

# Incremental parameter evaluation from incomplete data with application to the population pharmacology of anticoagulants

Marcel O. Vlad<sup>\*†‡</sup>, Alexandru Dan Corlan<sup>§</sup>, Federico Morán<sup>¶</sup>, Peter Oefner<sup>||</sup>, and John Ross<sup>\*†</sup>

<sup>\*</sup>Department of Chemistry, Stanford University, Stanford CA 94305-5080; <sup>†</sup>Institute of Mathematical Statistics and Applied Mathematics, Casa Academiei Romane, Calea 13 Septembrie 13, Bucharest, 050711 Romania; <sup>§</sup>Cardiology Research Unit, University Emergency Hospital, Romanian Academy of Medical Sciences, 169 Splaiul Independentei, Bucharest 050098, Romania, and Department of Enzymology, Institute of Biochemistry of the Romanian Academy, Splaiul Independentei 296, Bucharest 060031, Romania; <sup>¶</sup>Departamento de Bioquímica y Biología Molecular I, Universidad Complutense Madrid, 28040 Madrid, Spain and Centro de Astrobiología (CSIC-INTA), Carretera de Ajalvir, Km 4, 28850 Torrejón de Ardoz, Madrid, Spain; and <sup>||</sup>Institute of Functional Genomics, University of Regensburg, Josef-Engert Strasse 9, 93053 Regensburg, Germany

Contributed by John Ross, December 6, 2007 (sent for review October 18, 2007)

We develop a method for parameter evaluation from incomplete data. Improved estimates of the desired parameters are evaluated step by step, from experiment to experiment by using both Bayesian and informational methods. We make dynamical, improved predictions while the experiments are still going on and keep and interpret information about local fluctuations, which is lost on applying global techniques. The input of information in small packets leads to semi-analytic methods for data processing. An evolution criterion for parameter evaluation, similar to Fisher's theorem of population selection, is derived. We develop direct processing methods, which can be applied to low dimensional systems, semi-analytic methods based on direct or double logarithmic phase expansions, steepest descent approaches, variation and perturbation methods. The techniques are illustrated by developing a method of long-term planning of treatments with oral anticoagulants based on limited clinical data. The efficiency of treatment by oral anticoagulants depends strongly on various anthropometric and genotypic factors, which lead to large variations of the clinical response. We use the clinical data, which accumulates from medical consultations, for extracting improved, incremental information about the statistical properties of the kinetic and anthropometric parameters for a given patient, which in turn is used for making repeated, improved clinical predictions as the treatment proceeds.

Various scientific problems involve the evaluation in time and/or space of parameters, which are not direct observables. The input information is contained in observables, which depend on the desired parameters, deterministically or stochastically. In many cases the dynamics of the process is not properly known and the available experimental data is either/both incomplete and inaccurate. Estimation theory (1) is a branch of statistics and signal processing that deals with these types of problems. Even though there are applications in physics and chemistry, estimation theory is usually not part of the training of a physicist or a chemist. In this paper we develop an approach which bridges the gap between statistical physics and estimation theory, by emphasizing a probabilistic scientific approach (2). Suitable examples of our approach are reaction-diffusion processes in random environments where the rate and transport coefficients are random, large biochemical networks in *in vivo* conditions, genetic systems involving multiple genes, gene regulation networks, etc. In this paper, we present a case study of the population pharmacology of anticoagulants.

## Statistical Mechanical Approach to Incremental Estimation

We consider a process characterized by a set of parameters  $\mathbf{w} = (w_1, \dots, w_m)$ ;  $\mathbf{w} = \mathbf{w}(\rho)$  is in general a field, that is, it depends on a state vector  $\rho$  such as a position vector,  $\rho = (\mathbf{r})$ , in real space, the time,  $\rho = (t)$ , the space-time continuum,  $\rho = (\mathbf{r}, t)$ , or an abstract state space. Parameters  $\mathbf{w}$  may be time and space-

dependent rate or transport coefficients for a transformation and transport process occurring in a random environment, or growth and migration coefficients for human populations in genetics. The experimental observables  $\mathbf{c} = (c_1, \dots, c_m) = \mathbf{c}(\rho)$  are also fields depending on a position vector  $\rho$ , such as concentrations, population densities, gene frequencies, etc. We assume that the experiments are incremental, that is, they involve more and more numbers of state vectors. The incremental nature of an experiment is obvious if the position vector is the time,  $\rho = (t)$ ; in this case, successive measurements  $q = 1, 2, \dots$ , take place at different times and as time increases, each measurement adds new information. The experiments can be carried out in an incremental way, even if the possible position vectors cannot be ordered. For example, in genetic studies of a given mutation, information from different countries can be collected and processed incrementally.

From successive measurements  $q = 1, 2, \dots$ , we collect incremental information about the vector  $\mathbf{c} = \mathbf{c}(\rho)$  at different positions  $\rho = \rho_1, \rho_2, \dots$ , and we want to estimate the properties of the parameters  $\mathbf{w} = \mathbf{w}(\rho)$ . In general, the data collected is not enough for the evaluation of  $\mathbf{w} = \mathbf{w}(\rho)$ . Our purpose is to fill out the gaps in the data and come up with an estimation approach, which replaces our ignorance by an unbiased guess. Such an unbiased guess can be based on the optimization of the information gain between two successive measurements. We denote by  $\mathbf{W} = [\mathbf{w}(\rho_u)]_{u=1, \dots, \sigma_\Sigma}$  and  $\mathbf{C} = [\mathbf{c}(\rho_u)]_{u=1, \dots, \sigma_\Sigma}$  the matrices of the parameters and of the observables, where  $\sigma_\Sigma$  is the total number of position vectors used in our description of the system and introduce the corresponding probability densities  $P_X^{(q)}(\mathbf{X})$ ,  $\mathbf{X} = \mathbf{W}, \mathbf{C}$  with the normalization conditions  $\int P_X^{(q)}(\mathbf{X}) d\mathbf{X} = 1$ ,  $\mathbf{X} = \mathbf{W}, \mathbf{C}$ ; we also introduce the joint probability densities  $R_{\mathbf{W}\mathbf{C}}^{(q)}(\mathbf{W}; \mathbf{C})$  with  $\int R_{\mathbf{W}\mathbf{C}}^{(q)}(\mathbf{W}; \mathbf{C}) d\mathbf{W} d\mathbf{C} = 1$  as well as the conditional probabilities  $B_{\mathbf{W}|\mathbf{C}}^{(q)}(\mathbf{W}|\mathbf{C})$  and  $B_{\mathbf{C}|\mathbf{W}}^{(q)}(\mathbf{C}|\mathbf{W})$  with  $\int B_{\mathbf{W}|\mathbf{C}}^{(q)}(\mathbf{W}|\mathbf{C}) d\mathbf{W} = 1$ ,  $\int B_{\mathbf{C}|\mathbf{W}}^{(q)}(\mathbf{C}|\mathbf{W}) d\mathbf{C} = 1$ , respectively. The information gained for the random matrices  $\mathbf{W}$  and  $\mathbf{C}$  from the  $q$ th to the  $(q + 1)$ th measurement are given by:

$$\mathcal{I}_X^{(q+1)} = \int P_X^{(q+1)}(\mathbf{X}) \ln [P_X^{(q+1)}(\mathbf{X}) / P_X^{(q)}(\mathbf{X})] d\mathbf{X},$$

with  $\mathbf{X} = \mathbf{W}$  or  $\mathbf{C}$ . [1]

Author contributions: M.O.V. and A.D.C. designed research; M.O.V., A.D.C., F.M., P.O., and J.R. performed research; M.O.V., A.D.C., F.M., P.O., and J.R. analyzed data; and M.O.V., A.D.C., F.M., P.O., and J.R. wrote the paper.

The authors declare no conflict of interest.

<sup>†</sup>To whom correspondence may be addressed. E-mail: marceluc@stanford.edu or john.ross@stanford.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0711508105/DC1](http://www.pnas.org/cgi/content/full/0711508105/DC1).

© 2008 by The National Academy of Sciences of the USA

We search for the extremum of  $\mathcal{H}_W^{(q+1)}$ , subjected to two types of constraints:  $\int P_W^{(q+1)}(\mathbf{W})d\mathbf{W} = 1$ , which expresses the normalization condition for  $P_W^{(q+1)}(\mathbf{W})$ , and  $P_C^{(q+1)}(\mathbf{C}) = \int R_{WC}^{(q+1)}(\mathbf{W}; \mathbf{C})d\mathbf{W} = \int B_{C|W}^{(q+1)}(\mathbf{C}|\mathbf{W})P_W^{(q+1)}(\mathbf{W})d\mathbf{W}$ , which expresses the fact that the probability density  $P_C^{(q+1)}(\mathbf{C})$  of the measured quantities can be expressed as an integral of the joint probability density  $R_{WC}^{(q+1)}(\mathbf{W}; \mathbf{C}) = B_{C|W}^{(q+1)}(\mathbf{C}|\mathbf{W})P_W^{(q+1)}(\mathbf{W})$  of the observables  $\mathbf{C}$  and the parameters  $\mathbf{W}$  over all possible values of the parameters  $\mathbf{W}$ . This type of extremum problem is similar to the information approach to nonequilibrium statistical ensembles of the Zubarev–McLennan type (3–5). Its solution is:

$$P_W^{(q+1)}(\mathbf{W}) = [\mathcal{Z}^{(q+1)}[\lambda^{(q+1)}(\mathbf{C})]]^{-1}P_W^{(q)}(\mathbf{W}) \exp\left\{-\int B_{C|W}^{(q+1)}(\mathbf{C}|\mathbf{W})\lambda^{(q+1)}(\mathbf{C})d\mathbf{C}\right\}, \quad [2]$$

where

$$\mathcal{Z}^{(q+1)}[\lambda^{(q+1)}(\mathbf{C})] = \int_{\mathbf{W}} P_W^{(q)}(\mathbf{W}) \exp\left\{-\int B_{C|W}^{(q+1)}(\mathbf{C}|\mathbf{W})\lambda^{(q+1)}(\mathbf{C})d\mathbf{C}\right\}d\mathbf{W}, \quad [3]$$

are partition functionals and  $\lambda^{(q+1)}(\mathbf{C})$  are test functions, originating from Lagrange multipliers, which are the solutions of the functional equations

$$-\frac{\delta}{\delta\lambda^{(q+1)}(\mathbf{C}')} \ln \mathcal{Z}^{(q+1)}[\lambda^{(q+1)}(\mathbf{C})] = \langle B_{C|W}^{(q+1)}(\mathbf{C}'|\mathbf{W}) \rangle_{\mathbf{W}}. \quad [4]$$

The ensemble average  $\langle \dots \rangle_{\mathbf{W}}$  is taken over all possible values of  $\mathbf{W}$ .

Eqs. 2–4 provide a prescription for the incremental evaluation of the statistical properties of the parameters  $\mathbf{W}$ . From measurement to measurement, the probability densities  $P_W^{(q+1)}(\mathbf{W})$  are scaled by multiplicative factors, which depend on the parameters  $\mathbf{W}$ , and then renormalized by a partition functional, which plays the role of preserving the normalization condition for the probability densities. We represent the scaling factors as products of amplitude and phase terms:  $T_W^{(q)}(\mathbf{W}) = \mathcal{A}^{(q)} \exp(-\varphi^{(q)}(\mathbf{W}))$ .  $T_W^{(q)}(\mathbf{W})$  are nonnegative; to preserve the nonnegativity of the probability densities, the exponentials  $\exp(-\varphi^{(q)}(\mathbf{W}))$  are real, and from Eq. 2 it follows that the phase factors are given by  $\varphi^{(q)}(\mathbf{W}) = \int B_{C|W}^{(q+1)}(\mathbf{C}|\mathbf{W})\lambda^{(q+1)}(\mathbf{C})d\mathbf{C}$ . They are real and can be either positive or negative; due to the presence of the partition functionals in Eq. 2, the values of the amplitude factors are irrelevant. Eqs. 2 and 3 turn into a simpler form:

$$P_W^{(q+1)}(\mathbf{W}) = [\mathcal{Z}^{(q+1)}]^{-1}P_W^{(q)}(\mathbf{W}) \exp[-\varphi^{(q)}(\mathbf{W})], \quad [5]$$

$$\mathcal{Z}^{(q+1)} = \int_{\mathbf{W}} P_W^{(q)}(\mathbf{W}) \exp[-\varphi^{(q)}(\mathbf{W})]d\mathbf{W}. \quad [6]$$

