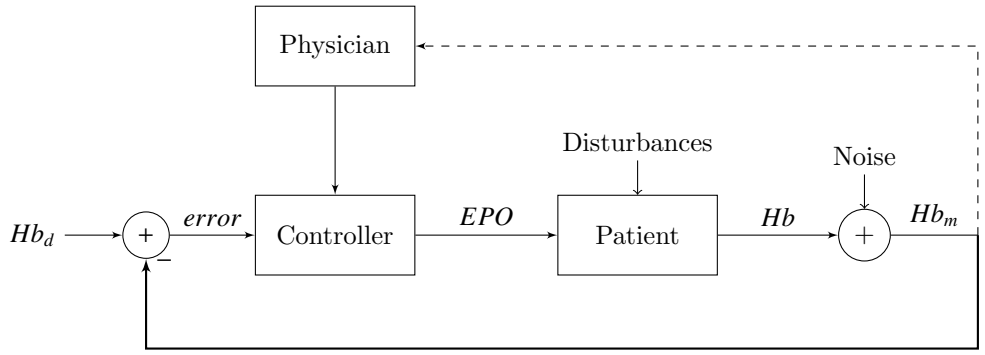

Figure 3: Closed-loop feedback system representing Patient and Physician interaction with erythropoietin and hemoglobin where Hb_d is the desired hemoglobin level and Hb_m is the measured hemoglobin level.

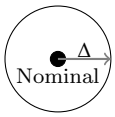

Figure 4: Pictorial view of nominal model (center dot) with bounded uncertainty (radius Δ)

output. This representation can be defined either as a (rational) complex valued function:

$$G(z) \doteq \sum_{i=0}^{\infty} g_i z^i$$

or as a minimal state-space realization:

$$G \equiv \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right)$$

In sequel, for stable system G, we denote $T_G : \ell^\infty[0,\infty) \to \ell^\infty[0,\infty)$ the Toeplitz matrix and $\Gamma_G : \ell^\infty(-\infty,-1] \to \ell^\infty[0,\infty)$ the Hankel operator associated with G.

The following result is used to establish the existence of a linear-time invariant (LTI) systems subject to the given *a priori* information.

Theorem 1. *(Carathéodory-Fejér). Given a matrix valued sequence $\{L_i\}_{i=0}^{N-1}$, there exists a causal, discrete-time, LTI operator $L(z) \in \mathcal{BH}_\infty$ such that*

$$L(z) = L_0 + L_1 z + L_2 z^2 + \cdots + L_{N-1} z^{N-1} + \ldots \qquad (1)$$

if and only if $(T_L^N)^T T_L^N \leq I$ [38]

where $L(z) \in \mathcal{BH}_\infty$ defines bounded functions analytic in the unit circle while (T_L^N) and $(T_L^N)^T$ are, respectively, the Toeplitz and its conjugate transpose and I is the identity matrix with the appropriate dimensions.

3.4. The Model

The aim is to derive a discrete linear-time invariant (LTI) model relating hemoglobin concentration to erythropoietin administration using the available *a priori* and *a posteriori* information. The available *a priori* information on the model consists of three positive constants: $K > 0$, $\rho > 1$, and $\epsilon > 0$, which respectively are, the upper bound on a certain gain associated with plant, the lower bound on the relative stability margin of the plant, and the upper bound on measurement noise. From the time domain measurement data, the Carathéodory-Fejér interpolation theory of Theorem 1 is used to establish such models. The aim is to determine the necessary and sufficient conditions for there to exist a function $h \in \mathcal{BRH}_\infty$ such that

$$h(z) = h_0 + h_1 z + h_2 z^2 + \ldots + h_n z^n + z^{n+1} h^n(z) \qquad (2)$$

for some $h^n(z) \in \mathcal{BRH}_\infty$, where $h \in \mathcal{BRH}_\infty$ belongs to the *class* of real valued analytic functions bounded in the unit circle. Then from [38], such a function, h, exists if and only if:

$$I - HH^* \geq 0 \qquad (3)$$

where H is the Toeplitz matrix defined in terms of impulse response h, H^* is its conjugate transpose, and I is the identity matrix with proper dimensions [39].

