New criteria for assessing fit quality in dynamic contrast-enhancedT1-weighted MRI for perfusion and permeability imaging

Daniel Balvay, Frédérique Frouin, Guillaume Calmon, Bertrand Bessoud, Edmond Kahn, Nathalie Siauve, Olivier Clément, Charles A Cuenod

Magnetic Resonance in Medicine 54:868–877 (2005)

New criteria for assessing fit quality in dynamic contrast-enhanced T_1-weighted MRI for perfusion and permeability imaging

Daniel Balvay[1-3], Frédérique Frouin[1], Guillaume Calmon[2], Bertrand Bessoud[3], Edmond Kahn[1], Nathalie Siauve[3,4], Olivier Clément[3,4], Charles A. Cuenod[3,4]

1U678 INSERM/UPMC, APHP, CHU Pitié Salpêtrière, Paris, France.
2General Electric Healthcare, Vélizy, France.
3Universite´ Paris Descartes/LRI, APHP, CHU Necker, Paris, France.
4APHP, CHU HEGP, Service de Radiologie, Paris, France.
*Correspondence to: Daniel Balvay, U678 INSERM, 91 Bd de l'Hopital, 75634 Paris Cedex 13, France. E-mail: ~~balvay@imed.jussieu.fr~~ <new> daniel.balvay@inserm.fr
Received 31 September 2004; revised 31 May 2005; accepted 2 June 2005.

First published: 09 September 2005
DOI 10.1002/mrm.20650

Abstract

Contrast-enhanced (CE) MRI provides in vivo physiological information that cannot be obtained by conventional imaging methods. This information is generally extracted by using models to represent the circulation of contrast agent in the body. However, the results depend on the quality of the fit obtained with the chosen model. Therefore, one must check the fit quality to avoid working on physiologically irrelevant parameters. In this study two dimensionless criteria - the fraction of modeling information (FMI) and the fraction of residual information (FRI) - are proposed to identify errors caused by poor fit. These are compared with more conventional criteria, namely the quadratic error and the correlation coefficient, both theoretically and with the use of simulated and real CE-MRI data. The results indicate the superiority of the new criteria. It is also shown that these new criteria can be used to detect oversimplified models.

Physiological parameters of tissue perfusion and capillary permeability can be obtained from contrast-enhanced (CE) MRI (1-4). These parameters can be estimated by deconvolution methods (5, 6) or compartmental analysis (7, 8). The identification of such parameters has become an important aim in imaging research, in the fields of oncology (9, 10), neurology (1, 11, 12) and cardiology (13-15), to facilitate both the detection (16, 17) and characterization of abnormalities (2, 18, 19) and the evaluation of treatment efficacy (20, 21).

Compartmental analysis estimates physiological parameters by regression (22). The parameters of the chosen compartmental model are adjusted to obtain a modeled response that is as faithful as possible to the observed tissue dynamics, generally based on least-squares regression. When the analysis is applied to each pixel, it generates maps of estimated physiological parameters.

However, when regression is used, nothing guarantees that the modeled response will effectively match the observed tissue dynamics. The quality of fit must therefore be determined in order to avoid erroneous interpretation of physiologic parameters.

One strategy that is commonly used to verify the quality of fit is to visually compare modeled responses and measured temporal data. We will refer to this method as "visual analysis." This approach is reliable, but it is also subjective and cannot be applied to the multitude of fits obtained in parametric imaging. Another strategy is to calculate the quadratic error (Q) or the correlation coefficient (R^2) (23-25); however, these criteria, which are based on measuring the gap between modeled and measured data, depend not only on the quality of fit but also on the random noise contained in the data set. If the quadratic error (or the correlation coefficient) is to be a reliable confidence indicator, the noise should be negligible or at least of constant amplitude. Yet both the amplitude of noise and the signal-to-noise ratio (SNR) depend on the experimental conditions and can vary from one pixel to another, depending on the concentration of contrast agent used. Thus, conventional confidence criteria are, in principle, not very well suited to studying the quality of fit in CE-MRI.

The aim of this work was to develop numerical criteria to assess the fit quality of a given model using the properties of the autocorrelation function to overcome the random noise issue.

The theoretical view of this problem is first approached by introducing the notion of quadratic modeling error (Q'), which corresponds to the conventional quadratic error (Q) that one would obtain for noiseless data. Q' then yields two normalized dimensionless criteria: the fraction of modeled information (FMI) and the fraction of residual information (FRI). The quadratic modeling error Q' is estimated from the fit residual on the basis of the following postulates: the modeling error is strongly correlated in time because it translates a temporal law of change; in contrast, by nature, random noise is poorly time-correlated. The autocorrelation function is thus used to distinguish the respective contributions of the modeling error and noise within the residual. Finally, the theoretical criteria Q', FMI, and FRI are estimated as Q'*, FMI*, and FRI*. This estimation method is then tested on simulated data containing various amplitudes of random noise. The stability of the estimated criteria relative to Gaussian white noise is also compared with the stability of the conventional criteria. The estimated criteria (FMI* and FRI*) and conventional criteria (Q, R^2) are then applied to observed CE-MRI data, on the basis of the fit quality, which is clearly confirmed by the visual reference method. In this study we compare the abilities of the new and conventional criteria to reproduce the visual classification ("good" and "poor" fits) with both criteria maps and the mean values of the criteria in regions corresponding to tumors. The impact of additional non-random noise on these new criteria is also discussed.

THEORY

Limits of the Quadratic Error

It can be assumed that the data $d(t)$ derived from a series of CE-MRI acquisitions are the sum of a deterministic phenomenon of interest $d_o(t)$ and random noise $b(t)$. At the end of this section, a generalization for additional temporally correlated noise is introduced. For each pixel, with i and j being indices of the position within the image, it is assumed that:

$$d_{ij}(t) = d_{oij(t)} + b_{ij}(t) \quad (1)$$

In practice, model-based data fitting decomposes some of these data into the sum of a modeled signal (m) and a residual (r):

$$d_{ij}(t) = m_{ij}(t, p^*) + r_{ij}(t) \quad (2)$$

where $p^* = \{p_1^*, \ldots, p_n^*\}$ represents the n parameters of the model that minimize the quadratic error Q. By definition, $Q = \|d - m\|^2 = \|r\|^2$. The aim of regression techniques is to model d_o as well as possible with m, by rejecting the noise b in the residual r. To determine how well this objective is reached, the modeling

error (e_{tot}) is defined as the difference between d_o and m: $e_{tot}(t)=d_o(t)-m(t,p^*)=r(t)-b(t)$ and the quadratic modeling error (Q') by its amplitude:

$$Q' = \|e_{tot}\|^2 = \|r - b\|^2 \quad (3)$$

The quality of fit improves as Q' decreases, and is perfect when $Q' = 0$.

The conventional quadratic error (Q) can then be expressed as:

$$Q = \|(d_o + b) - m\|^2 = \|e_{tot} + b\|^2 \quad (4)$$

The quadratic error therefore depends on both the modeling error and the noise amplitude. If the model is assumed to be robust (i.e., relatively noise-independent), then the modeling error is also relatively independent of noise, and Q can be approximated as ($\|e_{tot}\|^2 + \|b\|^2$). This approximation can be used to deduce the following three properties:

- 1) If the noise amplitude is constant, the fit improves as Q decreases. This property is used during regression.
- 2) If the noise amplitude is not constant, the quadratic error varies and therefore can no longer be used to compare different fits.
- 3) The value of Q does not perfectly reflect the quality of fit, because of its dependence on noise amplitude, which is not known beforehand.

Properties 2 and 3 hold true not only for Q but also for functions of Q, such as the correlation coefficient R^2 defined as:

$$R^2 = 1 - \|r\|^2/\|d\|^2 \quad (5)$$

Figure 1 illustrates the limitations of Q and R^2 in their capacity to compare qualities of fit. On two pixel enhancements it shows that lower values of Q and higher values of R^2 can, paradoxically, correspond to poorer fit.

