A Kinetic Neural Network Approach for Absolute Quantification and Change Detection in Positron Emission Tomography

Davide Poggiali
Dipartimento di Matematica "Tullio Levi-Civita",
Via Trieste 63, 35121 Padova, Italy

Padova Neuroscience Center
Via Giuseppe Orus 2b, Padova, Italy
davide.poggiali@unipd.it

Diego Cecchin
Dipartimento di Medicina, Unità di Medicina Nucleare,
Via N. Giustiniani 2, 35128 Padova, Italy

Padova Neuroscience Center
Via Giuseppe Orus 2b, Padova, Italy
diego.cecchin@unipd.it

Stefano De Marchi
Dipartimento di Matematica "Tullio Levi-Civita",
Via Trieste 63, 35121 Padova, Italy
stefano.demarchi@unipd.it

Abstract

When dealing with absolute quantification of the tracer kinetics in Positron Emission Tomography (PET), one of the most reliable and widely used family of methods are the physiologically-consistent compartmental Ordinary Differential Equation (ODE) models introduced in the 70s mainly by Sokolov and Patlak.

In this work we introduce the kinetic Neural Network (kNN), a novel ODE-based technique for absolute quantification inspired by compartmental models and Neural Networks commonly used in Deep Learning.

We applied this method on the PET Grand Challenge 2018 consisting in simulated PET images of the same subjects before and after an intervention. At last, we compared the derived parameter images by using Hostelling's t-squared statistic to evaluate the regions of the brain affected by the treatment.

Keywords: Kinetic Models, Positron Emission Tomography, Ordinary Differential Equations, Neural Networks.

1. Introduction to Tracer Kinetics Models

Positron Emission Tomography (PET) is a nuclear medical imaging technique that produces an image of metabolic processes in the body. The PET scanner detects pairs of gamma rays originated from the annihilation of a positron emitting radioactive element (as the 18-Fluore-deoxy-glucose or ^{18}FdG), injected into the body.

One of the peculiarities of PET stands in the possibility to reconstruct a list mode acquisition as a time series of images. This allows the researchers to study the evolution of the tracer kinetics [8, 4]. A **tracer** is a small quantity of substance used for the purpose of traking the kinetics of a traced molecule For instance ^{18}FdG is a tracer for glucose. In fact the tissues cannot distinguish such molecule from the glucose and hence it follows the same kinetics of the traced.

1.1. What are compartment models

A compartment model [1, 3, 5] is a type of mathematical model used for describing how materials or energies are transmitted among the compartments of a system. Each compartment is assumed to be a homogeneous entity. Hence a compartment represents, in medical application, a tissue (or homogeneous group of tissues) exchanging material at the same speed with another.
The kinetics of the material is described by an Ordinary Differential Equations (ODE) system with as unknown functions the concentration of the considered material in each compartment.

The Sokolov model [7] (1970s) is the main model studying the kinetics of the tracer ^{18}FdG in the brain. It uses a sequence of PET reconstructed images as input.

The model, represented in Fig. 1, is built under the hypothesis that the tracer, injected in the plasma C_p is excanged with the brain tissue C_f, and part of it, called C_b undergoes an irreversible chemichal bound. Defined the volume fraction of the compartment C_p as V_b, it's easy to compute the joint volume fraction of C_f and C_b as $1 - V_b$. Since it is impossible from a PET

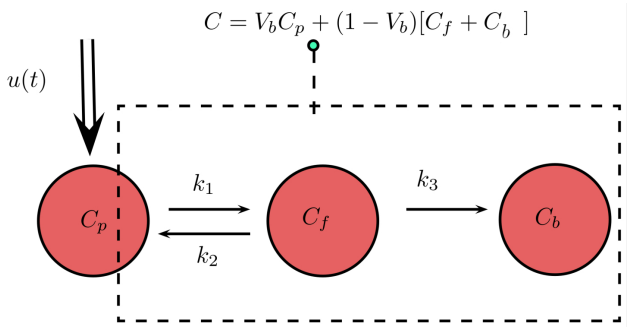

Figure 1: Grafical representation of the Sokolov Model.