With the *a priori* assumptions, it is assumed that the system generating the *a posteriori* experimental data belongs to the following system class of models, $\mathcal{S}$, such that

$$\mathcal{S} \doteq \{G(z) = H(z) + P(z)\} \qquad (4)$$

where $H(z) \in \mathcal{BH}_{\infty,\rho}(K)$ for some $\rho > 1$ representing the nonparametric and $P(z)$ the parametric portion of the system, respectively. Furthermore, we assume the parametric portion of the system, $P(z)$, belongs to some affine set $\mathcal{P}$ of the form:

$$\mathcal{P} \doteq \{P(z) = p^T G_p(z),\ p \in \mathcal{R}^{N_P}\} \qquad (5)$$

where $p \in \mathcal{R}^{N_P}$ are some unknown vectors of some *a priori* known component G_p and N_p is the number of unknown p parameters to be determined.

The following lemma is used to establish the existence of a family of systems in $\mathcal{S}$.

***Lemma* 1.** *Given K, ρ, and two other vector sequences (u,y) representing input and output measurements, respectively, there exists an operator $S \in \mathcal{S}$ such that* $\mathbf{y} = \mathcal{S}\mathbf{u}$ *if*

and only if there exists a vector $\mathbf{h}$ *such that the following conditions are satisfied [40] :*

$$M(\mathbf{h}) = \begin{bmatrix} KR^{-2} & (T_h^N)^T \\ T_h^N & KR^2 \end{bmatrix} \geq 0$$
$$|y - (T_u^N pP + T_u^N h)| \leq \epsilon \tag{6}$$

where $(P)_k \doteq [g_k^1 \; g_k^2 \; \dots \; g_k^{N_p}]$, g_k^i denotes the k^{th} Markov parameter of the i^{th} transfer function of the parametric portion and h_k are the Markov parameters of the nonparametric portion. T_u^N defines the Toeplitz matrix of N inputs $\mathbf{u}$, T_h^N defines the Toeplitz matrix corresponding to the non-parametric portion of the system response $\mathbf{h}$, while

$$R = \text{diag}\,[1 \; \rho \; \rho^2 \; \dots \; \rho^{N-1}]$$

is the diagonal matrix formed by the stability margin, ρ. Since the model is assumed to be in the *a priori* model class, $\mathcal{S}$, containing several other models capable of generating the measurement data with the assumed bounded noise, we can determine the nominal model by parameterizing the set $\mathcal{S}$ with a free parameter $Q(z) \in \mathcal{BH}_\infty$. A simple such parameter of choice is the zero parameter, thus $Q(z) = 0$ leads to the central model $S_{\text{central}} = H_0(z) + \mathbf{p}^T \mathbf{G}_p(z)$ where explicit state space realization of $H_0(z)$ can be determined [41].

3.5. Semi-blind robust identification

For individualized ESRD erythropoiesis therapy, it is our hypothesis that accounting for the effects of patients initial conditions can significantly enhance the predictive ability of individualized models. In the robust identification process introduced in section 3.3, we identify the system G by assuming that the experimental data are generated exactly from some initial time t_0 while ignoring the effects of basal erythropoiesis inputs prior to t_0. Neglecting such information could lead to artificially high prediction error as seen in our previous work [27]. To account for the effects of initial conditions, we consider a recently developed *semi-blind* robust identification [28]. For the purpose of this study, we encapsulate the effect of prior inputs (erythropoiesis) on the system into some unknown non-zero variable $\mathbf{x}_0$. Subsequently by evoking Lemma 1, equation (6) can be rewritten as

$$M(\mathbf{h}) = \begin{bmatrix} KR^{-2} & (T_h^N)^T \\ T_h^N & KR^2 \end{bmatrix} \geq 0$$
$$|y - (T_u^N pP + T_u^N h + \mathbf{x_0})| \leq \epsilon$$
$$|\mathbf{x_0}| \leq K_G K_{u^-} \tag{7}$$

where K_G is the initial conditions system gain, K_{u^-} is the maximum initial input and are interpreted in a componentwise [28].