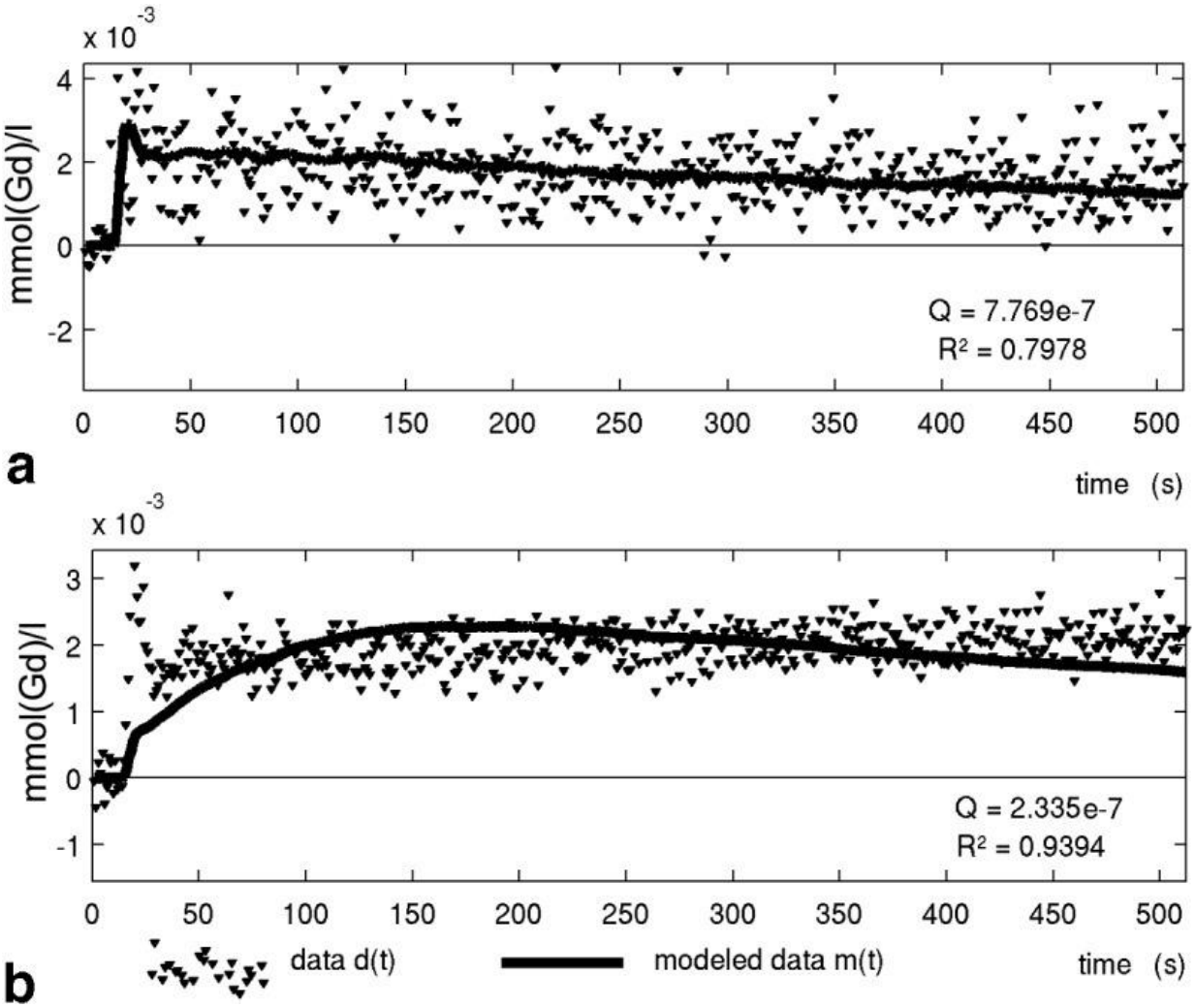

Figure 1. *Limitations of the quadratic error (Q) and correlation coefficient (R^2) for comparing two fits of pixel enhancements. The better fit* ***(a)*** *is not necessarily characterized by a lower Q and a higher R^2* ***(b)****.*

Deterministic Modeling Error

The modeling error e_{tot} can arise from three sources:

- 1) Accidental modeling of some of the random noise creates a statistical modeling error (e'). This "excess of modeling" is reflected by a statistical error in the parameters obtained by regression (dp'). The statistical error dp' of the parameters can be studied with existing methods (26-28).
- 2) Failure of the regression method, which leads to a local minimum quadratic error. This creates a regression error (e_r) that corresponds to the interval between the local minimum and the absolute minimum. Error e_r is deterministic, and is expressed as a deterministic parameter error (dp_r).
- 3) A model that is poorly adapted to the observed phenomenon yields a poor fit regardless of the parameters used. When the regression reaches the absolute minimum, a deterministic modeling error (e_m) arises. This modeling error corresponds to a deterministic parameter error (dp_m) if e_m disappears in a modified model. Otherwise, the modeling error indicates that the model is qualitatively inappropriate. Thus

$$e_{tot} = e' + e_r + e_m \qquad (6)$$

This article is devoted to the last two items, which comprise the deterministic modeling error ($e = e_r + e_m$). This error is an index of reliability, in that it is associated with a deterministic parameter error ($dp = dp_r + dp_m$). It indicates the extent to which the modeled data m, and the corresponding parameters then represent the observed phenomenon d_o. It is assumed below that the model is robust (i.e., that e_{tot} is noise-independent), in which case $e' \ll e$, and the modeling error is equivalent to the deterministic modeling error ($e_{tot} = e + e' \approx e$).

If a fit is to be appropriate, the deterministic modeling error e must be small relative to the modeled phenomenon d_o. The first dimensionless criterion—the Fraction of Modeled Information (FMI)—is defined as follows:

$$FMI = 1 - \|e\|^2/\|d_o\|^2 \tag{7}$$

This criterion is constructed by analogy with R^2. It represents the non-modeled information (e) relative to d_o. For a robust regression, the modeled data (m) and the modeling error (e) can be assumed to be independent. Then $\|d_o\|^2$ can be approximated by $\|m\|^2 + \|e\|^2$, and FMI by $\|m\|^2/\|d_o\|^2$. This explains the use of the term "FMI."

A second dimensionless criterion - the Fraction of Residual Information (FRI) - represents the non-modeled information (e) relative to the total residual, and is defined as follows:

$$FRI = \|e\|^2/\|r\|^2 = Q'/Q \tag{8}$$

This criterion is associated with the ability to extract a modeling error from the residual. It identifies a modeling error e, even when this error is small relative to the observed phenomenon d_o.

In an ideal modeling process (i.e., one in which $m = d_o$), conventional criteria are noise-dependent,

$$R^2 = 1 - \|b\|^2/\|d\|^2$$

$$Q = \|b\|^2$$

whereas the new criteria, which are not random-noise-dependent, reach known maxima and minima:

$$FMI = 1\,; FRI = 0$$

Estimation of the Deterministic Modeling Error

The phenomenon d_o is unknown when we deal with observed data (d). The criteria Q', FRI, and FMI must therefore be estimated in the presence of noise. We propose an estimation based on autocorrelation. Temporal autocorrelation (29) of a signal x of M samples, for a time-lag $k \cdot T$ (where T is the sampling interval) is defined as

$$R_{xx}(k) = \frac{1}{M-|k|}\sum_{i=1}^{M-k} X(t_i) \cdot X(t_{i+k}) \tag{9}$$

This function can be used for $M >> 1$. R_{xx} is a symmetrical function (i.e., $R_{xx}(-k) = R_{xx}(k)$) and, when zero:

$$R_{xx}(0) = \|x\|^2/M \tag{10}$$

Assuming that $r = e + b$, for each time lag k (Fig. 2a):