image to distinguish the compatments, the PET-derived measured signal C at each time will be a weigthed sum between the tracer concentration of all the compartments, with the volume fractions as weights.
Under these assumptions we can write this model as the following ODE system:

$$\begin{cases} \dot{C_p} = & u(t) & -k_1 C_p & +k_2 C_f & \\ \dot{C_f} = & & k_1 C_p & -k_2 C_f & -k_3 C_f \\ \dot{C_b} = & & & & k_3 C_f \\ C(t) = & & V_b C_p(t) + & (1 - V_b) & (C_f(t) + C_b(t)) \end{cases} \tag{1}$$

An ODE model describing the kinetics of the tracer can be written in a more general form. Under the assumptions that the tracer:

1. can be measured independently from the traced
2. follows the dynamics of the traced and do not perturb the state of the system (small quantity)
3. is kinetically indistinguishable from the traced

the model goes under the category of Linear Time-Invariant model in state-space form (LTI state-space) and can be witten as follows

$$\begin{cases} \dot{x} &= Ax + Bu \quad , \quad x(0) = x_0 \\ y(t) &= Cx(t) \end{cases} \tag{2}$$

where u is the input, y the neasured output and the system is determined by (A, B, C).

For instance, the Sokolov model decribed in (1) can be rewritten as follows:

$$\begin{gathered} A = \begin{pmatrix} -k_1 & k_2 & 0 \\ k_1 & -(k_2+k_3) & 0 \\ 0 & k_3 & 0 \end{pmatrix}, \\ B = (1,0,0), \\ C = (V_b, 1-V_b, 1-V_b) \end{gathered} \tag{3}$$

with $x = [C_p, C_f, C_b]$, and (with a small notation abuse) $y(t) = C(t)$.

1.2. Analytical solution of compartment models

The system (A, B, C) is equivalent to the direct input-output relation

$$y(t) = Cx(t) = (C\ \exp(tA)\ B) * u(t)\,, \tag{4}$$

where $*$ is the convolution operator. This direct formula comes from the general solution of the ODE

$$\dot{x} = Ax + Bu$$

which is given by

$$x(t) = \exp(At) \int_0^t \exp(-A\tau) B u(\tau) d\tau =$$

$$= \int_0^t \exp(A(t-\tau)) B u(\tau)\ d\tau = (\exp(At)B) * u(t).$$

In fact, since the vector C is constant, it follows

$$y(t) = Cx(t) = (C \exp(At)B) * u(t) = \phi(t) * u(t)$$

where $\phi(t) = C \exp(At)B$ is called **impulse response function** (IRF). The IRF is the solution in case $u(t)$ is the unitary pulse, i.e. the Dirac function $\delta(t)$. It will be shown in Section 2.2 that the explicit solution can be used for estimating the parameters contained in (A, B, C) given the input-output signals $(u(t), y(t))$.

2. A Kinetic Neural Network approach

If we observe the representation of the model in Fig. 1, we understand that in priciple[1] any oriented graph with a knot as input and another as output can be eligible for a kinetic model.
In this work we use the graph stucture of a Neural Network as a kinetical model

[1] given for granted some assuptions on model identifiability that we don't discuss in this report for the sake of conciseness.

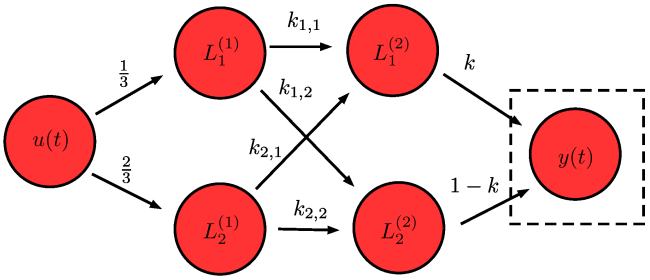

Figure 2: Grafical representation of the simple 2x2 kNN model.