4. Application and Results

In the literature, ESRD patients are classified as either *normal responders* or *poor responders* to EPO therapy [11, 17, 42]. Normal responder patients are those that achieve the targeted hemoglobin level with an average EPO dosage while poor responder patients are defined as those that receive substantially high amount of EPO medication yet, very little change is noticeable in their hemoglobin levels. Predicting the hemoglobin values of these poor responder groups is a major challenge to clinicians since it is not exactly known what constitutes such behavior in these patient groups. By inspecting the graphical behaviors of each patient data obtained from the University of Louisville Chronic Dialysis Unit, we identified certain patients as medically *interesting*. These patients exhibit behavior similar to or worse than those of *poor responders* patients. In addition to the behavior of *poor responders*, these patients' hemoglobin levels widely overshoot and undershoot the targeted range with little or no change in EPO administration.

To utilize the semi-blind identification framework proposed in section 3.5, we need to characterize the possible past input values into a set, $\mathcal{U}_-$. Such set contains all possible EPO (erythropoiesis) values prior to the identification process. In principle, we could assume such a set contains values less than or equal to the maximum possible EPO value ($u^- \in \mathcal{U}_-$, where $u^- \leq |u|_{max}$); however, such assumption is too coerced and too conservative since the range of EPO dosage varies widely depending on the period of therapy. We can obtain a much tighter bound by considering the effects of change in the EPO to hemoglobin levels; consequently the problem becomes identifying an operator S mapping the *change* in erythropoietin, $u_k = Epo_k - Epo_{k-1}$ to hemoglobin Hb_{k+1}. With the above assumption, it should be realized that the *a priori* information on the system should therefore include an integrator in the system, subsequently contributing to the parametric portion of the system. Figure 5 shows such an integrator contribution to the system where $\mathbf{u_k}$ is the difference in the current and previous EPO dosage and $\mathbf{Hb_{k+1}}$ is the effect of such difference on the patient and p is the constant unknown parameter of the parametric portion of the model $P(z)$. To improve the optimization process of the identification procedure, the input data were uniformly rescaled to the interval [-1,1].

The following *a priori* information on both the system and the noise are used in the study:

$$\begin{aligned} &\mathcal{S} = \{G(z) = H(z) + P(z), \quad P(z) = p\tfrac{z}{z-1}, \quad H \in \mathcal{BH}_{\infty,\rho}(K)\} \\ &\mathcal{N} = \{\eta \in \ell^\infty : |\eta| \leq \epsilon\}, \quad \epsilon = 0.31 \\ &\mathcal{U}_- = \{u \in \ell^\infty : |u_k| \leq u_{max}\} \end{aligned} \tag{8}$$

where $P(z)$ and $H(z)$ are the parametric and nonparametric contributions to the system, respectively. The bound on measurement error of $\epsilon = 0.31$ is determined based on informations obtained on hematocrit measuring device used

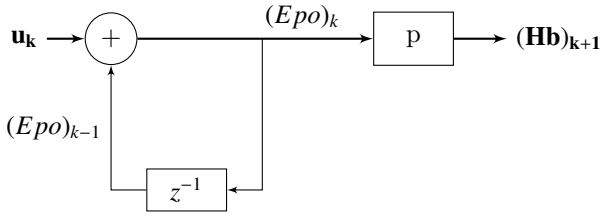

Figure 5: Representation of an integrator where $\mathbf{u_k}$ is the input to the system. It is the change in EPO level while $\mathbf{(Hb)_{k+1}}$ is the output hemoglobin as effected by change in input EPO input.

Patient#	K	ρ	N_t	Respondent Group
1	0.295	1.0604	15	×
28	0.280	1.0900	17	×
30	0.266	1.0700	18	×
6	0.280	1.0214	15	✓
37	0.176	1.0870	15	✓

Table 1: Some *a priori* parameter information for the selected patients of *interesting*, patients #1, #28, and #30 and *normal responder*, patients #6 and #37. N_t represents the number of data points used in the identification and we indicated with a check mark whether patient belongs to a responder or non-responder groups.

at the dialysis facility [43]. Table 1 summarizes the remaining *a priori* information on the selected patients. The gain, K, is determined based on the LMIs of Equation (7), the decay rates, ρ, are heuristically determined to obtain reasonable results. Though there may exist some optimization algorithm to determine the optimal decay rate values, ρ, such procedure is not known to us at the time of this study.