$$R_{rr}(k) = R_{ee}(k) + R_{eb}(k) + R_{be}(k) + R_{bb}(k) \tag{11}$$

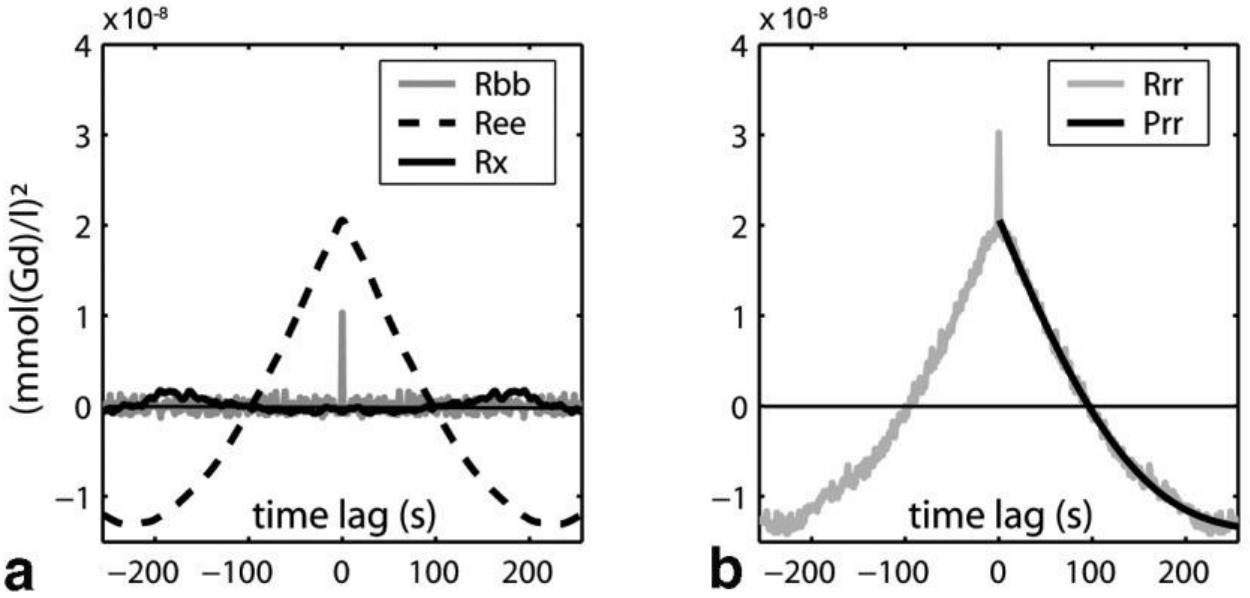

Figure 2. *Estimation of the quadratic modeling error Q′ by autocorrelation of the residual (r).* ***a:*** *The autocorrelation of residual R_{rr} is the sum of autocorrelations of the modeling error R_{ee} and noise R_{bb}, and of an intercorrelation term R_x.* ***b:*** *Estimation of $Q' = \|e\|^2$ by fitting of R_{rr} with polynomial P_{rr}. $R_{ee}(k > 0)$ is estimated by the polynomial P_{rr}, which is extrapolated to zero.*

Assuming that e is deterministic and that b is white noise, then, when $k > 0$, $E\{R_{bb}(k)\} = E\{R_{be}(k)\} = E\{R_{eb}(k)\} = 0$, where $E\{.\}$ is the statistical expectation. Then, $E\{R_{rr}(k)\} = E\{R_{ee}(k)\} = R_{ee}(k)$ when $k > 0$. R_{rr} is therefore an unbiased estimator of R_{ee} when $k > 0$. This is no longer true when $k = 0$, as R_{bb} (0), which corresponds to the mean noise amplitude, is no longer negligible. However, if e is deterministic and the sampling frequency is adequate, d_o and m, and therefore e, vary only moderately with time. Thus, $e(t_{i+k})$ approximates $e(t_i)$ for low values of k, and, as a result, $R_{ee}(0) = \|e\|^2/M$ can be approximated by using $R_{ee}(k)$ for low values of k.

The term $R_{ee}(k > 0)$ is fitted by a polynomial function $P_{ee}(k > 0)$, and then P_{ee} is extrapolated by continuity to zero in order to estimate Q' (Fig. 2b):

$$\|e\|^{2*} = M \cdot P_{rr}(0) \qquad (12)$$

By applying this calculation to the data set, $\|d_o\|^2$ is estimated by $\|d_o\|^{2*} = M \cdot P_{dd}(0)$. Thus, FMI and FRI can be estimated as follows:

$$FMI^* = 1 - \frac{\|e\|^{2*}}{\|d_o\|^{2*}}, FRI^* = \frac{\|e\|^{2*}}{\|r\|^{2*}} \qquad (13)$$

Extension to Non-White Noise

The above theory and estimation technique can be extended to a larger family of random noise, namely decorrelated noise for time-lags larger than $k_o << M$. In such conditions the autocorrelation can be fitted with P_{ee} *and* P_{rr} only for $k > k_o$ and then extrapolated to zero.

If there is additional temporally correlated noise $bc(t)$ in the data, our new criteria (FMI and FRI) exclude the contribution of random noise, contrary to conventional criteria. However, in this case, a deterministic modeling error cannot be distinguished from correlated noise. The modeling error then has to be replaced by the more general "modeling mismatch" (e_{mm}), which is the sum of the modeling error and temporally correlated noise.

$$e_{mm} = r - b = e_{tot} + bc = e' + e + bc \quad (14)$$

For robust modeling, $Q' = \|r - b\|^2$ corresponds to deterministic phenomena ($e + bc$) that interfere with the temporal information corresponding to the model parameters. This deterministic phenomenon, which may (e) or may not (bc) be modeled, corresponds to a deterministic error of physiological parameters. FMI and FRI can still be used as reliability criteria. In addition, the principle of extraction of random noise b from the residual with autocorrelation is unchanged. However, the question of temporally correlated noise detection is beyond the scope of this paper. Thus, for a reliable interpretation of fit quality attributable only to the modeling error, the null hypothesis bc has to be assumed.

MATERIALS AND METHODS

Clearly, references are required to test the new fitting criteria (Q'*, FMI*, FRI*). A first study was performed by using simulated data, in which d_o was known, and by calculating the theoretical criteria Q', FMI, and FRI as references. A second study was performed on real data. Since we had no objective references, the results had to be confirmed by visual analysis of temporal signals. Although it is qualitative, visual analysis usually allows fits to be classified as obviously "good" or "poor." Such fits were obtained on observed data by using two distinct models: a first model that was not adjusted to the data (M1), and a second, adjusted model (M2). In both studies the results obtained with the new criteria (Q'*, FMI*, FRI*) were compared with those obtained with the conventional criteria (Q, R^2).

In these two studies, with simulated and observed data, the following two compartmental models were used:

- 1. A one-compartmental two-parameter model (M1) that solely described perfusion exchanges within the tissue (29). This model assumes that leakage of contrast agent through the capillary walls is negligible during the measurements. It yields a perfusion parameter F (ml/100g/min) and a blood volume fraction parameter v_b:

  $$\frac{dq_b}{dt} = F \cdot \left(q_{AIF} - \frac{q_b}{V_b}\right), q_t = q_b \quad (15)$$

  where q_b and q_{AIF} are the quantities of contrast agent in the capillary and arterial compartments, respectively, selected in an equal volume of observation (voxel volumes), and q_t is the total quantity of contrast agent in the tissue volume.

- 2. A two-compartmental model (M2) with four parameters (30). This model describes perfusion within the tissue, but also takes into account transcapillary distribution between capillary plasma and the extracapillary extracellular space. This model is used in the case of contrast agents that cross capillary walls during measurement. In addition to perfusion F (ml/100g/min) and blood volume fraction v_b, it provides the permeability-surface area product PS (ml/100g/min) and the extracapillary and extracellular (interstitial) volume fraction v_e:

$$\frac{dq_b}{dt} = F \cdot \left(q_{AIF} - \frac{q_b}{V_b}\right) - PS \cdot \left(\frac{q_b}{(1 - Hct) \cdot V_b} - \frac{q_e}{V_e}\right)$$

$$\frac{dq_e}{dt} = PS \cdot \left(\frac{q_b}{(1 - Hct) \cdot V_b} - \frac{q_e}{V_e}\right)$$

$$q_t = q_b + q_e \quad (16)$$

where q_b, q_e, and q_{AIF} are the quantities of contrast agent in the capillary, interstitial, and arterial compartments, respectively, which were selected in an equal volume of observation (voxel volumes). q_t is the total quantity of contrast agent in the tissue volume, and *Hct* is the hematocrit (fixed at 0.45). M1 and M2 are also called "single flow-limited" and "membrane-limited" compartmental models, respectively (31). The fittings were based on the Levenberg-Marquard regression method, which is available in the Matlab® software.

Study of Simulated Data

This study was undertaken to verify the reliability of the estimates (Q'*, FRI*, FMI*) of theoretical criteria (Q', FRI, FMI) for the different SNRs. This also allowed us to compare the estimated criteria (Q*, FMI*) with the corresponding conventional criteria (Q, R^2).