and we call it **Kinetic Neural Network** (kNN). Amongst all possible kNN models, we choose the simple 2x2 model whose representation can be seen in Fig. 2, described by the following model:

$$\begin{cases} \dot{L}_1^{(1)} = \frac{1}{3}u(t) - k_{1,1}L_1^{(1)} - k_{1,2}L_1^{(1)} \\ \dot{L}_2^{(1)} = \frac{2}{3}u(t) - k_{2,1}L_2^{(1)} - k_{2,2}L_2^{(1)} \\ \dot{L}_1^{(2)} = k_{1,1}L_1^{(1)} + k_{2,1}L_2^{(1)} \\ \dot{L}_2^{(2)} = k_{1,2}L_1^{(1)} + k_{2,2}L_2^{(1)} \\ C = kL_1^{(2)} + (1-k)L_1^{(2)} \end{cases} \tag{5}$$

It has to be noted that $1/3$ and $2/3$ are chosen arbitrarily to guarantee reachability and identifiability, and that all the parameters are supposed to lie in the interval $[0, 1]$.

The model (5) can be rewritten in LTI form (2) with (A, B, C):

$$\begin{aligned} A &= \begin{pmatrix} -(k_{1,1}+k_{1,2}) & 0 & 0 & 0 \\ 0 & -(k_{2,1}+k_{2,2}) & 0 & 0 \\ k_{1,1} & k_{2,1} & 0 & 0 \\ k_{1,2} & k_{2,2} & 0 & 0 \end{pmatrix}, \\ B &= (1/3, 2/3, 0, 0), \\ C &= (0, 0, k, 1-k). \end{aligned} \tag{6}$$

Using this model the physiological consistence is lost in favor of a number of (voxel-by-voxel) parameters representing the tracer kinetics in the given model. Such parameters can be used for comparing PET images of different groups of subjects or the same subject before and after a medical treatment.

2.1. Computing ϕ

To get the ODE model's explicit solution we decided to compute $\phi = C\ \exp(tA)\ B$ symbolically with the Python package SymPy [6], by using the Padè approximation [2] of the exponential matrix $\exp(tA)$.
The Padè approximation of a squared, real-valued matrix M is obtained by computing the matrices

$$N_{p,q}(M) = \sum_{j=0}^{p} \frac{(p+q-j)!\ p!}{(p+q)!\ j!\ (p-j)!} M^j$$

$$D_{p,q}(M) = \sum_{j=0}^{q} \frac{(p+q-j)!\ q!}{(p+q)!\ j!\ (q-j)!} (-M)^j$$

with $p, q \in \mathbb{N}$ properly chosen. The approximant is hence given by

$$\exp(M) \approx F_{p,q}(M) = N_{p,q}(M)^{-1} D_{p,q}(M).$$

From literature [2] it is known that if the matrix has norm $||M|| < 5.4$ then $||\exp(M) - F_{13,13}(M)|| < 10^{-16}$ and hence $p = q = 13$ is a reasonable choice.

2.2. Identification of the model

The model is evaluated on at Volume Of Interest (VOI) level or on a voxel basis with PET data and fitted numerically using the Gradient Descent technique.
From last Section we explained how to obtain an approximate value for $\phi(t)$ as a symbolic function of the model parameters set $\theta = \{\theta_1, \ldots, \theta_n\}$. From that values we can obtain an approximation of the expected output

$$y_\theta = \phi_\theta * u$$

that we have to compare to the real measured output y. Namely, we want to find the paramater set θ tha minimizes the functional

$$J(\theta) = \frac{1}{2n_t} \sum_{i=1}^{n_t} (y_{\theta,i} - y_i)^2$$

with n_t the numer of timepoints. As optimization algorithm, we choose the Gradient Descent technique

$$\theta \leftarrow \theta - \alpha\ \nabla_\theta J$$

starting from a randomly chosen set θ_0 of parameters in $[0, 1]$. To get $\nabla_\theta J$ we calculate

$$\frac{\partial J}{\partial \theta_j} = \frac{1}{n_t} \sum_{i=1}^{n_t} (y_{\theta,i} - y_i) \frac{\partial y_{\theta,i}}{\partial \theta_j} = \frac{1}{n_t} \sum_{i=1}^{n_t} (y_{\theta,i} - y_i) \frac{\partial \phi_{\theta,i}}{\partial \theta_j} * u(t)$$

that can be obtained numerically with realtive ease, given our knownlegde of ϕ_θ at a symbolic level. The symbolic variables are at last evaluated numerically using numpy in Python, in order to obtain the "optimal" set of parameters θ for each level (VOI or voxel).