Fifty-six ESRD patient data were obtained from the University of Louisville Chronic Dialysis Unit for analysis. Typically, each patient data consists of 52 weekly hemoglobin measurements with 0-3 times per week of EPO administration. However, six of the patients have data less than 15 either due to missing appointments or receiving kidney transplants and they were subsequently omitted in this study. This is a common occurrence in medical studies. Using the aforementioned robust identification in conjunction with the *semi-blind* robust identification procedures, we obtained models ranging from first to third order models for the remaining 50 patients. Due to space limitations, we show the results for five patients: three of which are of medically *interesting* groups (Figures 6, 8, and 7) and two patients of *respondent* groups (Figures 9 and 10). A one-step-ahead prediction is performed in predicting the hemoglobin levels in each case.

Running the proposed algorithm for few weekly measured hemoglobin data with the corresponding erythropoietin and after performing model reduction by eliminating uncontrollable and unobservable states, we obtained the following transfer function equations for the selected patients.

$$G^1(z) = \frac{1.03z^3 + 1.398z^2 + 0.8412z - 0.04125}{z^3 + 0.3172z^2 - 0.501z - 0.8161} \tag{9}$$

$$G^{28}(z) = \frac{0.5161z^2 - .3610z + 0.472}{z^2 - .368z + .632} \tag{10}$$

$$G^{30}(z) = \frac{1.018z + 0.2443}{z - 1} \tag{11}$$

$$G^{6}(z) = \frac{0.8243z^2 + 0.8323z + 0.1615}{z^2 - 0.1139z - 0.8861} \tag{12}$$

$$G^{37}(z) = \frac{1.046z + 0.08025}{z - 1} \tag{13}$$

Using the time delay properties of the z-transform $\mathcal{Z}^{-1}\{z^n G(z)\} = g_{k+n}$, Equations (9) to (13) are equivalently represented as follows:

$$\begin{aligned} Hb^1_{k+3} = \; & -0.3172Hb_{k+2} + 0.501Hb_{k+1} + 0.8161Hb_k \\ & +1.03u_{k+3} + 1.398u_{k+2} + 0.8412u_{k+1} - 0.04125u_k \end{aligned} \tag{14}$$

$$\begin{aligned} Hb^{28}_{k+2} = \; & 0.368Hb_{k+1} - 0.632Hb_k \\ & +0.5161u_{k+2} - 0.3610u_{k+1} + 0.472u_k \end{aligned} \tag{15}$$

$$Hb^{30}_{k+1} = Hb_k + 1.018u_{k+1} + 0.2443u_k \tag{16}$$

$$\begin{aligned} Hb^{6}_{k+2} = \; & 0.1139Hb_{k+1} + 0.8861Hb_k \\ & +0.8243u_{k+2} + 0.8323u_{k+1} + 0.1615u_k \end{aligned} \tag{17}$$

$$Hb^{37}_{k+1} = Hb_k + 1.046u_{k+1} + 0.08025u_k \tag{18}$$

where Hb_k and u_k are the hemoglobin and the change in EPO at time k, respectively.

4.1. Performance measure

The forecasting power of the proposed approach can be seen in Figures 6 to 10 for the selected patients of #1,#28, #30, #6, and #37. Figure 6 shows hemoglobin prediction of patient #1 to the change in erythropoietin using Equation (14). In this Figure, "+" denotes the experimental data used for the identification while "*" denotes additional experimental data used for validation purposes. The predicted output of the model is denoted by "o". In addition, we use a dotted vertical line to demarcate data used for identification and those used for validation. As shown there, the model is able to predict very well considering only 15 data points are used to generate the model. Similarly, Figure 8 shows the response of patient #30 using Equation (15) with 18 input-output data points used for determining the model. Figure 9 shows similar results with 15 data points used for creating the model for patient #6. The number of data points used for the identification is selected heuristically for optimal predicted results; however, since the procedure is interpolatory, increasing the number of data points used in the identification improves on the predicted results. For each of the selected patients, the maximum prediction error defined in Equation (19)

Patient#	$error_{max}$	$error_{rms}$
1	1.07	0.28
28	0.68	0.37
30	1.35	0.40
6	0.65	0.26
37	0.99	0.35