Noiseless tissue-enhancement data (d_o) were constructed with model M2, using observed arterial entry data obtained by CE-MRI and the same 512 time samples (see the Study of Real MRI Data section). The physiological parameters were chosen so as to generate tissue responses of the same order as those obtained by the fitting of real series of CE-MRI acquisitions. The reference parameters were chosen as follows: F_o = 29.5 ml/100 g/min; v_{bo} = 1.8%; PS_o = 0.2 ml/100 g/min; and v_{eo} = 35%. Initial tissue enhancement (d_o) was then modified by adding Gaussian white noise (b) of variable amplitudes. The range of noise amplitude was chosen so that the SNR in the data set (SNR = $\|d_o\|/\|b\|$) would be between 0.3 and 5. For the different noise amplitudes, the full simulated data set ($d = d_o + b$) was fitted using the simplified model M1, to obtain a constant nonzero deterministic modeling error corresponding to the model error ($e = e_m$). The estimated criteria (Q'*, FRI*, FMI*) and the conventional criteria (Q, R^2) were then calculated for each noise level and compared with the theoretical reference criteria (Q', FRI, FMI). The simulations were first performed qualitatively for the reference set of physiological parameters, with SNR increments of 0.1. Then for each SNR, 1000 sets of physiological parameters were selected using a uniform randomized generator in the range of 0.5 P_o to 1.5 P_o, where $P_o = \{F_o, v_{bo}, PS_o, v_{eo}\}$ with an SNR increment of 0.5.

Study of Real MRI Data

We verified that fits of different visual qualities were correctly classified by quantitatively distinct criteria.

MRI Acquisition

Experimental CE-MRI studies were undertaken on human prostatic adenocarcinomas (PC3) implanted into male nude mice (32). Images were acquired in 23 mice harboring tumors 8 mm in diameter. A rapid-clearance blood pool macromolecular contrast agent (Vistarem®, Guerbet, France) was used (33). CE-MRI was performed after initial scanning studies, with a dose of 0.045 mMol(Gd)/kg of contrast agent infused via a caudal venous catheter. Series of 512 2D fast spoiled gradient-recalled (FSPGR) images (TR/TE = 15

ms/2.2 ms, FA = 60°, thickness = 5 mm, matrix size = 256 × 128, FOV = 70 × 30 mm, half *k*-space, temporal resolution = 1.1s) were acquired on a 1.5-T clinical device (Signa®, General Electric, Velizy, France). After each acquisition the dynamic sequence was applied to a standard range of six tubes of Vistarem® (P792) diluted in water to concentrations of 0.112–2.24 mmol(Gd)/l (34).

Parametric Imaging

Arterial (left ventricle) and tissue regions (tumor) were selected manually and converted into concentrations of contrast agent by using the standard range (34). The tissue region was masked to eliminate pixels that did not behave as tissues irrigated by the left ventricle (elimination of pixels with excessively early enhancement and those with inadequate intensity or enhancement). The physiological parameters, using M1 and M2, and criteria (Q, R^2, FMI*, FRI*) were then calculated from all unmasked pixels. To determine the detection limit of deterministic modeling errors (e) based on the criteria, it was first necessary to know the values adopted by these criteria at zero modeling errors, i.e., when the residuals are reduced to the level of experimental noise. The temporal signals of pixels located outside the tissue are examples of the experimental noise used as reference. The residual criteria Q and FRI* calculated from these pixels yielded the minimal values of Q and FRI* that were capable of revealing a deterministic modeling error.

Designing Fits of Different Quality

To generate two sets of fitted data of different quality (based on the reference visual method), the data were fitted with models M1 and M2. It was then verified that there was a visually evident difference in fit quality between the two models.

Criteria Analysis

Each model was used to generate images of the physiological parameters and images of the associated criteria Q, R^2, FMI*, and FRI*. The different maps were examined to verify that the agreement between the classification of fits by maps of the new criteria (FMI* and FRI*) and by visual analysis was better than the classification between the maps of conventional criteria (Q, R^2) and visual analysis. The tumor regions were then manually delimited to include the whole tumors. The mean criteria and parameters were calculated for each tumor. Pairs of conventional criteria (Q, R^2) and estimated criteria (FMI* and FRI*) were then plotted to examine how well the different criteria were able to distinguish between fits of different visual quality.

The relevance of the M1 fits was estimated with the new criteria in "short" studies, corresponding to the first 10% of samples of the entire series. The coherence of the criteria information was analyzed by comparing physiologic maps generated for short series using M1, and maps generated from the whole series using M2. To be able to compare series with different number of samples M, the residual criteria Q, Q' and Q'* were scaled by M. For reasons of homogeneity, Q, Q', and Q'* are thus scaled by M throughout the paper.

RESULTS

Results for Simulated Data

As shown in Fig. 3a, for a given set of physiological parameters P_o, the quadratic modeling error Q' remained nearly constant (Q'_o = 3.17e-8) when the SNR fell, except for very low SNR values (<0.5), for which Q' varied randomly. This indicates that the hypothesis $e' << e$ was relevant over a large range of SNR values (0.5–5.0). The stability of Q' allowed us to study changes in Q and Q'* relative to random noise

b alone. For a globally constant modeling error, the quadratic error *Q* increased rapidly as SNR fell. The estimated quadratic modeling error (*Q'**) remained stable when SNR varied. It was centered on the theoretical quadratic modeling error (*Q'*). Variations of *Q'** were always smaller than the drift of *Q*. Similar trends were obtained for the correlation criteria R^2, FMI, and FMI* (Fig. 3b) with a theoretical deterministic FMI* of 0.98. The estimated fraction of residual information (FRI*) was close to the theoretical residual fraction (FRI) regardless of the SNR (Fig. 3c).