Eqs. 5 and 6 can be also derived by using a Bayesian method (6). The conditional probability density  $B_{W|C}^{(q)}(\mathbf{W}|\mathbf{C})$  at the  $q$ th measurement can be considered as *a posteriori*, unconditional probability density for the next,  $(q + 1)$ th measurement,  $B_{W|C}^{(q)}(\mathbf{W}|\mathbf{C}) = P_W^{(q+1)}(\mathbf{W})$ . From the Bayes rule  $B_{W|C}^{(q)}(\mathbf{W}|\mathbf{C})P_C^{(q)}(\mathbf{C}) = B_{C|W}^{(q)}(\mathbf{C}|\mathbf{W})P_W^{(q)}(\mathbf{W})$ , combined with the expression for  $P_C^{(q)}(\mathbf{C})$  as a marginal probability,  $P_C^{(q)}(\mathbf{C}) = \int B_{C|W}^{(q)}(\mathbf{C}|\mathbf{W})P_W^{(q)}(\mathbf{W})d\mathbf{W}$ , we come to Eqs. 5 and 6, where the phase factor  $\varphi^{(q)}(\mathbf{W})$  now depends both on  $\mathbf{W}$  and  $\mathbf{C}$ . We have

$\varphi^{(q)}(\mathbf{C}|\mathbf{W}) = -\ln B_{C|W}^{(q)}(\mathbf{C}|\mathbf{W})$ . The Bayesian approach leads to different estimates, depending on the fluctuating values of the observables  $\mathbf{C}$ . The usual choice is to consider the average values of the observables  $E(\mathbf{C}) = \langle \mathbf{C} \rangle$ ; if this choice is made, the phase factor  $\varphi^{(q)}(\mathbf{W})$  depends only on the values of the parameters  $\mathbf{W}$ .

In both cases, the iterations in Eqs. 5 and 6 can be carried out explicitly, resulting in

$$P_W^{(q+1)}(\mathbf{W}) = [\mathcal{Z}_{\Sigma}^{(q+1)}]^{-1}P_W^{(1)}(\mathbf{W}) \exp[-\Phi^{(q)}(\mathbf{W})], \quad [7]$$

$$\mathcal{Z}_{\Sigma}^{(q+1)} = \int_{\mathbf{W}} P_W^{(1)}(\mathbf{W}) \exp[-\Phi^{(q)}(\mathbf{W})]d\mathbf{W}, \quad [8]$$

where  $\Phi^{(q)}(\mathbf{W}) = \sum_{q'=1}^q \varphi^{(q')}(\mathbf{W})$  are global phase factors. For large numbers of experiments, we expect that the contributions  $\varphi^{(q')}(\mathbf{W})$  of the different experiments are smaller compared to the total phase factor,  $\Phi^{(q)}(\mathbf{W}) \gg \varphi^{(q')}(\mathbf{W})$ . Furthermore, for a large number of experiments, we expect that the information gain acquired in a given experiment  $\mathcal{H}_W^{(q+1)} = \int P_W^{(q+1)}(\mathbf{W}) \ln[P_W^{(q+1)}(\mathbf{W})/P_W^{(q)}(\mathbf{W})]d\mathbf{W} \geq 0$ , which is always nonnegative (2), is smaller than the total information gain,  $\mathcal{H}_W^{(q+1)} = \sum_{q=2}^{q+1} \mathcal{H}_W^{(q)}$ . In general,  $\mathcal{H}_W^{(q+1)}$  is different from the information divergence computed from the probability density  $P_W^{(q+1)}(\mathbf{W})$  at the  $(q + 1)$ th experiment and the probability density  $P_W^{(1)}(\mathbf{W})$  at the first experiment. These assumptions of smallness lead to semi-analytic approaches for the incremental processing of experimental data.

Eqs. 5 and 6 have a similar structure as the evolution equations for gene frequencies in population genetics, which suggests the existence of an evolution criterion similar to Fisher’s fundamental theorem of natural selection (7–9). We introduce the distortion factor  $\rho^{(q)}(\mathbf{W}) = \exp(-\varphi^{(q)}(\mathbf{W})) - 1$ , which is a measure of the rate of change of the probability density of  $\mathbf{W}$  from the  $q$ th to the  $(q + 1)$ th experiment, and is the analog of the intrinsic rate of growth of a genotype from the  $q$ th to the  $(q + 1)$ th generation. Similarly,  $-\varphi^{(q)}(\mathbf{W})$  is the analog of the intrinsic rate of growth of a genotype; for small changes of the probability density, we have  $\rho^{(q)}(\mathbf{W}) = -\varphi^{(q)}(\mathbf{W})$ . The variation of the average values  $\rho^{(q)}(\mathbf{W})$  or  $\varphi^{(q)}(\mathbf{W})$  are mean fitness variables that provide information about the evolution of the process of incremental evaluation of state probabilities. By using Eqs. 5 and 6, we can derive the following relation for the variation of the average value of the distortion  $\rho^{(q)}(\mathbf{W})$  from experiment to experiment:

$$\langle \rho^{(q+1)}(\mathbf{W}) \rangle_{q+1} - \langle \rho^{(q+1)}(\mathbf{W}) \rangle_q = \frac{\langle \Delta \rho^{(q)}(\mathbf{W}) \Delta \rho^{(q+1)}(\mathbf{W}) \rangle_q}{1 + \langle \rho^{(q)}(\mathbf{W}) \rangle_q}, \quad [9]$$