Table 2: Model maximum and root mean square error for the selected patients

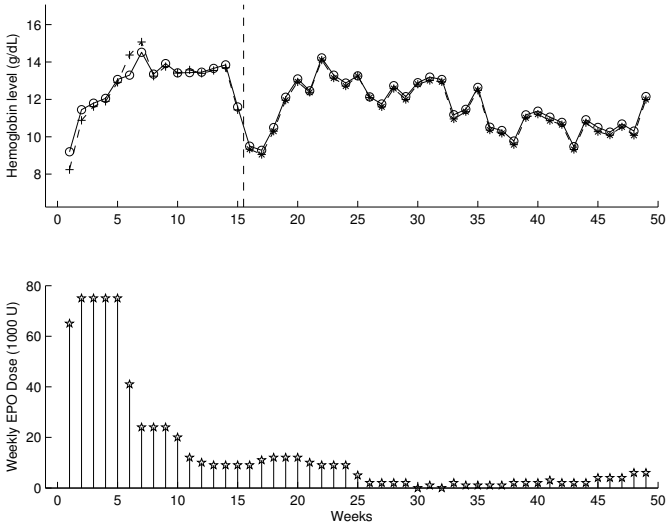

Figure 6: Patient #1 [Top] Predicted results with vertical demarcated line separating identification and prediction data. (+) denotes experimental data used in the identification while, (*) denotes additional experimental data not used in the identification. (o) indicates predicted output with 3^{rd} order model. [Bottom] Administered erythropoietin dose in x1000 U per week.

as well as the average error defined in Equation (20) are determined,

$$|error|_{max} = |y_{data}(i) - y_{pred.}(i)|_{\infty} \tag{19}$$

$$error_{RMS} = \sqrt{\frac{\sum_{i=1}^{N}(y_{data}(i) - y_{pred.}(i))^2}{N}} \tag{20}$$

where y_{data} and $y_{pred.}$ are the actual hemoglobin data and model predicted values, respectively. Table 2 provides a summary on both the maximum prediction errors and average errors as well as the number of data points used for the model identification for each of the selected patients. While the maximum prediction error values may seem high, especially for the *interesting* patients, it should be emphasized that there were no patient specific attributes (weights,race, gender, etc.) considered when determining these models. From the root means square prediction error, however, one can conclude those error values are acceptable given the width of the target range (11-12 g/dL).

5. Conclusion

In this paper, the problem of developing individualized patient specific model for anemic of End-Stage Renal Disease is attempted. Using robust identification combined

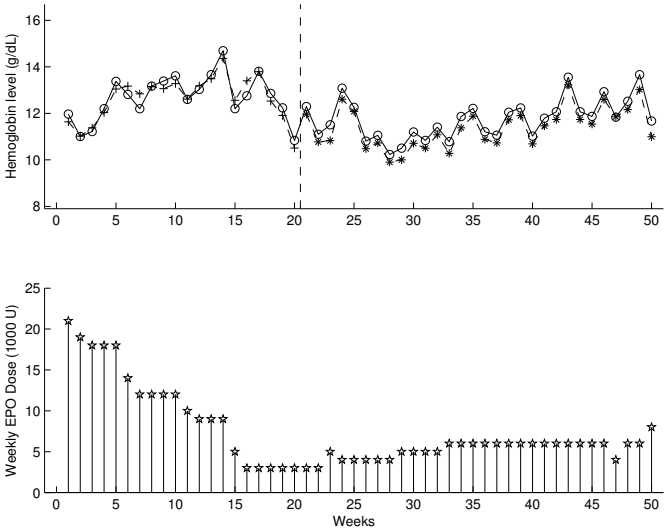

Figure 7: Patient #28 [Top] Predicted results with vertical demarcated line separating identification and prediction data. (+) denotes experimental data used in the identification while, (*) denotes additional experimental data not used in the identification. (o) indicates predicted output with 2^{nd} order model. [Bottom] Administered erythropoietin dose in x1000 U per week.

