Estimation and Validation of Semiparametric Dynamic Nonlinear Models

Yves Rolain, *Fellow, IEEE*, Wendy Van Moer, *Senior Member, IEEE*, Johan Schoukens, *Fellow, IEEE*, and Tom Dhaene, *Member, IEEE*

Abstract—An approach for measurement-based modeling of nonlinear devices is proposed. The method that is commonly used for linear time-invariant systems, namely, parametric modeling and nonparametric verification, is hereby extended to a class of nonlinear systems. The applicability of the method is illustrated on the baseband modeling of a radio-frequency amplifier over a wide power and frequency range.

Index Terms—High frequency, measurement, modeling, nonlinear systems, semiparametric.

I. Introduction

DURING THE last years, there has been an increasing interest in modeling of nonlinear behavior of subsystems that are "close" to being linear. Mainly, these systems were designed to be linear. However, a description of the deviation from this ideal behavior is vital for the evaluation or simulation of the performance of the global system to which the considered system belongs. In telecommunication applications, for example, data error rate is linked to the in-band distortion of the power amplifier that is used in the transmitter. As a consequence, a nonlinear model for the operation of the device in baseband (the neighborhood of the fundamental frequency) is needed by RF system designers to analyze, optimize, and tune the system.

For such a system, there is a lot of prior model information available. The model class that is considered here is further restricted to the class of the systems that, when excited by a periodic excitation, produce a periodic output with the same period. This class of systems is called the periodic in, same period out (PISPO) system. It is clear that this definition excludes "nasty" behaviors such as chaos or bifurcation [1].

In applications such as modeling for telecommunication, the class of signals that is used during device operation is known in advance. The class of narrowband modulated signals (signals with a modulation bandwidth of a few percent proportional to the carrier frequency) covers most practical applications and will be used here.

To obtain a model that is useful in a design context, one needs more than "just a model." It should be validated in the frequency band and power range where the model will be used later on. Therefore, a robust visualization and validation tool for the comparison of the measured system response to the estimated model response is mandatory. In this paper, a measurement-based nonparametric validation tool that is based on the Volterra theory is proposed.

Parametric models for PISPO systems based on measured input/output characteristics were already obtained earlier, following two main tracks: the polynomial models (including Volterra models) and the behavioral models such as neural nets or wavelet models [2], [3]. The approach used here is a combination of both approaches: It extends Volterra-type models with neural netlike kernel functions to obtain a parsimonious model for devices operating under hard nonlinear conditions.

The proposed three-step approach consisting of experiment design, measurement visualization, and finally model extraction and validation is experimentally illustrated on the identification of a power amplifier in a frequency band from 500 to 2500 MHz and for an input power range from −15 to 5 dBm.

II. Measurement and Experiment Design

The signal has to persistently excite the device over the user-specified frequency and power ranges for it to become an eligible excitation signal for model identification. Both spectrally rich signals and sine waves can be used to meet this.

Spectrally rich excitations, such as random phase multisines [4] or noise signals, are extensively used for baseband devices up to a few megahertz. In RF measurements that rely on the harmonic sampling principle, the use of noise excitation is not an option, as a harmonic sampler requires a periodic excitation signal to properly downconvert the signal. Multisine signals enable a fast device characterization over the analysis frequency band in one single take. Inclusion of a power sweep then realizes the requested experiment. The back side is that the generation of these multisine signals requires highly specialized RF generators to cover the modulation bandwidth that is considered here. For the extraction of the nonlinear model that is used in this paper, the use of a spectrally rich signal significantly complicates the processing.

Ease of generation and model extraction are the driving forces in using sine waves as an excitation signal. Of course, both the frequency and the power have to be swept now. This results in a much higher number of experiments to be performed: A separate experiment is required for each (power, frequency) pair.

However, the information that can be obtained by both excitation signals is different, as the multisine allows measurement

Fig. 1.   LSNA-based measurement setup.