where  $\langle \dots \rangle_q = \int \dots P_W^{(q)}(\mathbf{W})d\mathbf{W}$ . Eq. 9 is a generalized Fisher theorem that states that the variation of the growth factor from the  $q$ th to the  $(q + 1)$ th experiment is proportional to the correlation function of the growth factor at the  $q$ th and the  $(q + 1)$ th experiment evaluated in terms of  $P_W^{(q)}(\mathbf{W})$ . The case for which the growth factors are the same for all experiments,  $\rho^{(q+1)}(\mathbf{W}) = \rho(\mathbf{W})$ , is analogous to constant selection in population genetics and  $\langle \rho(\mathbf{W}) \rangle_{q+1} - \langle \rho(\mathbf{W}) \rangle_q = \langle [\Delta \rho(\mathbf{W})]^2 \rangle_q [1 + \langle \rho(\mathbf{W}) \rangle]$ , that is, the variation from experiment to experiment of the average growth factor is proportional to its dispersion. Because the dispersion is nonnegative, it follows that the average distortion factor increases until it reaches a maximum value, which corresponds to the absolute maximum value of  $\rho(\mathbf{W})$ ,  $\max[\rho(\mathbf{W})]$  (if there is one). If there is a single value  $\mathbf{W}_{\max}$  of  $\mathbf{W}$ , which corresponds to  $\max[\rho(\mathbf{W})]$ , then in the long run, for large numbers  $q$  of experiments,  $P_W^{(q)}(\mathbf{W})$  tends toward  $\delta(\mathbf{W} - \mathbf{W}_{\max})$ ; this situation is similar to the evolution of a population under constant selection pressure, where the genotype with the biggest fitness prevails in the long run. In general, the distortion factor

depends on the experiment and thus the estimates of the parameters do not tend towards deterministic values.

The direct application of Eqs. 5 and 6 is limited to the analysis of a limited number of parameters. For large systems, we have developed semi-analytical methods, which deal with cumulants and moments, reducing the dimension and complexity of calculations. These methods are based on the assumption of dealing with a large number of experiments for which the individual phase factors  $\varphi^{(q)}(\mathbf{W})$  of the different experiments are smaller than the total phase factor,  $\Phi^{(q)}(\mathbf{W}) \gg \varphi^{(q)}(\mathbf{W})$ . Based on this inequality, we developed various field theoretical approaches (10, 11) for computing the partition functionals, from which we derived iterative equations for the cumulants of  $\mathbf{W}$ , which can be used in numerical computations. In the following, we summarize the basics of the methods developed and give a few details regarding one of them.

If the phase factors depend weakly on the parameters  $\mathbf{W}$ , then  $\varphi^{(q)}(\mathbf{W})$  can be approximated by linear and quadratic terms of their Taylor expansions. We assume that  $\varphi^{(q)}(\mathbf{W}) \approx \sum_{\alpha,u} \chi_{\alpha,u}^{1(q)} w_{\alpha}(\rho_u) + \sum_{\alpha_1,u_1;\alpha_2,u_2} \chi_{\alpha_1,u_1;\alpha_2,u_2}^{2(q)} w_{\alpha_1}(\rho_{u_1}) w_{\alpha_2}(\rho_{u_2})$ , where,  $\chi_{\alpha,u}^{1(q)} = \partial_{w_{\alpha}(\rho_u)} \varphi^{(q)}$ ,  $\chi_{\alpha_1,u_1;\alpha_2,u_2}^{2(q)} = \frac{1}{2} \partial_{w_{\alpha_1}(\rho_{u_1})} \partial_{w_{\alpha_2}(\rho_{u_2})} \varphi^{(q)}$  are susceptibility functions. The linear phase approximation retains only the linear terms, and the partition functionals can be computed exactly for any probability densities; moreover, the functional iteration can be solved explicitly. The quadratic approximation leads to Gaussian integrals and the partition and characteristic functionals can be also computed analytically. We developed two different approaches: (i) If the initial distributions have a number of well separated maxima, then the method of steepest descent can be used for evaluating the partition functionals. (ii) We assumed that the initial distributions can be expressed as mixtures of Gaussians and computed the resulting path integrals. These Gaussian methods involve only cumulants of first- and second-order and all higher order cumulants are zero, resulting in a reduction of the number of variables and simpler numerical computations.

Gaussian methods fail to represent properly the statistical properties of nonnegative random variables. These limitations may be circumvented by using logarithmic transformations of the desired parameters  $\theta_{\alpha}(\rho_u) = \ln[w_{\alpha}(\rho_u)/w_{\alpha}^0(\rho_u)]$ , where  $w_{\alpha}^0(\rho_u)$  are reference parameters. We assume that the phase factors are analytic in  $\theta_{\alpha}(\rho_u)$  and the Gaussian methods are replaced by multidimensional log-normal methods. All the results obtained in the Gaussian case can be easily translated to the log-normal case.

We give a few details about the linear phase approximation; the results can be easily extended to the case of log-linear phase approximation, where the phase factors are proportional to the logarithms of the desired parameters. In the linear phase approximation, there is a simple relation between the partition functionals, and the characteristic functions of the parameters  $\mathbf{W}$ :

$$\mathcal{Z}^{(q+1)}(\chi^{(q)}) = \int_{\mathbf{W}} P_{\mathbf{W}}^{(q)}(w_{\alpha}(\rho_u)) \exp \left[ - \sum_{\alpha,u} \chi_{\alpha,u}^{(q)} w_{\alpha}(\rho_u) \right] \prod_{\alpha,u} dw_{\alpha}(\rho_u), \quad [10]$$

that is, the partition function at the  $(q + 1)$ th measurement is the Laplace (or Fourier) transform of the estimate of the probability density of the parameters at the previous measurement. The characteristic function of  $P_{\mathbf{W}}^{(q)}$  can be defined also as a Laplace (or Fourier) transform and can be easily computed in terms of the partition function. We obtain

$$\mathcal{Z}^{(q+1)}(\kappa) = \mathcal{Z}^{(q+1)}(\chi^{(q)} + \kappa) / \mathcal{Z}^{(q+1)}(\chi^{(q)}) = \mathcal{Z}^{(0)} \left( \sum_{\mu=0}^q \chi^{(\mu)} + \kappa \right) / \mathcal{Z}^{(0)} \left( \sum_{\mu=0}^q \chi^{(\mu)} \right). \quad [11]$$