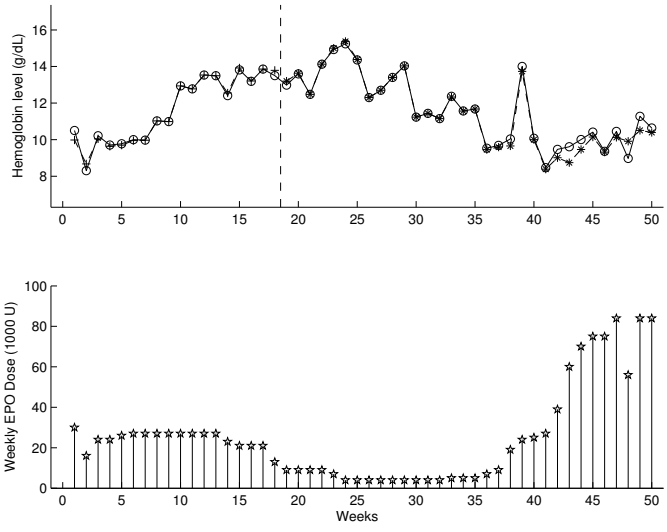

Figure 8: Patient #30 [Top] Predicted results with vertical demarcated line separating identification and prediction data. (+) denotes experimental data used in the identification while, (*) denotes additional experimental data not used in the identification. (o) indicates predicted output with 1^{st} order model. [Bottom] Administered erythropoietin dose in x1000 U per week.

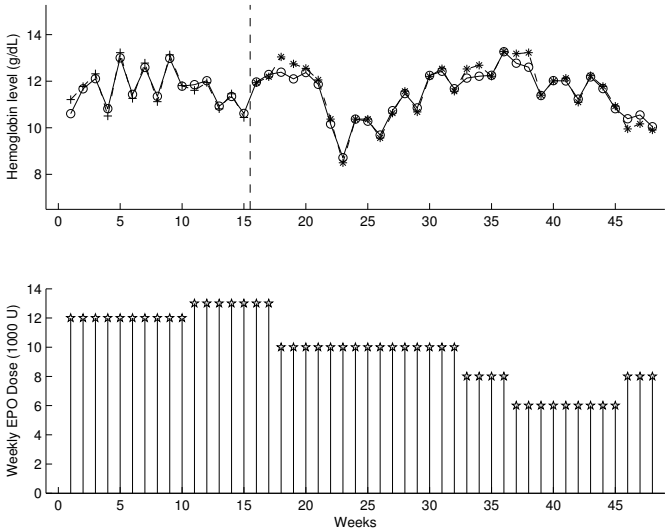

Figure 9: Patient #6 [Top] Predicted results with vertical demarcated line separating identification and prediction data. (+) denotes experimental data used in the identification while, (*) denotes additional experimental data not used in the identification. (o) indicates predicted output with 2^{nd} order model. [Bottom] Administered erythropoietin dose in x1000 U per week.

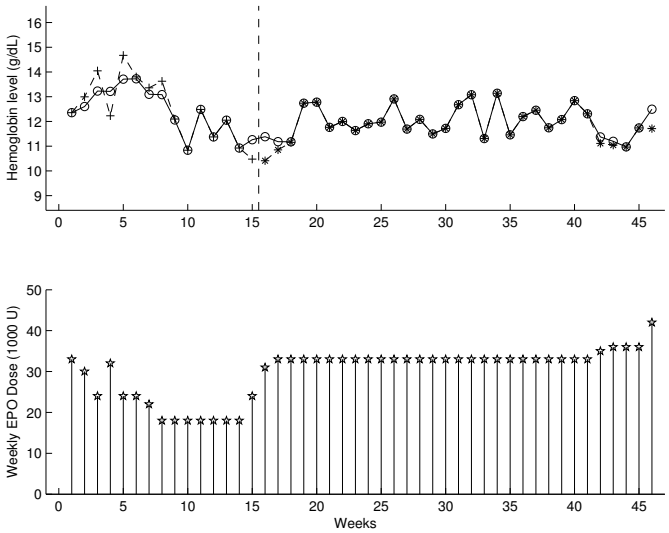

Figure 10: Patient #37 [Top] Predicted results with vertical demarcated line separating identification and prediction data. (+) denotes experimental data used in the identification while, (*) denotes additional experimental data not used in the identification. (o) indicates predicted output with 1^{st} order model. [Bottom] Administered erythropoietin dose in x1000 U per week.

with the recently developed *semi-blind* robust identification techniques, the procedure provides a dynamical model capable of predicting individualized response to EPO medication. A benefit of the methodology is that it requires a very minimal dataset and *a priori* information in developing the models, which makes it useful in clinical settings where, often times, very little information is known about a given patient. For each of the selected patients, less than 20 weekly data were used to produce each patient model Figures 6,8, 9, 7, and 10 show the predictive ability of the model developed for the selected patients. Results for each patient model showed an acceptable predicting error, even though other patient attributes such as weight, demography, race, etc. were not considered in the study. The obtained result is important because it suggests that the robust identification techniques can be used in clinical environments to enhance personalized drug dosage optimization efforts in dialysis units.