Fig. 2.   Magnitude of the measured $\text{ETF}_1(f_0, |U(f_0)|)$ as a function of the input power and frequency.
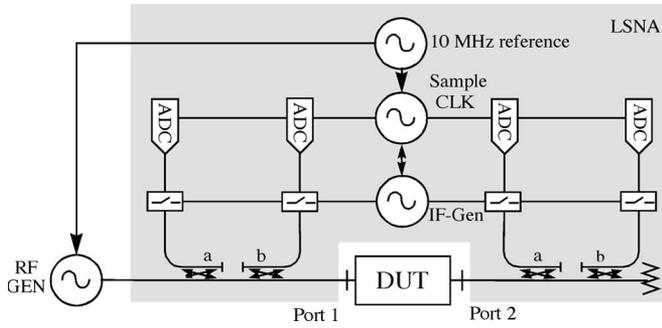
of intermodulation products that contribute at frequencies that are close to the carrier and therefore appear without attenuation at the output of the device due to the bandpass nature of the device under test (DUT).

A multicarrier signal is, therefore, a first choice whenever it is technologically realizable. However, in the RF and microwave example considered here, it is almost impossible to generate the arbitrary wideband signal that is needed to cover the whole frequency band. A first brand of arbitrary waveform generators covering the frequency band from dc to 2 GHz became available only recently and is not yet widely used. Sine wave generators, on the other hand, are the most common signal sources in the gigahertz frequency range and will therefore be used in this paper.

The measurement setup is build up around the large signal network analyzer (LSNA) that can be assimilated to a Fourier analyzer for microwave applications. The schematic diagram of the setup is shown in Fig. 1. The incident and reflected waves at both device ports are separated by couplers. The signals are then downconverted using a harmonic sampler and are digitized by VXI-based Agilent E1437 digitizers. These devices sample at 20-MHz sampling frequency, with a spurious-free dynamic range exceeding 90 dB. The data are then directly transferred to MATLAB for the processing. Note the presence of a synchronization between the RF generator and the LSNA. This phase reference is mandatory to keep the whole system phase-coherent and to remove leakage errors when the spectrum of the measured time waveforms is calculated.

III. MEASUREMENT VISUALIZATION

Once the measurements are taken at n_F frequencies and n_P power levels, one possesses $n_F n_P$ input and output spectra describing the DUT. The major question now is how to extract the information out of these data. To this end, a nonparametric approach that is comparable to the frequency response function (FRF) of the linear time-invariant (LTI) systems is used.

Assume that the Volterra theory gives a qualitatively valid description of the frequency mixing. This is a very weak hypothesis as the system is assumed to be a PISPO system.

A sine wave excitation with a complex amplitude equal to $U(f_0)$ will be used as an example to illustrate the proposed
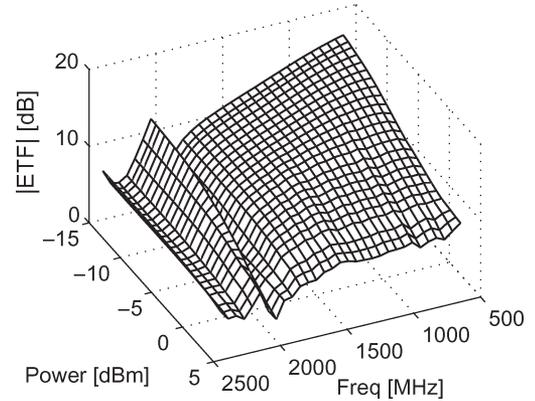
method. Consider the response at frequency f_0, as it is obtained using a Volterra representation [5]

$$Y(f_0) = \sum_{k=0}^{L} K_k(f_0) U^{k+1}(f_0) \overline{U}^k(f_0). \tag{1}$$

Note that only output contributions that fold back to the initial input frequency are considered here. All these terms have the same structure, namely, $U(f_0)|U(f_0)\overline{U}(f_0)|^2$. Therefore, the previous equation can be rewritten as

$$Y(f_0) = U(f_0) \left\{ \sum_{k=0}^{L} K_k(f_0) \left\| U(f_0) \right\|_2^{2k} \right\}. \tag{2}$$