where  $\mathcal{Z}^{(q+1)}(\kappa) = \langle \exp[-\sum_{\alpha,u} \kappa_{\alpha,u} w_{\alpha}(\rho_u)] \rangle^{(q+1)}$  is the characteristic function of the probability density  $P_{\mathbf{W}}^{(q+1)}(w_{\alpha}(\rho_u))$  and  $\kappa_{\alpha,u}$  are Laplace variables conjugated to  $w_{\alpha}(\rho_u)$ . From the characteristic functions of the probability densities of the parameters  $\mathbf{W}$ , we can evaluate various statistical properties, such as moments of different types, cumulants, linked averages, etc. Through inverse Laplace transformation, it is possible to evaluate the probability densities  $P_{\mathbf{W}}^{(q)}$  for different measurements. The cumulants of the parameters  $\mathbf{W}$  can be easily computed, by expanding  $\ln \mathcal{Z}^{(q+1)}(\kappa^{(q)})$  in a Taylor series and evaluating the coefficients of different powers. The shifts in the Laplace variables in Eqs. 11 lead to hierarchical relations typical in nonequilibrium statistical mechanics; the cumulants of a given order at a given measurement depend on cumulants of all orders, attached to a previous measurement. A simple case is that of initial Gaussian distributions for which Eqs. 11 show that, in the linear phase approximation, an initial Gaussian distribution leads to Gaussian distributions for all other steps; moreover, the cumulants of the first order (average values) vary from step to step and the variations are proportional to the elements of the initial covariance matrix

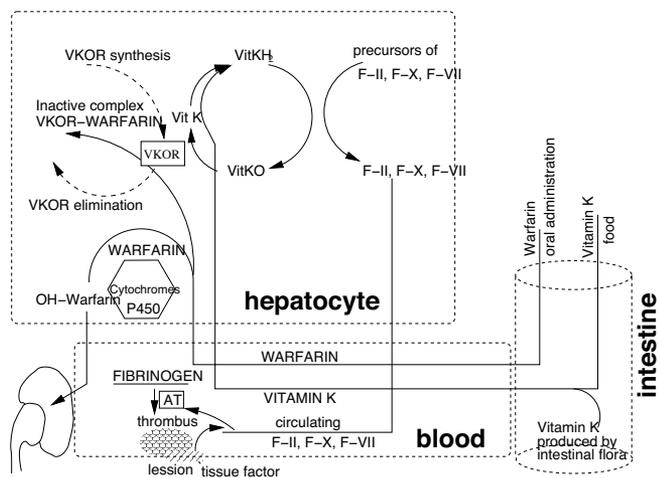
$$\langle \langle w_{\alpha_1}(\rho_{u_1}) \rangle \rangle^{(q+1)} = \langle \langle w_{\alpha_1}(\rho_{u_1}) \rangle \rangle^{(0)} + \sum_{\alpha_2,u_2} \langle \langle w_{\alpha_1}(\rho_{u_1}) w_{\alpha_2}(\rho_{u_2}) \rangle \rangle^{(0)} \sum_{\mu=0}^q \chi_{\alpha_2,u_2}^{(\mu)}. \quad [12]$$

The cumulants of the second order are the same for all measurements,  $\langle \langle w_{\alpha_1}(\rho_{u_1}) w_{\alpha_2}(\rho_{u_2}) \rangle \rangle^{(q)} = \langle \langle w_{\alpha_1}(\rho_{u_1}) w_{\alpha_2}(\rho_{u_2}) \rangle \rangle^{(0)}$ . In general, Eqs. 11 cannot be extended beyond the linear phase approximation; if the initial probability density of the parameters is Gaussian and the phase factors for the different measurements are quadratic in the parameters, then the probability densities  $P_{\mathbf{W}}^{(q)}$  can be evaluated by direct computations:  $P_{\mathbf{W}}^{(q)}$  are also Gaussian for any measurement but in this case both the averages and the elements of the covariance matrix vary from measurement to measurement.

### Application to Population Pharmacology of Oral Anticoagulants

Oral anticoagulants are widely used (12). However, their narrow therapeutic range and the variability of an individual's response to anticoagulant treatment due to genetic and nongenetic factors often result in adverse thromboembolic and bleeding events (12, 14). A great deal of information is needed for optimal prediction, which requires costly and advanced testing and patient cooperation and is impossible to gather in most cases. Our suggestion is to replace the partial knowledge about the patient by an unbiased estimation carried out by using our technique.

Blood coagulation involves chains of reactions in which proenzymes are activated to become reactive enzymes in the next reaction in a cascade, resulting in the transformation of fibrinogen to fibrin. The kinetics of the process has been recently described by a comprehensive model (15). The proenzymes and enzymes are coagulation factors, indicated by Roman numerals. A number of cofactors are involved, including calcium, phospholipids (constituents of the platelet membranes), Vitamin K<sub>1</sub>, which is a necessary cofactor for the hepatic carboxylation of glutamic acid residues in a number of pro- and anticoagulant proteins by  $\gamma$ -glutamyl carboxylase (GGCX: MIM 137167), thereby enabling calcium binding and attachment of the pro- and anticoagulants to phospholipids. During the carboxylation reac-



**Fig. 1.** Schematic representation of our pkpd model. We use two compartments: the digestive tract and the internal compartment. Administration of warfarin with saturated, rapid absorption, is modeled as a train of rectangular impulses (see *SI Appendix* for details).

tion, vitamin K<sub>1</sub> is oxidized to vitamin K epoxyde and must be regenerated by the vitamin K epoxyde reductase (EPHX1; MIM 132810 - VKOR) for carboxylation to continue. A subunit of EPHX1, named VKORC1 (MIM 608547) (16), is the target of anticoagulant drugs, such as warfarin. Anticoagulant drugs impair the function of vitamin K epoxyde reductase complex reducing the efficiency of the gamma carboxylation process. Vitamin K<sub>1</sub> is further reduced to a hydrogenated form, Vitamin K<sub>1</sub>H<sub>2</sub>, by another enzyme, Vitamin K<sub>1</sub> reductase, which is not sensitive to warfarin.