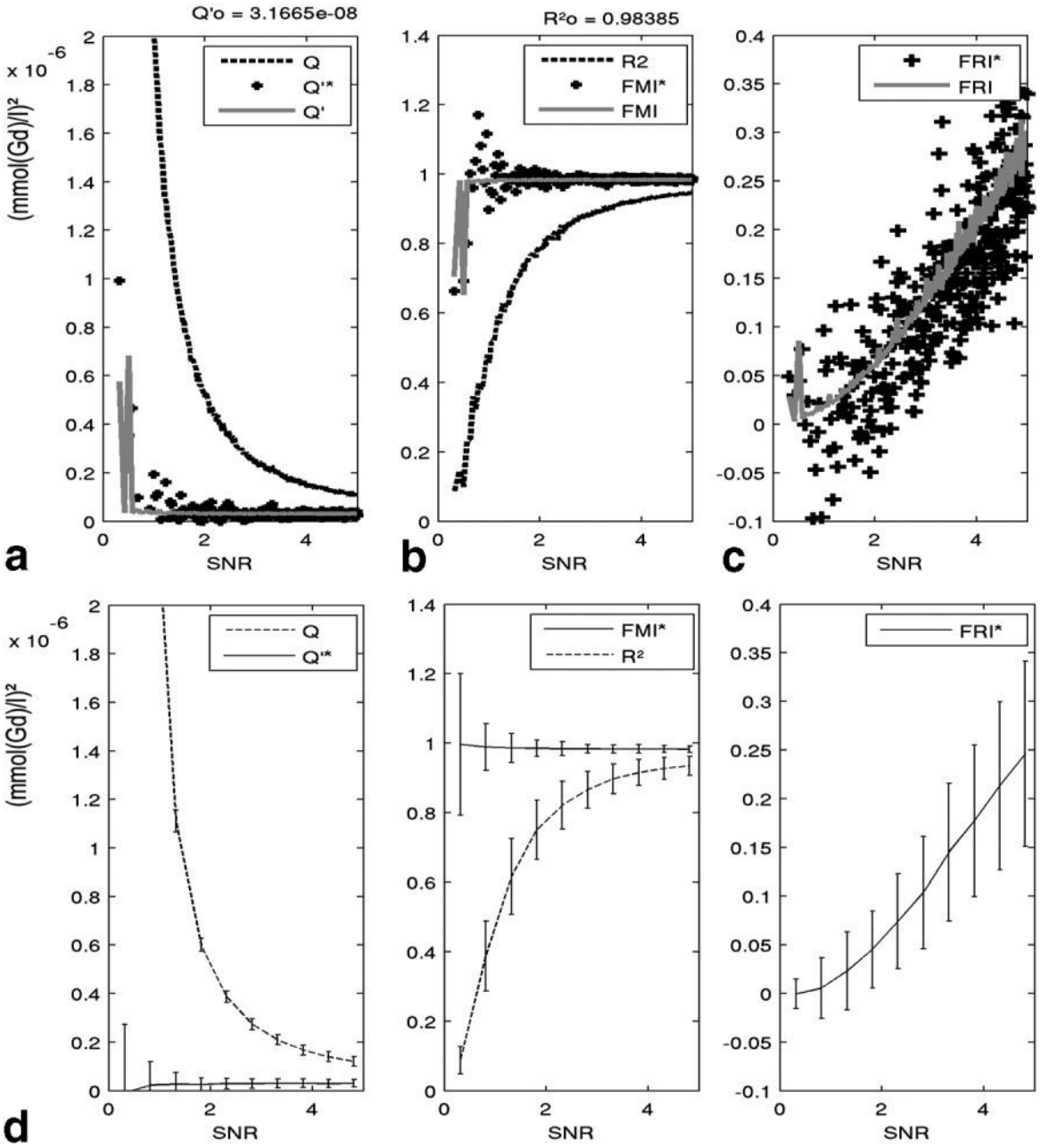

Figure 3. *Impact of the noise level on the criteria for simulated data. Data were simulated by using model M2 with the set of parameters of reference P_o and a Gaussian white noise generator. Data were fitted using M1.* ***a:*** *Quadratic errors (Q, Q'* and Q'), vs. the SNR of the data. Without noise the modeling error (Q'_o) is not zero.* ***b:*** *Correlation criteria fractions (R^2, FMI, FMI*) vs. SNR.* ***c:*** *Theoretical and*

estimated fractions of residual information (FRI, FRI) vs. SNR.* ***d:*** *Means and SDs of conventional (Q, R^2) and estimated criteria (Q'*, FMI*, FRI*) vs. SNR.*

These results were confirmed for 1000 sets of physiological parameters in the range $0.5 \times P_o$ to $1.5 \times P_o$. Figure 3d shows the means and standard deviations (SDs) of the new estimated and conventional criteria for this distribution of parameters. It shows that the sensitivity of FRI* for detecting modeling errors increases as SNR increases.

Results for Real MRI Data

Visual Analysis of Quality of Fit

A visual analysis of the fits of tissue enhancements obtained with the two models, examined for a large number of pixels, confirmed that model M1 yielded visually unsatisfactory fits for all mice, while model M2 provided visually satisfactory fits (Fig. 4). Therefore, one can estimate the capacity of the criteria to identify fits of different qualities in the visual reference method by studying their capacity to identify the model used for the fittings.

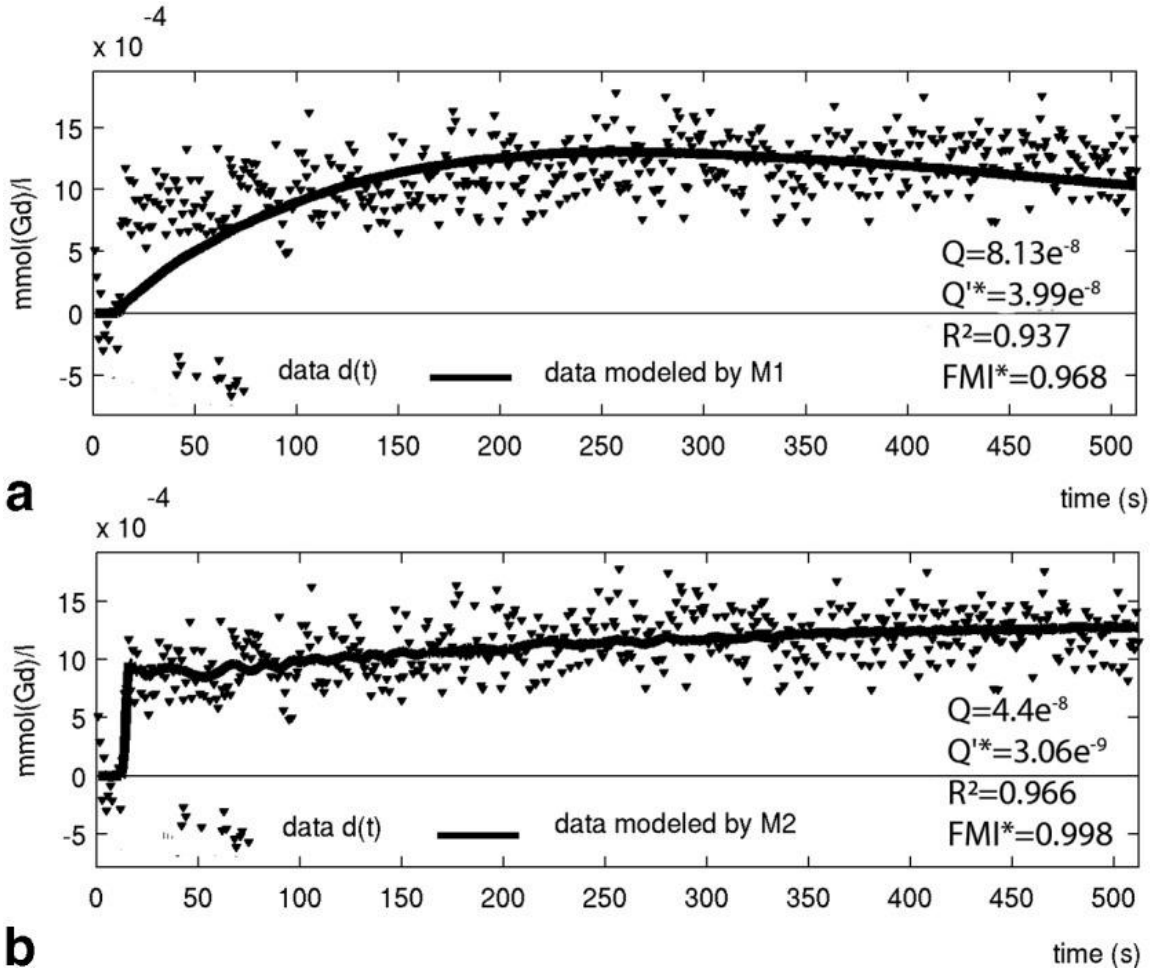

Figure 4. *Example of two fits obtained with models M1 and M2 on a pixel of tumor. The curves of measured concentrations are fitted with* ***(a)*** *a poorly adapted monocompartmental model (M1) and* ***(b)*** *a well-adapted bicompartmental model (M2).*

Analysis of criteria maps.

After the arterial and tissue regions (Fig. 5a) were selected, the pixels in the tissue region were masked. The physiological parameters were calculated by regression using the two models. In each map in Fig. 5b, three types of masked region are observable as blue pixels around the tumor (lower left: pulmonary region; right: region outside the tissue region, corresponding to air; within the tumor: region of low enhancement relative to noise (estimated SNR < 0.25)). The parametric maps of tissue perfusion and the blood volume fraction obtained with M1 (top) and M2 (bottom) are very different. This illustrates the impact of fit quality on the results. The maps of conventional criteria (Q, R^2) were poorly informative (see Fig. 6c). They indicated that the values of Q obtained with M2 are lower than the values obtained with M1 (while the values of R^2 are higher). However, this improvement yields poor sensitivity (since the relative variations are small), and it is not known whether it corresponds to a real improvement or to accidental modeling of the noise.