The last factor on the right-hand side is a polynomial with complex coefficients that has the (real) power of the input as an independent variable. Dividing this equation by $U(f_0)$, one obtains

$$\text{ETF}_1\left(f_0, |U(f_0)|\right) = \frac{Y(f_0)}{U(f_0)} = \left\{ \sum_{k=0}^{L} K_k(f_0) |U(f_0)|^{2k} \right\}. \tag{3}$$

This quantity is defined here as an energy transfer function and would be equal to the FRF if the system is linear. Due to the nonlinearity of the system, $\text{ETF}_1(f, |U(f_0)|)$ is no longer a constant but rather a smooth complex-valued surface that is phase coherent with the input signal. The shape of this surface (both magnitude and phase) gives a lot of insight into the behavior of the device, as shown in Figs. 2 and 3, where the magnitude and the phase of this quantity are shown as a function of the input power and frequency.

This approach nicely extends to multicarrier excitation signals or contributions that shift the frequency of the output. As an example, consider that a second harmonic $(3f_0)$ is also present in the input signal. The Volterra equation for this two-tone input signal can be obtained in analogy to the single sinewave and becomes

$$Y(f_0) = U(f_0)\text{ETF}_1\left(f_0, |U(f_0)|\right) + U(3f_0)\overline{U}^2(f_0)$$
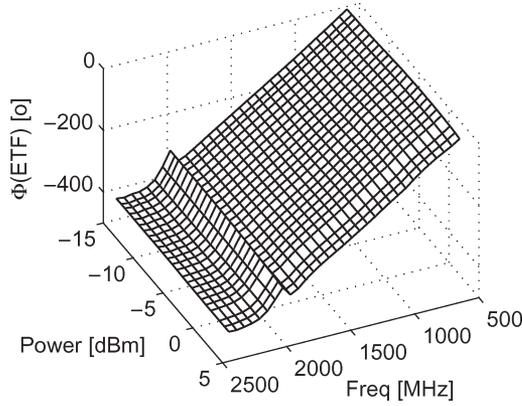$$\times \text{ETF}_{31}\left(f_0, |U(f_0)|, |U(3f_0)|, \left|U(3f_0)\overline{U}^3(f_0)\right|\right) \tag{4}$$

Fig. 3. Phase of the measured $\text{ETF}_1(f_0, |U(f_0)|)$ as a function of the input power and frequency.

with

$$
\text{ETF}_{31}\left(f_0, |U(f_0)|, |U(3f_0)|, \left|U(3f_0)\overline{U}^3(f_0)\right|\right)
$$
$$
= \sum_{k=0}^{L}\sum_{l=0}^{L-k}\sum_{p=0}^{L/2-k-l} \gamma_{pkl}(f_0)\,|U(3f_0)|^{2l}\,|U(f_0)|^{2k}
$$
$$
\times \Re\left(U(3f_0)\overline{U}^3(f_0)\right)^p. \tag{5}
$$

The function $\text{ETF}_{31}(f_0, U(f_0), U(3f_0))$ appears again to be a 3-D polynomial function of real independent variables. Even if this can be strange from an intuitive point of view, the reason for it is quite obvious. Remember that, for a real-valued time signal, $\overline{U}(f) = U(-f)$. This means that the factor $U(3f_0)\overline{U}^2(f_0)$ in (4) contributes to the spectrum at f_0. The spectral contributions that appear in ETF_{31} therefore cannot contribute to the output frequency at all. Hence, besides the power contributions $|U(3f_0)|^2$ and $|U(f_0)|^2$, the only possible contributions with zero frequency are $U(3f_0)\overline{U}^3(f_0)$, $\overline{U}(3f_0)U^3(f_0)$ or powers thereof. As the Volterra kernels can always be symmetrized, these contributions always appear in complex conjugate pairs, whose sum is again a real variable. A smooth surface that is phase coherent with the input spectrum results again. Evaluating $Y(f_0)/U(f_0)$ and $Y(f_0)/(U(3f_0)\overline{U}(f_0)\overline{U}(f_0))$ only yields a coherent surface in the range of (input power, frequency), where one energy transfer function (ETF_1 or ETF_{31}) dominates the other contributions. This yields a lot of insight in the operation of the device.