Optimization of anticoagulant treatment aims at reducing adverse reactions that result from random causes related to the pathology of the case and the variability of the individual response to the anticoagulant. The risk is determined as the proportion of individuals that develop thromboembolism or bleeding at a given activity of the anticoagulant drug. The standard measure is the “International Normalized Ratio” (INR) (17), which is the ratio of a patient’s prothrombin time, i.e., the time a blood sample takes to clot upon addition of thromboplastin, to a control sample, raised to the power of the International Sensitivity Index (ISI) value, which normalizes for the variation observed between various prothrombin time assays due to problems with the purity of the thromboplastin concentrate. The causes of variability in response to oral anticoagulants are quite well understood. Nevertheless, integration of this information with INR measurements to achieve the desired INR range fails frequently (12). Because the INR is measured invasively, it is determined at intervals of days or weeks. Consequently, it is necessary to predict risk using a minimum of information on patient response.

For clinical implementation, we have to take into account not only biochemical but also pharmacokinetic, genetic, and physiological variables. A probabilistic description increases the complexity of computations exponentially, and thus we tried to simplify the biochemical part of the model. After numerous trials, we came up with a relatively simple model of the populational pharmacokinetics and pharmacodynamics (pkpd), which incorporates a simplified description of the biochemistry and pharmacokinetics of coagulation as well as some physiological and populational aspects. Despite its simplicity, this model manages to give a satisfactory description of experimental and clinical data reported in the literature. A representation of the model is given in Fig. 1. The evolution equations, the literature

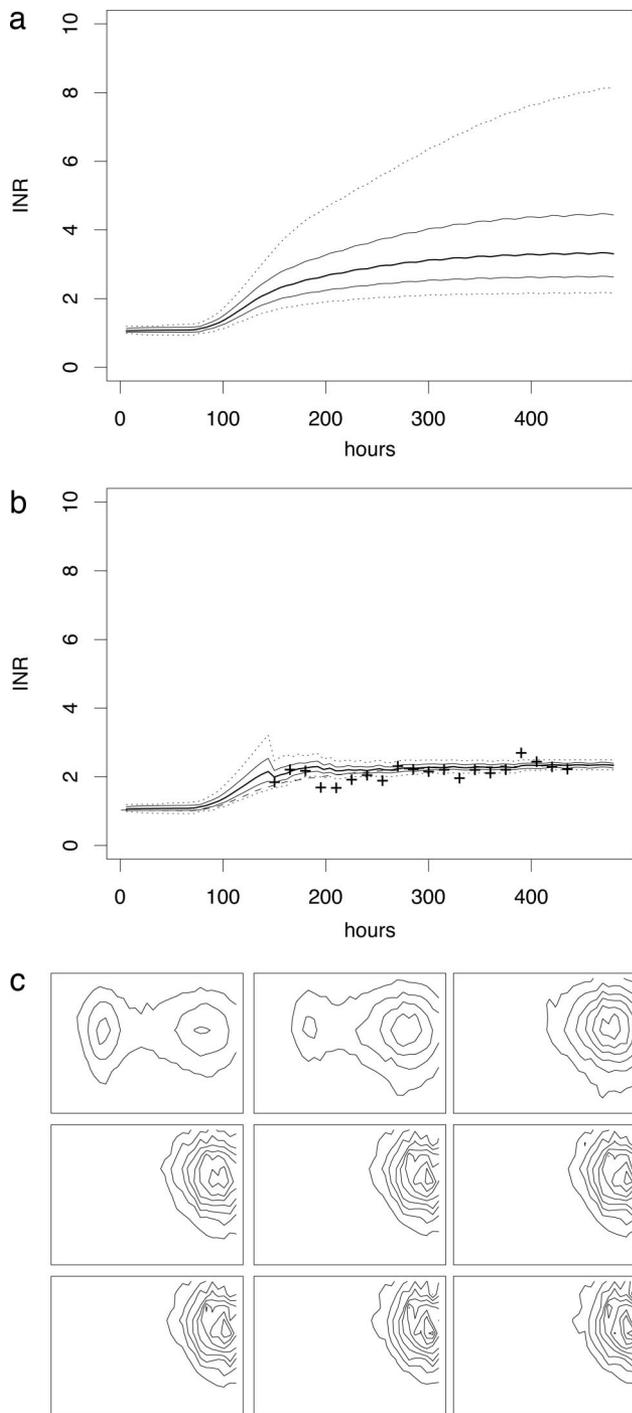
references, as well as the computer code used are given in [supporting information \(SI\) Appendix](#) and *SI Code*.

The relevant variables form a random vector whose evolution is modeled by a system of ordinary differential equations with random parameters **W** that describe the kinetics of the biochemical reactions influenced by warfarin, supplemented by genetic expressions for genotypic frequencies derived from the Hardy–Weinberg equations, as well as other quantitative relations among anthropometric parameters (see the online material). Warfarin is a racemic mixture of two enantiomers, R and S warfarin, usually administered once a day. It is inactivated through hydroxylation in the liver by cytochrome P450 enzymes. S-warfarin, the more active enantiomer, is hydroxylated primarily by the CYP450 isoform CYP2C9 (MIM 601130), which blocks irreversibly (13) the microsomal enzyme vitamin-K epoxyde reductase (VKORC1) that reduces Vitamin-K epoxyde. Vitamin K is oxidized during  $\gamma$ -carboxylation of glutamate residues of a number of proteins, rendering them biologically active. From the coagulation factors that are released into the blood stream, we take explicitly into account only two (factors II and VII), because their titers directly influence the speed of clot formation (17). Warfarin, in therapeutic dosages, inhibits reduction of vitamin-K epoxyde sufficiently to lead to an inhibition of the  $\gamma$ -carboxylation reaction that inhibits the rate of active factor release. Coagulation factors are eliminated from circulation with variable rates (14) and their titer is reduced, thus prolonging coagulation time.

Genetic sources of interindividual variability to warfarin treatment are polymorphisms in CYP2C9 (MIM 601130) and in VKORC1 and in the coagulation factors VII and II (16, 18). We assume that the alleles corresponding to these loci are at genetic equilibrium and thus the phenotypic frequencies are evaluated in the model by using the Hardy–Weinberg equations. Other factors are the daily absorption of vitamin K from food and intestinal bacterial sources, overall vitamin K reserves, individual blood volume, and anthropometric variables (height, weight, body mass index, and distribution volume). The relations between these anthropometric quantities are described by empirical relations (see *SI Appendix* and references therein). The INR is evaluated from the titers of the coagulation factors II, VII by using a simplified allometric law introduced by Watala (17).