2.3. Statistics

Provided that the set of parameter obtained in the previous Section is at voxel level, we have to compare statistically two images "before" and "after" the intervention of n parameters. To perform this we decided to use the Hotelling test.
Given two matrices X and Y containing repectively n_X and n_Y observations of m variables and their covariance matrices Σ_X and Σ_Y we compute the matrix

$$\Sigma = \frac{(n_X - 1)\Sigma_X + (n_Y - 1)\Sigma_Y}{n_X + n_Y - 2}$$

that can be used to get the T^2 (scalar) value, defined as

$$T^2 = \frac{n_X n_Y}{n_X + n_Y}(\bar{X} - \bar{Y})\Sigma^{-1}(\bar{X} - \bar{Y})$$

with $\bar{X}$ and $\bar{Y}$ to indicate the mean across the first axis. The T^2 values are related to the F-statistic as follows:

$$F = \frac{n_X + n_Y - m - 1}{(n_X + n_Y - 2)m}T^2$$

and from the F-value we get the probability of the null hypothesis (X and Y belong to the same normal distribution) by computing the Cumulative Distribution Function of F using the apposite scipy function in Python.

3. Application: the PET Grand Challenge

PET Grand Challenge 2018[2] has been a competition proposed by King's College (UK) and Imperial College (UK) researchers aimed to compare different techinques in the study of dynamic PET processing.
A set of 10 simulated PET images was given, 5 after 5 before an 'intervention'. The goal was to identify the areas that underwent a change during the intervention, and the magnitude of such change per area.
Every simulated scan lasted 90 minutes after the tracer injection, all the images were alingned to a template (then no motion correction was needed) and the nature of the tracer was not known (so there was not a preferable model). Since the PET were simulated, the ground truth was known to the organizer, in order to declare a winner.

The simulated data provided for the challenge was chosen to test the method proposed before.
Since the matrix A from (6) was made of parameters in $[0, 1]$ and the acquisition time was less or equal than 1.5 hours, we conclude that

$$||At||_\infty \leq 4.5 \;\; \forall t \leq 1.5$$

[2] http://www.petgrandchallenge.com/

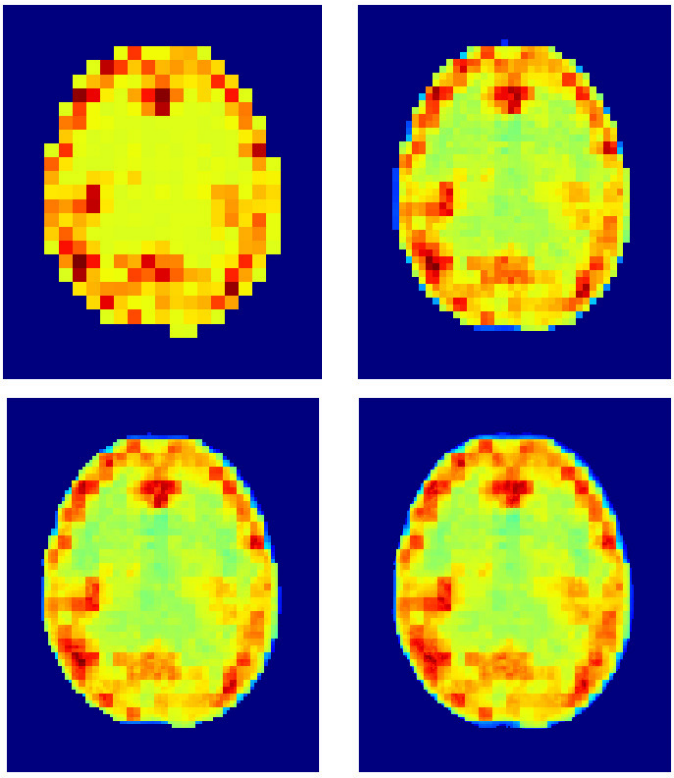

Figure 3: Progressive iterative scheme: the same silces of the $k_{1,1}$ parameter image at level: 8 (upper left), 4 (upper right), 2 (down left) and full resolution image at level 1 (down right).

thus assuring the accuracy of the Padè approximation of the exponential Matrix $\exp At$ as explained in Section .