These ideas can be extended to a general response with more than two excited lines at the cost of a complex mathematical formalism that is out of the scope of this paper.

IV. IDENTIFICATION AND MODEL VALIDATION

In the context of this paper, an output error noise model is used. This ensures consistency of the estimates [6]. Using repeated experiments, the variance of the noise source can be used to improve the estimators' efficiency.

For the considered class of PISPO systems, a Volterra-like description is flexible enough to yield appropriate models (in mean-square sense) but fails to describe device saturation (also called compression). This is very similar to the observa-tion that a polynomial cannot give a high-quality description for a static system that is driven into deep saturation. Since the systems used in this context are driven up to deep saturation or high compression levels, the plain Volterra series is bound to the impossible.

To get around this problem, one can start from the equations derived in the previous paragraph. Reading in between the lines, a simple and accurate model should be obtained by replacing the polynomial function ETF by some other real function which adequately matches the effects of amplitude saturation. This new set of functions should behave like a polynomial for moderate power levels, but saturate outside the measured power range.

Here, the contribution of type $U(f)$ is replaced by $U(f)/(g + |U(f)|)$, with g as a real gain factor. For the first example used earlier, the model used in the identification of the semiparametric model at test frequency f_l becomes

$$
Y(f_l) = \sum_{k=0}^{L} K_{lk}\left\{\frac{U(f_l)}{g_l + |U(f_l)|}\right\}^{k+1}\left\{\frac{\overline{U}(f_l)}{g_l + |U(f_l)|}\right\}^{k}. \tag{6}
$$

To extend this model to the general case discussed earlier, each nonzero input spectral line is assigned one separate gain factor.

Three successive modeling steps are used, and each step delivers a model that can be used on its own. The number of parameters contained in the model decreases in each step.

1) In the first step, a separate parametric model is separately estimated for each of the n_F test frequencies using the model of (6). This model is, from now on, called non-parametric versus frequency.
2) To obtain a more parsimonious model in the second step, a parametric polynomial representation of the gain g versus frequency is introduced. In this hybrid parametric/nonparametric representation, the $(L+1)n_F$ values K_{lk} and the parametric representation of g are now estimated together using the data of the $n_F n_P$ experiments. Because of its hybrid structure, this model is called semi-parametric versus frequency.
3) To obtain a parsimonious model, a parametric LTI model is used to model each K_l-factor versus frequency. The parametric models for the K_l and g are simultaneously extracted, based on the $n_F n_P$ available measurements.

A. Nonparametric Model versus Frequency

The proposed model is parametric in the input power but non-parametric in the frequency variable f_l. One model is therefore evaluated at each measured frequency f_l. The estimation for each frequency reduces to the optimization of the following cost function:

$$
L_N\left(Y(f_l), U(f_l), \theta\right) = e_w^H e_w
$$
$$
\text{with} \quad e_{wk} = \frac{Y_k(f_l) - M\left(\theta, U_k(f_l)\right)}{W_k(\theta_p)} \tag{7}
$$

where N is the number of power levels measured at each frequency, $\theta = [K_{l1}, \ldots, K_{lL}, g_l]$ is the parameter vector, and

$M(\theta, U(f_l))$ is the model equation, as defined in (6). The weight $W_k(\theta)$ is set equal to the experimental standard deviation of the equation error $Y_k(f_l) - M(\theta_p, U_k(f_l))$ that is obtained from repeated experiments. Note that this model is linear in K_{lj}. g_l is the only parameter that enters the model in a nonlinear way. The model is, hence, a semilinear model, as considered in [4]. This significantly simplifies the optimization. A 1-D nonlinear optimization remains to be solved after the elimination of the linear parameters. The estimates for this first step, which is nonparametric over the frequency, are

$$\hat{\theta}_l = \underset{\theta}{\mathrm{argmin}} \left(L_N \left(Y(f_l), U(f_l), \theta \right) \right). \qquad (8)$$