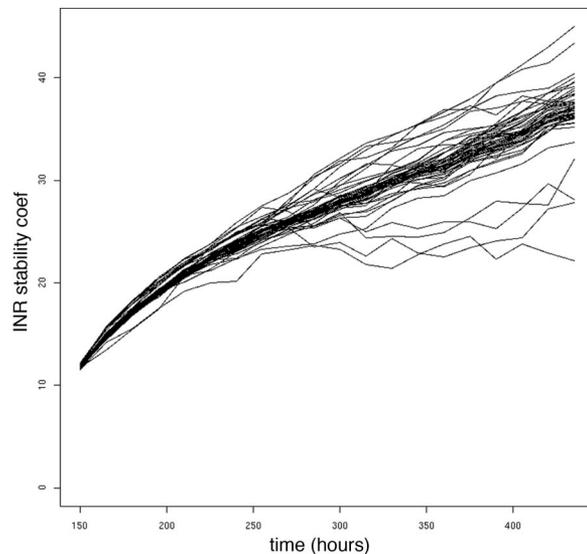
The random parameters **W** are: gender-dependent height, body mass index, CYP activity, VKORC1 activity, and baseline factor synthesis levels for coagulation factors II and VII. The anthropometric variables are subjected to intrinsic random variations. The kinetic parameters are primarily determined by the genotypes of the individuals; because, in many cases, these genotypes are not known, they produce apparent random variations. In the literature, there is empirical data that relates the required dosages of anticoagulants necessary for reaching a given INR value for different types of individuals classified by gender, genotype and anthropometric parameters. This type of clinical data can be input in our model for simulating “standard” (reference) cases, which correspond to given sets of parameters. Processing of the clinical data gives a probability density  $p(\mathbf{W})$ , and values of INR, which are random functions of time, that represent the expected behavior of the population. Fig. 2a represents the expected populational effect for 10<sup>5</sup> simulated patients, the central line is the median of the instantaneous INR, the thin continuous line enclose 0.66 of the area under the probability density of the instantaneous INR, whereas the dotted lines enclose 0.99 of the area.

The representation of the clinical data in the form given in Fig. 2a is needed for testing the accuracy of our method. The results of these computations are stored in a database, from which we randomly select different cases; for a given case, “measured” INR values are generated for successive times by adding Gaussian noise to the database values; these simulated “measure-



**Fig. 2.** Simulation results. (a) Expected populational effect of a drug regimen on the INR in time.  $10^5$  virtual patients are generated. A fixed 5 mg/day regimen is administered to each simulated case and the INR is calculated every hour for every case. The thick line represents the median INR value as a function of time. The thin continuous lines contain 0.66 of the probability distribution. The dotted lines contain 0.99 of the distribution. (b) Effect of INR measurements on probability distribution of predicted INR. One of the INR traces (dashed line, close to the median line) is selected as a reference and the INR measurements are simulated at equal intervals in time (crosses). The same conventions are used as in a. (c) Successive estimates of the marginal probability densities of two  $\mathbf{W}$  parameters: CYP2C9 (on x axis) and VKORC1 (on y axis).

ments” are represented by crosses in Fig. 2b. Each cross represents a consultation and the corresponding INR value is the additional incremental information, which should be processed



**Fig. 3.** The stability coefficient of the estimated values of the INR for 50 sets of simulations, each involving  $10^5$  virtual patients, as a function of time.

in order to make clinical predictions. We used the Bayesian version of the method presented above, where the vector of observables  $\mathbf{C}$  has one component, the INR. At the first checkup, we know nothing about the parameters  $\mathbf{W}$ . We start from an initial *a priori* density  $p_0(\mathbf{W})$ , from which, by applying Eqs. 7 and 8 repeatedly, we get improved estimates of the statistical properties of  $\mathbf{W}$ , represented by successive probability densities  $p_1(\mathbf{W}), \dots, p_q(\mathbf{W}), \dots$ . These probability densities are used for computing successive estimates for the probability densities of the INR. In addition to the random values of the INR, represented by crosses, in Fig. 2b we show the median values of the incremental evaluations of the INR and the corresponding envelopes for 0.66 and 0.99 confidence intervals, respectively. We did a large number of simulations and noticed that the method has remarkable accuracy. See *SI Appendix*, which contains 50 different realizations of Fig. 2 that display the results of 50 different sets of computations, each involving  $10^5$  simulated patients; all these figures illustrate the accuracy of the method. Fig. 2c shows a two-dimensional projection of the successive estimates of the probability densities  $p_1(\mathbf{W}), \dots, p_q(\mathbf{W}), \dots$ . In most cases, these probability densities become tighter and tighter. Irrespective of the behavior of  $p_q(\mathbf{W})$ , the accuracy of reproducing the values of the INR is excellent.

For assessing the efficiency of our method, we evaluate the relative contribution of the random fluctuations to the estimated values. For each of the 50 different sets of computations, each involving  $10^5$  simulated patients, presented in *SI Appendix*, we compute the stability  $\vartheta_{\text{INR}}(t) = \langle \text{INR}(t) \rangle / [\langle \text{INR}^2(t) \rangle]^{1/2}$  of the evaluated INR with respect to fluctuations; that is, the ratio between the average value of the estimated INR,  $\langle \text{INR}(t) \rangle = \langle \text{INR}(t) \rangle$ , and the square root of the dispersion,  $\langle \text{INR}(t) - \langle \text{INR}(t) \rangle \rangle^2 = \langle \text{INR}^2(t) \rangle$ . Fig. 3 shows that, as time increases, the stability function  $\vartheta_{\text{INR}}(t)$  has a tendency of increasing; that is, as more and more information from successive medical checkups is implemented in the computations, the relative fluctuations of the estimated INR are becoming smaller and smaller.