The image size was 182x218x182x23, which means 3D images of 182x218x182 voxels at 23 different times. The challenge required a voxelwise fitting of the model, hence about $7 \cdot 10^6$ independent optimization Gradient Descents had to be performed. Since this is potentially time-consuming, we adopted two strategies:

1. A progressive refining of the parameters, by undersampling the images of a factor 8, 4, 2 and then 1 (the original image size) per axis. The solution at lower level was resampled to the upper level and used as an initial guess. See Fig. 3.

2. The computation for each voxel stopped once the voxel residual was lower than the relative tolerance, so less voxels are left to fit at each iteration.

3. The tolerance increases of order with the level so that finer granularity requires less computation time.

After the model fitting, all the parameters were normalized by the mean cerebellar value, since Cerebellum is usually considered as a stable area of tracer uptake. Normalized images were given as voxelwise input to the Hotelling test, which resulted in a 3D 182x218x182 p-value image. Voxels of p-value less than the significance level 0.05 were saved in a binary mask, containing the voxels were the parameters were found to change significantly across the 5 simulated subjects.

At present, no result are present in the webpage, hence it is not possible for now to evaluate the accuracy of the resulting mask. A consistency analysis of the parameters between different subjects has been performed, revealing parameter's standard deviations between 0.01 and 0.02, and variation coefficients are between 15% and 30%. Since the parameters values have order of magnitude of 10^{-2}, we can affirm that the estimated paramaters lie in an acceptably short range and are coherently estimated between the different subjects.

4. Conclusions and future work

In this work we proposed a new approach in the modelling of tracer kinetics in PET, along with an application of such method.
In the context of this work a general, graph-based framework for PET parameter estimation has been written in Python, allowing in future studies VOI or voxel-level parameter estimation, regardless of the tracer injected.

In order to allow the usage of the framework in future studies it will be important to acheive its validation against traditionally used methods.

Other possible models and applications still has to be tested: for instance we can use the same framework in change detection in brain tumor PET, allowing to detect the areas of significant variation in the same subject acquired before and after neurosurgery. Another interesting task would be to test a larger network in the ODE model.

5. Acknoledgments

Work done within the Research Network RITA (Rete ITaliana di Approssimazione). The authors want to thank Emma Perracchione and Francesco Rinaldi from the Dept. of Mathematics of Padova for the precious they gave.

References

[1] D.H. Anderson, *Compartmental Modeling and Tracer Kinetics*, Springer Berlin Heidelberg, 1983.

[2] M. Arioli, B. Codenotti, and C. Fassino, *The Padè method for computing the matrix exponential*, Linear Algebra and its Applications, 240 (1996), pp. 111–130.

[3] R. E. Carson, P. Herscovitch, M. E. Daube-Witherspoon, *Quantitative Functional Brain Imaging with Positron Emission Tomography*, Academic Press, 1998

[4] C. Cobelli, E. Carson, *Introduction to modeling in physiology and medicine*, AP, 2008

[5] C. Cobelli, D. Foster, G. Toffolo, *Tracer kinetics in biomedical research*, Springer, 2001.

[6] A. Meurer, C. P. Smith, M. Paprocki, et al., *SymPy: symbolic computing in Python*, PeerJ Computer Science, 3 (2017), p. e103.

[7] L. Sokoloff, M. Reivich, C. Kennedy, et al. *The [14C] deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat.*, J Neurochem. 1977;28(5):897–916

[8] G. L. Zeng, A. Hernandez, D. J. Kadrmas, and G. T. Gullberg, *Kinetic parameter estimation using a closed-form expression via integration by parts*, Physics in medicine and biology, 57 (2012), pp. 5809–21.