B. Semiparametric Model versus Frequency

In this second step, the parameters g_l are replaced by a parametric polynomial model

$$g(f) = \sum_{m=0}^{G} \gamma_m f^m. \qquad (9)$$

Optimization is now performed together over the n_F frequencies and is still linear in the parameters K_{lk}. Using the elimination scheme of the aforementioned nonparametric case for the parameters that linearly appear in the model equation, only the parameters γ_m need to be optimized.

C. Parametric Model versus Frequency

Finally, the nonparametric kernel values K_{lk} are also replaced by a parametric LTI model $K_l(f)$. The parameters that linearly appear in the model are again eliminated as earlier, leaving a reasonable amount of parameters to be optimized. Of course, the resulting set of parameters involves the estimation of LTI systems. This problem is nonlinear in the parameters but is readily solved using the techniques in [4].

For the validation of the model, the same technique is used as for the visualization of the measurement, but this time, modeled and measured validation outputs are displayed on the same surface plots together with the magnitude of the complex error between the model and measurement.

V. SAMPLE SYSTEM

A GSM band power amplifier of type MAR6 (Mini-Circuits) is modeled. The supply voltage is set to 4 V, whereas the device output is terminated in a 50-Ω load. Absolutely calibrated[1] incident and reflected wave spectra at both ports of the DUT are measured by the LSNA [7].

The amplifier input is excited by a sine wave. The frequency is stepped from 600 to 2500 MHz in 20–MHz steps. The input power is stepped from -15 to 5 dBm in steps of 0.2 dB. At 5 dBm, the 2-dB compression point is reached in the passband of the device. Sample variances for the measured spectra are obtained by using only five repeated measurements. Even if this
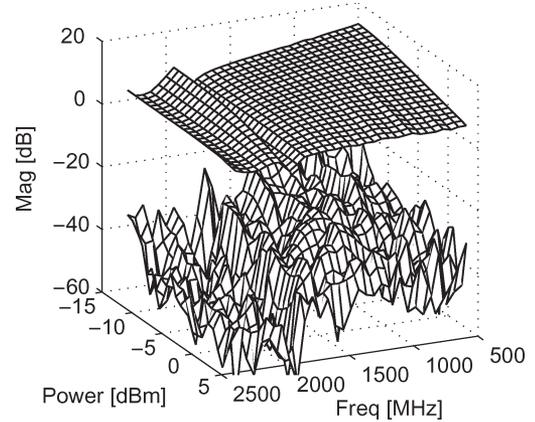

Fig. 4. Magnitude of (upper surface) the measured output and the nonparametric model output and (lower surface) the norm of the complex residual between them for the nonparametric model.

yields only a crude approximation of the variance, this can be proven to be sufficient in estimating an FRF with a minimal loss of statistical efficiency [4]. The relation

$$\mathrm{ETF}_1 \left(f_0, |U(f_0)| \right) = \frac{Y_m(f_0)}{U_m(f_0)} \qquad (10)$$

is analyzed. In this equation, the subscript m denotes a measured quantity. The full model equation is given in (6).

A total of 9696 spectral measurements are now condensed in one measured complex ETF_1 surface: The magnitude and phase of the surface are shown in Figs. 2 and 3. Note that the magnitude of the response has a large gain at both 900 and 1800 MHz, which is quite obvious for a dual-band GSM amplifier. The gain compression is much higher close to 900 MHz than to 1800 MHz. Phase dependence shows an important phase nonlinearity close to 1800 MHz.