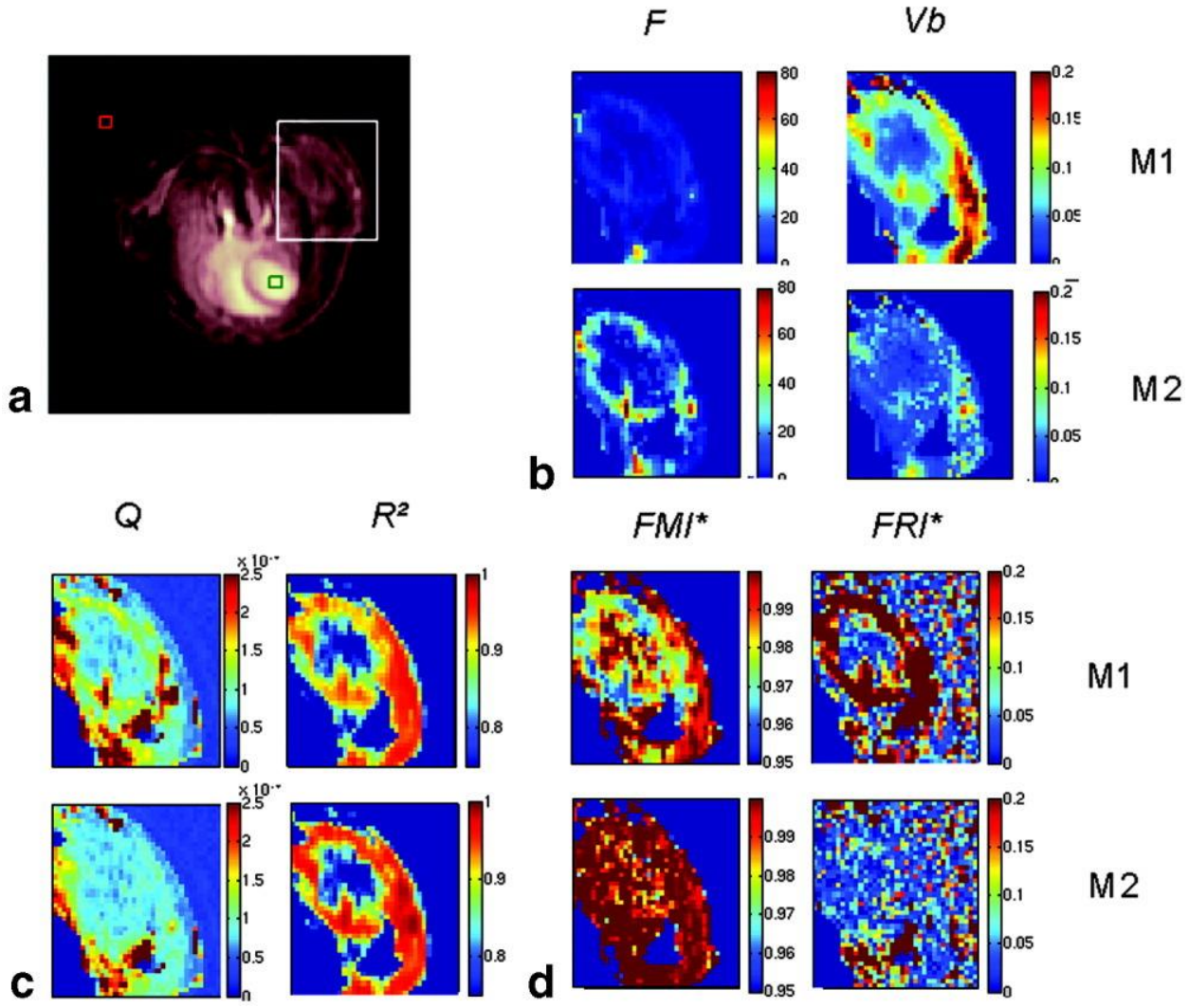

Figure 5. *Criteria maps constructed using models M1 and M2.* ***a:*** *Image with the selected arterial (left ventricle) tumor and noise regions.* ***b:*** *Maps of perfusion (F) and blood volume fraction (v_b).* ***c:*** *Maps of the quadratic error (Q) and correlation coefficient (R^2).* ***d:*** *Maps of FMI* and FRI*.*

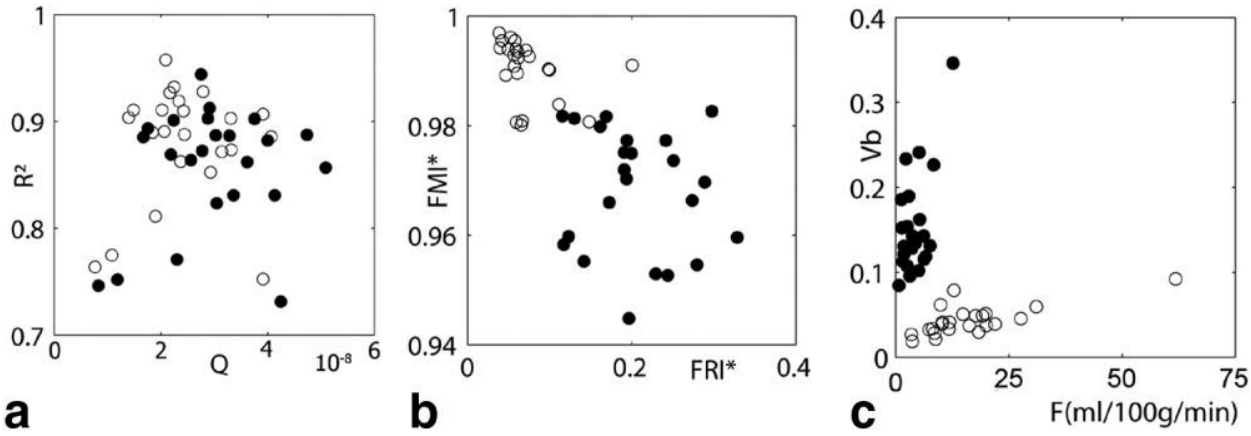

Figure 6. *Mean criteria and parameters of 23 tumors obtained with well-adapted M2 models (○) and poorly adapted M1 models (•).* ***a:*** *Conventional criteria (Q, R^2)* ***b:*** *New estimated criteria (FRI*, FMI*)* ***c:*** *Physiological parameters (F, v_b).*

The maps of estimated criteria (FRI*, FMI*) were far more sensitive than the maps of conventional criteria (Q, R^2) to detect improvements in the fits obtained with model M2 (Fig. 5d). The maps of estimated criteria therefore adequately distinguish fits of visually different qualities and thus allow one to correctly select the physiological maps obtained with model M2 (higher modeled information, lower residual information). They show that some spatial information on the tumor structure remains in the residual when M1 is used, but is taken into account when M2 is used.

The new criteria detected major modeling errors observed with M1 (especially in the hypervascularized periphery of the tumor) and small modeling errors observed with M2, in keeping with the results of the visual analysis. The FRI* maps indicated that the modeling error fell below the detection limit when model M2 was used (Fig. 5d). Indeed, the FRI* values in tissue pixels correspond to the FRI* values found in the air regions where there is no information.

Analysis of Mean Criteria

The mean conventional tumor criteria (Q, R^2) cannot distinguish parametric maps obtained with M1 from maps obtained with M2. As shown in Fig. 6a, the improvement in the criteria obtained using M2 appears negligible compared to intersubject variations. As a result, the mean conventional criteria cannot distinguish between the physiological results obtained with M1 and M2, which are in fact very different (Fig. 6c). As can be seen, each estimated criterion (FRI* or FMI*) permits a fit obtained with M1 to be distinguished from a fit obtained with M2, despite the variability of the mean SNR between the acquisitions (Fig. 6b). Fits obtained with M2 that are judged satisfactory by visual analysis are characterized by an FRI* close to zero and an FMI* close to one. Fits obtained with M1 that are unsatisfactory by visual analysis are characterized by higher FRI* and lower FMI* values. On average, the estimated criteria show a clear improvement in modeling when M2 is used, and detect the poor relevance of the physiological parameters obtained with model M1.

Analysis of Short Series

The new criteria maps obtained with the series of the first 51 s indicated that M1 was sufficient (Fig. 7d). The FMI* maps showed values close to one, and FRI* maps indicated that no more information was found for noise in air. This result was confirmed by the physiologic maps, which were close to those obtained

with M2 for the whole series (Fig. 7b). Conventional criteria maps are shown in Fig. 7c. On the basis of the R^2 maps, the M1 results for early series appear less relevant than the M1 results for the whole series (Fig. 5c), which is obviously not the case. Only the FMI* and FRI* maps are able to show that M1 is relevant for first-pass studies.