These simulations suggest that our approach provides a reliable method for incremental prediction of future behavior of the INR from incomplete clinical data. The method will fail in the case of unexpected random events, such as the accidental intake of food containing unusually large amounts of vitamin K; for such unusually rare events, special stochastic methods must be used. Nevertheless our approach is a step in the right direction.

Ansell *et al.* (12) summarizes the biochemical and genetic aspects of the variability of anticoagulant response. Hamberg *et al.* (19) describe an alternative populational approach to warfarin pharmacology that emphasizes the role played by the age of the patients. The protein synthesis function of the liver decreases with age, resulting in a change of the therapeutic effect of anticoagulants. Our focus is on building a minimal model, which, in addition to the most relevant biochemical processes, takes into account other relevant information and makes it possible to make simple clinical predictions based on incomplete information.

### Broader Implications

A referee pointed out interesting similarities between our approach and latent variable models (LVM) as used, for example, in behavioral sciences or econometrics (20–22) and on data assimilation techniques (DA) applied in atmospheric science and signal processing (23–25). Strictly speaking, there is no overlapping between our approach and these two techniques. LVM and DA techniques are based on traditional methods of mathematical statistics, whereas our approach is based on methods from

theoretical physics and chemistry. We choose a physical approach because: (i) it makes it possible to integrate in a simple way the results of the estimation into physico-chemical models; (ii) it opens the way for implementing simplified semi-analytic approaches from statistical field theories of theoretical physics, resulting in significantly shorter computation times. Neither of these features is available for LVM and DA. Despite these differences, some analogies do exist and their further investigation is strongly recommended. We are especially interested in combining our method with the filtering techniques used in DA. We also intend to use some features from LVM for analyzing biochemical networks; in this context, the implementation of diagrammatic methods similar to the one used for the applications to behavioral sciences of LVM seems to be promising.

**ACKNOWLEDGMENTS.** We thank Profs. Peter Wolynes, Robert Silbey, and an anonymous referee for their useful suggestions. This project was supported in part by the National Science Foundation, BayGene, the Alexander von Humboldt Foundation, CEEEX Grant M1-C2-3004/2006-Response and ANCS Nr.2-CEX0611-18/2006-Biomat, and VIASAN 451/2004 of the Romanian Ministry of Research and Education and Grant BFU2006-01951-BMC of the Ministry of Education and Science of Spain.

1. Kay SM (1993) *Fundamentals of Statistical Signal Processing Estimation Theory* (Prentice-Hall, Englewood Cliffs, NJ), Vol 1.
2. Jaynes ET (2003) *Probability Theory: The Logic of Science*, ed Bretthorst GL (Cambridge Univ Press, Cambridge, UK).
3. Zubarev DN (1974) *Nonequilibrium Statistical Thermodynamics* (trans from Russian) (Consultants Bureau, New York).
4. Vlad MO, Mackey MC (1995) Maximum information entropy approach to non-markovian random jump processes with long memory: Application to surprisal analysis in molecular dynamics. *Physica A* 215:339–360.
5. Vlad MO, Cerofolini GF, Oefner P, Ross J (2005) Random activation energy model and disordered kinetics, from static to dynamic disorder. *J Phys Chem B* 109:21241–21257.
6. Press SJ (2003) *Subjective and Objective Bayesian Statistics* (Wiley, New York), 2nd Ed.
7. Fisher RA (1999) in *The Genetical Theory of Natural Selection* (Oxford Univ Press, Oxford), pp 22–47.
8. Kimura M (1957) On the change of population fitness by natural selection. *Heredity* 12:145–167.
9. Vlad MO, *et al.* (2005) Fisher's theorems for multivariable, time- and space-dependent systems, with applications in population genetics and chemical kinetics. *Proc Natl Acad Sci USA* 102:9848–9853.
10. Parisi G (1998) *Statistical Field Theory* (Perseus, New York).
11. Zinn Justin J (2002) *Quantum Field Theory and Critical Phenomena* (Oxford Univ Press, New York), 4th Ed.
12. Ansell J, *et al.* (2004) The pharmacology and management of the vitamin K antagonists. *Chest* 126:2045–2335.
13. Fasco MJ, Principe LM (1982) R- and S-Warfarin inhibition of vitamin K, vitamin K 2,3-epoxide reductase activities in the rat. *J Biol Chem* 257:4894–4901.
14. Hardmann JG, E Limbird L, eds (2002) *The Pharmacological Basis of Therapeutics* (McGraw-Hill, New York).
15. Luan D, Zai M, Varner JD (2007) Computationally derived points of fragility of a human cascade are consistent with current therapeutic strategies. *PLoS Comput Biol* 3:1347–1359.
16. Rieder MJ, *et al.* (2005) Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N Engl J Med* 352:2285–2293.
17. Watala C, Golanski J, Kardas P (2003) Multivariate relationships between international normalized ratio and vitamin K-dependent coagulation-derived parameters in normal healthy donors and oral anticoagulant therapy patients. *Thromb J* 7:1–10.
18. Quek SC, Low PS, Saha N, Heng CK (2006) The effects of three factor VII polymorphisms on factor VII coagulant levels in healthy Singaporean Chinese, Malay and Indian newborns. *Ann Hum Genet* 70:951–957.
19. Hamberg AK, *et al.* (2007) A PK-PD model for predicting the impact of age, CYP2C9, and VKORC1 genotype on individualization of warfarin therapy. *Clin Pharmacol Ther* 81:529–538.
20. Loehlin JC (2004) *Latent Variable Models: An Introduction to Factor, Path and Structural Equation Analysis* (Lawrence Erlbaum, Mahwah, NJ), 4th Ed.
21. Hancock GR, Smuelsen KM, eds (2007) *Advances in Latent Variable Mixture Models* (Information Age, Charlotte, NC).
22. Heckman JJ, Vytlacil EJ (1999) Local instrumental variable models for identifying and bounding treatment effects. *Proc Natl Acad Sci USA* 96:4730–4734.
23. Evensen G, (2007) *Data Assimilation: The Ensemble Kalman Filter* (Springer, Berlin).
24. Kalnay E (2003) *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge Univ Press, Cambridge, UK).
25. Wang B, Zou X, Zhu J (2000) Data assimilation and its applications. *Proc Natl Acad Sci USA* 97:11143–11144.