A nonparametric model containing four terms is estimated for the 96 test frequencies

$$Y(f_l) = \sum_{k=0}^{5} K_{lk} \left\{ \frac{U(f_l)}{g_l + \|U(f_l)\|^2} \right\}^{k+1} \left\{ \frac{\overline{U}(f_l)}{g_l + \|U(f_l)\|^2} \right\}^{k}. \qquad (11)$$

To visualize the quality of the model, the measured output of the system $Y_m(f_l, |U(f_0)|)$ and the modeled output $Y(f_1, |U(f_0)|)$ that is obtained using the noisy measured signal as an input to the model are plotted on top of each other in Fig. 4. Visually, the match is so good that the curves fall on top of each other. To indicate the level of agreement, the magnitude of the complex error $\|Y_m(f_1, |U(f_0)|) - Y(f_1, |U(f_0)|)\|$ is plotted on the same plot (lower curve). Over the whole band, the relative complex modeling error is down by 35 dB. In the context of this paper, this model quality is very good for a model that is valid over a wide range of frequencies and input power levels. As shown in the figure, the remaining differences are almost at the noise level.

The first four corresponding normalized K_{lk} and the g_l parameters that were estimated for each excited frequency are shown in Fig. 5. Even if the model response is very close to the measured response, the behavior of the gain factor g_l and
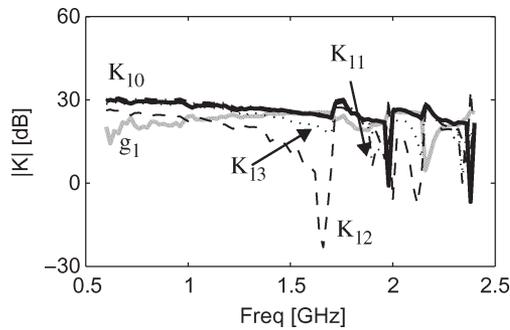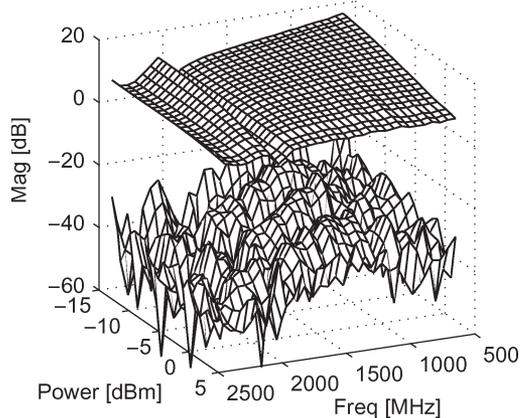
---

[1]For vectorial network analyzers that measure S-parameters, the calibration of the ratio of waves is sufficient. This is no longer the case for a nonlinear DUT, where calibrated waves are needed. This wave calibration is called "absolute."

Fig. 5. Normalized estimated K_{lk} and g_l for the nonparametric model.

Fig. 6. Magnitude of (upper surface) the measured output and the nonparametric model output and (lower surface) the norm of the complex residual between them for the semiparametric model.

the linear kernel K_{l0} contain several jumps and can hardly be interpreted in a physical context. This potentially indicates that gain and linear kernel are not totally independent.

In the second step, a semiparametric model with a parametric gain $g = \gamma_0$ is extracted. Again, the quality of the resulting model is shown using an overlay of the measured and modeled system responses (upper surface in Fig. 6) and the magnitude of their complex difference (lower surface in Fig. 6). When the residual in Fig. 6 is compared to the residual in Fig. 4, a slight increase can be noted, especially for frequency and parameter values that are close to the boundaries of the measured ranges. In fact, the difference is quite marginal when one considers that 96 parameters were removed from the model. This points in the direction of a very high correlation between the parameters g and K_0. The parametrization of g therefore acts as a regularization in the identification step. The first four normalized K_{lk} parameters that are separately estimated for all frequencies are shown in Fig. 7. In this plot, the kernels K now all have a smooth behavior as a function of frequency. The linear kernel K_0 gets close to the linear behavior of the DUT that has been measured using a vectorial network analyzer.