REFERENCES

[1] D. M. Spiegel. Anemia Management in Chronic Kidney Disease : What Have We Learned after 17 Years ? Seminars in Dialysis, 19(4):269–72, 2006.

[2] T. Tjornes. Response and prediction of response to recombinant human erythropoietin in patients with solid tumors and platinum-associated anemia. Journal of Oncology Pharmacy Practice, 5(1):22–31, 1999.

[3] S. Nurko. Anemia in chronic kidney disease: Causes, diagnosis, treatment. Cleveland Clinic of Medicine, 73(3), March 2006.

[4] J. Adamson. The erythropoietin-hematocrit relationship in normal and polycythemic man: implications of marrow regulation. Blood, 32(4):597–609, Oct 1968.

[5] J. W. Eschbach and J. W. Adamson. Anemia of end-stage renal disease (ESRD). Kidney International, 28(1):1–5, July 1985.

[6] F. Falderrabano. Erythropoietin in chronic renal failure. Kidney International, 50(4):1373–91, Oct 1996.

[7] NKF-QOQI. Clinical practice guidelines and clinical practice recommendations for anemia in chronic kidney disease. Am J Kidney Dis, 30(4):S192–S240, Oct 1997.

[8] E. F. Unger, A. M. Thompson, M. J. Blank, and R. Temple. Erythropoiesis-stimulating agents–time for a reevaluation. New England Journal of Medicine, 362(3):189–192, 2010. PMID: 20054037.

[9] R. Bellazzi, C. Siviero, and R. Bellazi. Mathematical modeling of erythropoietin therapy in uremic anemia. does it improve cost-effectiveness? Medical decision making and problem solving in hematology, 79:154–164, 1994.

[10] R. Bellazzi. Drug Delivery Optimization through Bayesian Networks: An Application to Erythropoietin Therapy in Uremic Anemia. Computers and Biomedical Research, 26(3):274–293, June 1993.

[11] M. K. Muezzinoglu. Approximate Dynamic Programming for Anemia management. PhD thesis, University of Louisville, March 2006.

[12] D. E. Uehlinger, F. A. Gotch, and L. B. Sheiner. A pharmacodynamic model of erythropoietin therapy for uremic anemia. Clinical Phamacology and Therapeutics, 51:76–89, 1992.

[13] D. Westwick and R. Kearney. Identification of physiological systems: a robust method for non-parametric impulse response estimation. Medical & Biological Engineering & Computing, 35(2):83–90, March 1997.

[14] R. Bellazi. Drug delivery optimization through bayesian networks. AMIA, pages 572–578, 1993.

[15] K. G. Gadkar, R. Gunawan, and F. J. D. III. Iterative approach

to model identification of biological networks. BMC Bioinformatics, 6, june 2005.

[16] J. M. Bailey and W. M. Haddad. Drug dosing control in clinical pharmacology. IEEE Control Systems Magazine, April 2005.

[17] R. Bellazzi, C. Siviero, and R. Bellazzi. Mathematical modeling of erythropoietin therapy in ure- mic anemia. does it improve cost-effectiveness? Hemoglobin, pages 154–164, 1994.

[18] A. E. Gaweda, A. Jacobs, M. E. Brier, and J. M. Zurada. Pharmacodynamic population analysis in chronic renal failure using artificial neural networks–a comparative study. Neural networks : the official journal of the International Neural Network Society, 16(5-6):841–5, 2003.

[19] A. Gaweda, A. Jacobs, G. Aronoff, and M. Brier. Intelligent control for drug delivery in management of renal anemia. In Machine Learning and Applications, 2004. Proceedings. 2004 International Conference on, pages 355–359, 2004.

[20] A. Gaweda, A. Jacobs, and M. Brier. Application of fuzzy logic to predicting erythropoietic response in hemodialysis patients. The International journal of artificial organs, 31(12):1035, 2008.

[21] A. E. Gaweda, M. K. Muezzinoglu, G. R. Aronoff, A. A. Jacobs, J. M. Zurada, and M. E. Brier. Individualization of pharmacological anemia management using reinforcement learning. Neural Networks, 18:826–834, 2005.