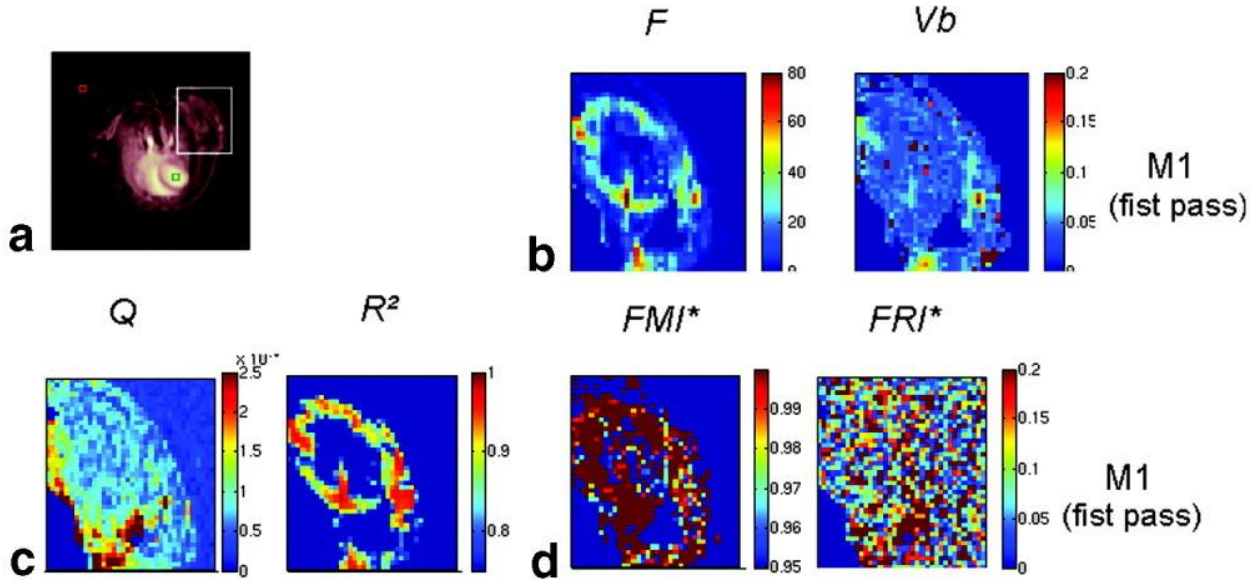

Figure 7. Results obtained with M1 on the first 51 images of the series. **a:** Image with the selected arterial (left ventricle) tumor and noise regions. **b:** Maps of perfusion (F) and blood volume fraction (v_b). **c:** Maps of the quadratic error (Q) and correlation coefficient (R^2). **d:** Maps of FMI* and FRI*.

DISCUSSION AND CONCLUSIONS

The limits of conventional criteria (Q, R^2) are theoretically evident, because the quality of fit and the SNR are not separated in Eqs. [4] and [5]. Our study using simulated data demonstrates that these theoretical limits generate serious problems for analyses of data from procedures such as CE-MRI. On average, the new criteria FMI* and FRI* are less dependent on random noise compared to the conventional criteria, over a wide range of SNR values, and are thus more focused on the quality of fit. The improvement provided by the new criteria for detecting modeling errors was subsequently verified with real data.

The results obtained with simulated and real data can be interpreted by using Eq. [4]. For simulated data, when d_o is constant, Q varies as $1/\mathrm{SNR}^2$. The autocorrelation function renders the noise component of Q essentially negligible. For real data, a significant decrease in the amplitude of e_{tot} (measured by FMI* and FRI*) corresponds to a nonsignificant decrease in Q because the random noise term remains constant. A typical error would be to choose M1 because Q and R^2 are not significantly improved by M2 when data are analyzed for an entire population of mice. However, the conventional criteria fail to show the improvement in fit because they are affected by the amplitude of random noise.

This does not mean that criteria using Q, such as R^2 or the scaled error, must not be used (35):

$$SFE = \sqrt{\frac{Q}{\|m\|^2}} \cdot 100\% \qquad (17)$$

These conventional criteria maps can be used; however, since they mix SNR and the modeling error, they are generally difficult to interpret. Indeed, they can be interpreted either as SNR or as fit-quality indicators,

depending on the balance between noise and the modeling error in the residual. The advantage of our new criteria is that random noise and quality of fit are separated. The R^2 maps can be readily interpreted as only SNR criteria if the corresponding FMI* maps are close to one (see Fig. 5). In this case the modeling error is indeed negligible, and Q is due solely to noise amplitude. In addition, FMI* and FRI* are estimators, and the quality of the estimates depends on the data and particularly the number of images in the series. In this respect, the detection limit of FRI* in a setting of Gaussian white noise can be associated with the Box-Pierce test (36). More generally, visual analysis of enhancement remains the qualitative reference method for determining appropriate criteria thresholds for accepting or rejecting the modeling. This qualitative approach is also important for understanding the nature of the modeling error detected by the criteria and for guiding the choice of the most suitable model. FMI* and FRI* can in fact be interpreted as quantifications of visual analysis, in that the operator generally compares the modeling data (m) with the general trend of the data (d_o*), in the same way as Q' is estimated. This is illustrated by the criteria in Fig. 4. Indeed, the visible fit error obtained with M1 is practically eliminated when M2 is used, and this is correctly associated with a 92% fall in Q', whereas Q falls by only 50%. This quantification implies that fits can be systematically verified in parametric imaging or on large series of regional studies.

The criteria FMI* and FRI* can be calculated quite rapidly. All of the parameter maps and criteria shown in Fig. 5 were obtained in less than 1 min on a 2 GHz Pentium® 4 computer running Matlab® software on 1140 pixels. The values of $\|d_o\|^{2*}$ required to calculate FMI* were also compared with those of $\|d\|^2$ in order to estimate the SNR in the data (SNR*). This estimate was used during the masking phase of the preprocessing to eliminate pixels with very low enhancement. These latter pixels are uninformative and processor time-consuming (see blue region within the tumor in Fig. 5).

The new criteria FRI* and FMI* indicate that model M1 is not relevant for the entire series because its gives absurd fits. Thus, FRI* and FMI* are useful for testing the flow-limited model, or more generally for testing the hypotheses used to simplify the model M2. The new criteria allow one to verify whether the tested model is able to explain the observations in a satisfactory manner, taking into account the random noise in the data set. This explanatory capacity depends on the model and tissue used, as well as the numerical integration of the model, the initial parameters of the chosen regression method, and the acquisition protocol. Errors can occur at all stages of the calculation, and the new criteria allow one to verify whether the results of the calculations are in line with the data set. For example, in studies of PC3 tumors in mice given Vistarem® bolus injections, the new criteria indicate that M1 is adapted to series of about 1 min (Fig. 7), and that M2 is better adapted to series of about 10 min (Fig. 5). This indicates that permeability information is expressed in the data, but not during the first minute. The ability of FMI* and FRI* to distinguish fits according only to the model a posteriori shows that the temporally correlated noise bc is small enough not to disrupt the results.

These new criteria for testing quality of fit complement other existing tools. For proper quantification of parametric values, it is commonly assumed that 1) the data are not degraded, 2) the data information is correctly extracted to model parameters, and 3) the parameters are interpretable. The main factor responsible for degraded data is the presence of temporally correlated noise (bc). The identification and elimination of noise bc is specific to each data set and must be done before modeling is performed, for example by using registration techniques for body motion, nonlinear conversion from MR intensity to the concentration of contrast agent, or partial volume correction for arteries with small diameters (37). The second item corresponds to the new criteria FMI* and FRI*. Finally, the third item corresponds to the question of identifiability, which involves verifying whether each physiological phenomenon (d_o) is represented by a unique set of parameters, and whether the SDs of these parameters are within an acceptable range. Identifiability can be studied with a linear approximation of m variations to estimate the Fisher information matrix (F_m) (38). F_m is generally used to detect overcomplex models. It can be combined with the new criteria FMI* and FRI* to find a model adapted to the complexity of the data. The relevance of the

data analysis does not mean that the parameters correspond to “real” physiological characteristics, however. For one set of data the correlation analysis between the modeling parameters and the corresponding gold standard measurements corresponds to a validation test. When reliable gold standard measurements are not available, validation can be approximated by replacing real data with realistic modeled data (39). Validation and verification are thus complementary studies. Indeed, proper validation always implies a good data analysis, while the analysis can be verified but not validated for each new data set. By testing the goodness of fit, the new criteria FMI* and FRI* clearly reinforce the quantitative credibility of physiologic parameters obtained by modeling CE-MRI data sets.

Acknowledgements

The authors thank David Young for translating this paper.