The smooth behavior of the functions now allows us to separately extract an LTI model in the Laplace plane for the five K functions. To this end, the frequency domain identification toolbox for MATLAB has been used. The estimated model frequency responses are shown together with the nonparametric "measurements" of the K-values in Figs. 8 and 9. Note that, at this time, the values are not normalized. This explains the
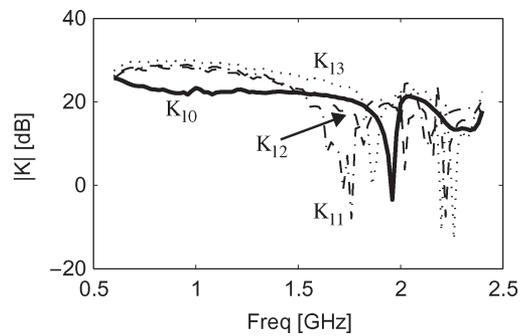
Fig. 7. Normalized estimated K_{lk} for the semiparametric model.
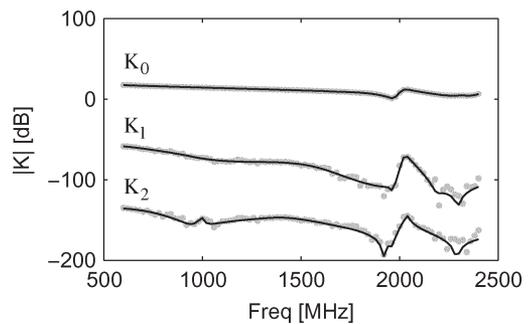
Fig. 8. Estimated (lines) and measured (gray dots) K_0, K_1, and K_2 for the parametric LTI model.
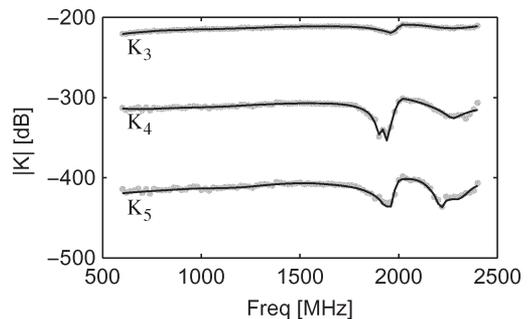
Fig. 9. Estimated K_3, K_4, and K_5 for the parametric LTI model.

very high dynamic range in between the different functions. The maximal order of the rational forms that are used is 7/7, and the model order is automatically estimated by the toolbox. All the proposed models pass the whiteness-of-residuals test provided. Based on this parametric model, the output spectrum of the model is again evaluated as earlier, which is based on the measured input signal. The result is shown in Fig. 10. Note that the mean residual error did increase by more or less 6 dB, whereas the number of parameters decreased by a factor of seven when compared to the semiparametric case.

To further validate the results, the model is used afterward on a validation data set with results that are similar to what has been shown earlier.

VI. Conclusion

A three-step approach is proposed to model nonlinear systems in the frequency domain. In the first step, a set of experiments is designed, and a large amount of data is acquired.
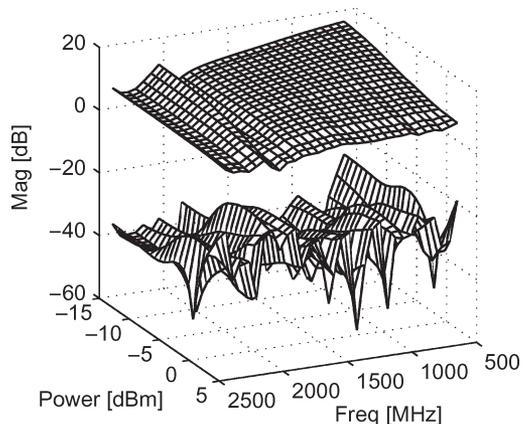
Fig. 10. Magnitude of (upper surface) the measured output and the nonparametric model output, and (lower surface) the norm of the complex residual between them for the parametric model.