[22] J. D. Martin-Guerrero, G. Camps-Valls, E. Soria-Olivas, A. J. Serrano-Lopez, J. J. Perez-Ruixo, and N. V. Jimenez-Torres. Dosage individualization of erythropoietin using a profile-dependent support vector regression. IEEE Transactions on Biomedical Engineering, 2003.

[23] S. Vozeh and J.-L. Steimer. Feedback Control Methods for Drug Dosing Optimization: Concept, Classification, and Clinical Application. Clinical Pharmacokinetics, 10:457–476, 1985.

[24] D. Linkens and S. S. Hacisalihzade. Computer control systems and pharmacological drug administration: a survey. Journal of medical engineering & technology, 14(2):41–54, 1990.

[25] S. S. Hacisalihzade. Control engineering and therapeutic drug delivery. IEEE Control Systems Magazine, June 1989.

[26] L. Ljung. System identification: theory for the user. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition, 1999.

[27] E. Akabua, T. Inanc, A. Gaweda, M. E. Brier, S. Kim, and J. M. Zurada. Robust identification approach to individualized anemia model. Journal of Applied Computer Science Methods & Mathematics, 3(2):65–75, 2011.

[28] W. Ma. Semi-Blind Robust Identification and Model (in)validation. PhD thesis, The Pennsylvania State University, December 2007.

[29] A. C. Guyton and J. E. Hall. Textbook of Medical Physiology. Elsevier Sanders, 11 edition, 2005.

[30] A. Gaweda. Improving management of anemia in end stage renal disease using reinforcement learning. In Neural Networks, 2009. IJCNN 2009. International Joint Conference on, pages 953–958, 2009.

[31] S. Fishbane and J. S. Berns. Hemoglobin cycling in hemodialysis patients treated with recombinant human erythropoietin. Kidney International, 68:1337–1343, 2005.

[32] L. Ljung. Model Validation and Model Error Modeling Model Validation and Model Error Modeling, 1999.

[33] J. Traub, G. Wasilkowski, and H. Wozniakowski. Information-Based Complexity. Academic Press, New York, 1988.

[34] A. Helmicki, C. Jacobson, and C. Nett. Control oriented system identification: a worst-case/deterministic approach in H_∞. Automatic Control, IEEE Transactions on, 36(10):1163 –1176, oct 1991.

[35] M. C. Mazzaro, P. A. Parrilo, and R. S. S. Pena. Robust identification:an approach to select the class of candidate models. International Journal of Control, 74(12):1210–1218, 2001.

[36] A. Onatski and N. Williams. Modeling model uncertainty. Int. seminar on macroeconomics, European Central Bank, 08 2002.

[37] J. Schoukens, Y. Rolain, and R. Pintelon. On the use of parametric and non-parametric noise-models in time- and frequency domain system identification. In Decision and Control (CDC), 2010 49th IEEE Conference on, pages 316 –321, dec. 2010.

[38] J. Chen and G. Gu. Control Oriented System Identification: An H_∞ Approach. John Wiley & Sons, Inc., New York, first edition, Jun 2000.

[39] T. Inanc. Mixed ℓ_1/H_∞ robust identification: Application to a flexible structure testbed. Master's thesis, Pennsylvania State University, 1996.

[40] P. A. Parrilo, R. S. S. Pena, and M. Sznaier. A parametric extension of mixed time/frequency robust identification. Automatic Control, IEEE Transactions on, 44(2):364–369, 1999.

[41] W. Ma, M. Yilmaz, M.Sznaier, and C. Lagoa. Semi-blind robust identification model (in)validation with application to macroeconomic modeling. IFAC, 2005.

[42] D. Miskulin, D. Weiner, H. Tighiouart, V. Ladik, K. Servilla, P. Zager, A. Martin, H. Johnson, and K. Meyer. Computerized decision support for EPO dosing in hemodialysis patients. American Journal of Kidney Diseases, 54(6):1081–1088, 2009.

[43] A. E. Gaweda, B. H. Nathanson, A. A. Jabos, M. G. George Aronoff, and M. Brier. Determining Optimum Hemoglobin Sampling for Anemia Management from Every-Treatment Data. American Society of Nephrology, (5):1939–1945, 2010.