References

REFERENCES

1. Toft PS, Kermode AG. Measurement of the blood–brain permeability and leakage space using dynamic MR imaging. Magn Reson Med 1991;1:357–367.
2. Lucht RE, Knopp MV, Brix G. Classification of signal-time curves from dynamic MR mammography by neural networks. Magn Reson Imaging 2001;19:51–57.
3. Clement O, Siauve N, Lewin M, de Kerviler E, Cuenod CA, Frija G. Contrast agents in magnetic resonance imaging of the liver: present and future. Biomed Pharmaco ther 1998;52:51–58.
4. Luypaert R, Boujraf S, Sourbron S, Osteaux M. Diffusion and perfusion MRI: basic physics. Eur J Radiol 2001;38:19–27.
5. Miles K, Dawson P, Blomley M. Functional computed tomography. Oxford: SIS Medical Media; 1997.
6. Lee TY. Functional CT: physiological models. trends in biotechnology 2002;20:s3–s10.
7. Jacquez JA. Compartmental analysis in biology and medicine. Ann Arbor: University of Michigan Press; 1985.
8. Sandberg IS. On mathematical foundations of compartmental analysis in biology, medicine, and ecology. IEEE Trans Circ Syst 1978;25:273–279.
9. Padhani AR. Dynamic contrast-enhanced MRI in clinical oncology: current status and future directions. J Magn Reson Imaging 2002;16: 407–422.
10. Knopp MV, Weiss E, Sinn HP, Mattern J, Junkermann H, Radeleff J, Magener A, Brix G, Delorme S, Zuna I, van Kaick G. Pathophysiologic basis of contrast enhancement in breast tumors. J Magn Reson Imaging 1999;10:260–266.
11. Abbott NJ, Chugani DC, Zaharchuk G, Rosen BR, Lo EH. Delivery of imaging agents into brain. Adv Drug Deliv Rev 1999;37:253–277.
12. Moran GR, Thornhill RE, Sykes J, Prato FS. Myocardical viability imaging using Gd-DTPA: physiological modeling of infracted myocardium, and impact on injection stategy and imaging time. Magn Reson Imaging 2002;48:791–800.
13. Fritz-Hansen T, Rostrup E, Sondergaard L, Ring PB, Amtorp O, Larsson HB. Capillary transfer constant of Gd-DTPA in the myocardium at rest and during vasodilation assessed by MRI. Magn Reson Med 1998;40: 922–929.
14. Vallee JP, Lazeyras F, Kasuboski L, Chatelain P, Howarth N, Righetti A, Didier D. Quantification of myocardial perfusion with FAST sequence and Gd bolus in patients with normal cardiac function. J Magn Reson Imaging 1999;9:197–203.
15. Noworolski SM, Fischbein NJ, Kaplan MJ, Lu Y, Nelson SJ, Carvajal L, Henry RG. Challenges in dynamic contrast-enhanced MRI imaging of cervical lymph nodes to detect metastatic disease. J Magn Reson Imaging 2003;17:455–462.

16. Chan JH, Tsui EY, Chau LF, Chow KY, Chan MS, Yuen MK, Chan TL, Cheng WK, Wong KP. Discrimination of an infected brain tumor from a cerebral abscess by combined MR perfusion and diffusion imaging. Comput Med Imaging Graph 2002;26:19–23.
17. Preul C, Ku¨hn B, Lang EW, Mehdorn HM, Heller M, Link J. Differentiation of cerebral tumors using multi-section echo planar MR perfusion imaging. Eur J Radiol 2003;48:244–251.
18. Brasch R, Turetschek K. MRI characterization of tumors and grading angiogenesis using macromolecular contrast media: status report. Eur J Radiol 2000;34:148–155.
19. Egmont-Petersen M, Hogendoorn PC, van der Geest RJ, Vrooman HA, van der Woude H, Janssen JP, Bloem JL, Reiber JH. Detection of areas with viable remnant tumor in postchemotherapy patients with Ewing's sarcoma by dynamic contrast-enhanced MRI using pharmacokinetic modeling. Magn Reson Imaging 2000;18:525–535.
20. Hawighorst H, Engenhart R, Knopp MV, Brix G, Grandy M, Essing M, Miltner P, Zuna I, Fuss M, Kaick Gv. Intracranial meningeomas: time and dose dependent effects of irradiation on tumor microcirculation monitored by dynamic MR imaging. Magn Reson Imaging 1997;15:423–432.
21. Bilgen M, Dogan B, Narayana PA. In vivo assessment of blood–spinal cord barrier permeability: serial dynamic contrast enhanced MRI of spinal cord injury. Magn Reson Imaging 2002;20:337–341.
22. Davidson J. Econometric theory. Oxford: Blackwell Publishers; 2003.678 p.
23. Bogin L, Margalit R, Mispelter J, Degani H. Parametric imaging of tumor perfusion using flow- and permeability-limited tracers. J Magn Reson Imaging 2002;16:289–299.
24. Zhu XP, Li KL, Kamaly-Asl ID, Checkley DR, Tessier JJ, Waterton JC, Jackson A. Quantification of endothelial permeability, leakage space, and blood volume in brain tumors using combined T1 and T2* contrast-enhanced dynamic MR imaging. J Magn Reson Imaging 2000;11:575–585.
25. Su MY, Jao JC, Nlcioglu O. Measurement of vascular volume fraction and blood–tissue permeability constants with a pharmacokinetic model: studies in rat muscle tumor with dynamic Gd-DTPA enhances MRI. Magn Reson Med 1994;32:741–724.
26. Ober RJ. The Fisher information matrix for linear systems. Syst Control Lett 2002;47:221–226.
27. Retout S, Duffull S, Mentre´ F. Development and implementation of the population Fisher information matrix for the evaluation of population pharmacokinetic designs. Comput Methods Programs Biomed 2001;65:141–151.
28. Asprey SP, Macchietto S. Designing robust optimal dynamic experiments. J Process Control 2002;12:545–556.29. Wu G. Sensitivity analysis of pharmacokinetic parameters in one compartment models. Pharmacol Res 2000;41:445–453.
30. Brix G, Bahner ML, Hoffmann U, Horvath A, Schreiber W. Regional blood flow, capillary permeability, and compartmental volumes: measurement with dynamic CT—initial experience. Radiology 1999;210:269–276.
31. Zheng D, Upton RN, Ludbrook GL, Martinez A. Acute cardiovascular effects of magnesium and their relationship to systemic and myocardial magnesium concentrations after short infusion in awake sheep. J Pharmacol Exp Ther 2001;297:1176–1183.
32. Kaighn M, Narayan K, Ohnuki Y, Lechner J, Jones L. Establishment and characterization of a human prostatic carcinoma cell line (PC-3). Invest Urol 1979;17:16:23.
33. Port M, Corot C, Rousseaux O, Raynal I, Devoldere L, Idee JM, Dencausse A, Le Greneur S, Simonot C, Meyer D. P792: a rapid clearance blood pool agent for magnetic resonance imaging: preliminary results. MAGMA 2001;12:121–127.
34. Pradel C, Siauve N, Bruneteau G, Clement O, Bazelaire CD, Frouin F, Wedge SR, Tessier JL, Robert PH, Frija G, Cuenod CA. Reduced capillary perfusion and permeability in human tumour xenografts treated with the VEGF signalling inhibitor ZD4190: an in vivo assessment using dynamic MR imaging and macromolecular contrast media. Magn Reson Imaging 2003;21:845:851.
35. Li KL, Zhu XP, Jackson A. Parametric mapping of scaled fitting error in dynamic susceptibility contrast enhanced MR perfusion imaging. Br J Radiol 2000;73:470–481.
36. Tagaris GA, Richter W, Kim S-G, Georgopoulos AP. Box-Jenkins intervention analysis of functional resonance imaging data. Neurosci Res 1996;27:289–294.

37. Adam JF, Elleaume H, Le Duc G, Corde S, Charvet AM, Tropres I, Le Bas JF, Esteve F. Absolute cerebral blood volume and blood flow measurements based on synchrotron radiation quantitative computed tomography. J Cereb Blood Flow Metab 2003;23:499–512.
38. Dokos S, Lovell NH. Parameter estimation in cardiac ionic models. Prog Biophys Mol Biol 2004;85:407–431.
39. Buckley DL. Uncertainty in the analysis of tracer kinetics using dynamic contrast-enhanced T1-weighted MRI. Magn Reson Med 2002;47:601–606.