In the second step, energy transfer functions are introduced to condense the data in a few characteristics that allow us to gain insight into the device operation using only the measured data. Building up this measurement-based knowledge is a vital and often overlooked part of nonlinear modeling. Finally, a modified Volterra-type model is proposed, which allows us to appropriately model systems in deep compression. Validation of the proposed models is performed using the visual tools. The modeling approach is illustrated on the identification of an RF power amplifier operating between 600 and 2500 MHz and between −15 and 5 dBm input power.

ACKNOWLEDGMENT

REFERENCES

[1] T. Poston and I. N. Stewart, *Catastrophe Theory and Its Applications*. London, U.K.: Pitman, 1978.

[2] Y. Rolain, W. Van Moer, and P. Vael, "Measuring the characteristics of modulated non-linear devices," in *Proc. 53rd ARFTG Conf.—Nonlinearity Characterization*, Anaheim, CA, Jun. 17–18, 1999, pp. 1–9.

[3] W. Van Moer, Y. Rolain, and A. Geens, "Measurement based nonlinear modeling of spectral regrowth," *IEEE Trans. Instrum. Meas.*, vol. 50, no. 6, pp. 1711–1716, Dec. 2001.

[4] R. Pintelon and J. Schoukens, *System Identification. A Frequency Domain Approach*. New York: IEEE Press, 2001.

[5] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York: Wiley, 1980.

[6] L. Ljung, *System Identification: Theory for the User*. New York: Prentice-Hall, 1999.

[7] W. Van Moer and Y. Rolain, "The large signal network analyzer: Why is it needed?" *IEEE Microw. Mag.*, vol. 7, no. 6, pp. 46–62, Dec. 2006.

Yves Rolain (M'90–SM'96–F'06) was born in Belgium in 1960. He received the degree in electrical engineering, the degree in computer sciences, and the degree in applied sciences from Vrije Universiteit Brussel (VUB), Brussels, Belgium, in 1984, 1986, and 1993, respectively.

He is currently with the Electrical Measurement Department, VUB. His main research interests include nonlinear microwave measurement techniques, applied digital-signal processing, parameter estimation/system identification, and biological agriculture.

Dr. Rolain is the recipient of the 2004 IEEE Instrumentation and Measurement Society Award.

Wendy Van Moer (M'97–SM'07) received the Engineer degree in telecommunication and the Doctor degree in applied sciences from the Vrije Universiteit Brussel (VUB), Brussels, Belgium, in 1997 and 2001, respectively.

She is currently a Postdoctoral Researcher with the Electrical Measurement Department, VUB. Her main research interests include nonlinear microwave measurements and modeling techniques.

Dr. Van Moer was the recipient of the 2007 IEEE Instrumentation and Measurement Outstanding Young Engineer Award.

Johan Schoukens (M'90–SM'92–F'97) was born in Belgium in 1957. He received the Engineer and Doctor degrees in applied sciences from the Vrije Universiteit Brussel (VUB), Brussels, Belgium, in 1980 and 1985, respectively.

He is currently a Professor with the VUB. His main research interest includes system identification for linear and nonlinear systems.

Tom Dhaene (M'02) was born in Deinze, Belgium, on June 25, 1966. He received the Ph.D. degree in electrotechnical engineering from Ghent University, Ghent, Belgium, in 1993.

From 1989 to 1993, he was a Research Assistant with the Department of Information Technology, Ghent University, where his research focused on different aspects of full-wave electromagnetic circuit modeling, transient simulation, and time-domain characterization of high-frequency and high-speed interconnections. In 1993, he joined the EDA company Alphabit (now part of Agilent). He was one of the key developers of the planar EM simulator ADS Momentum, ADS Model Composer, and ADS Broadband SPICE. Since 2007, he has been a Full Professor with the Department of Information Technology, Ghent University. He has authored or coauthored more than 120 peer-reviewed papers and abstracts in international conference proceedings, journals, and books. He is the holder of two U.S. patents.