Mathematical and Statistical modelling using the FMM approach

The case of the Electrocardiogram

Cristina Rueda Sabater

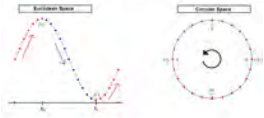
2021-11-09

Table of Contents

Table of Contents

Oscillatory Signal

A system in which a particle moves returning to its initial state after a certain period is an oscillatory system. A measure in that system is an oscillatory signal.



Single oscillatory signal or Circular signal

$\mu(t) = cos(\phi(t))$,
$0 \leq \phi(t) \leq \phi(t') \leq 2\pi, 0 < t \leq t' \leq 2\pi$

Oscillatory signal: one cycle

A signal in $(0, 2\pi]$ with multiple single oscillatory components.

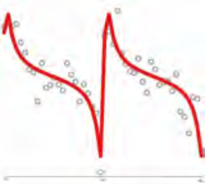
Oscillatory time series: several cycles

Is a sequence of oscillations, occurring during several cycles.

Biophysical signals usually do not oscillate as sinusoids.

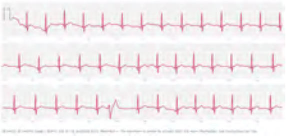
Examples of oscillatory biomedical signals

Oscillatory systems arise in different biological and medical fields.



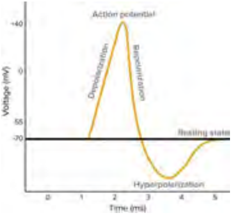
Gene expressions.

Circadian cycle/clock



Electrocardiogram (ECG).

Cardiac cycle.



Neuron voltage change.

Action Potential (AP) cycle.

Some of the questions and challenges

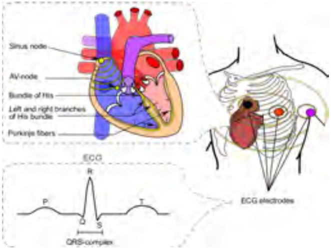
- What is the time of activation of a rhythmic gene?
- How is the morphology of the ECG signal for different pathologies?
- Can the morphology of the AP curves predict the genetic type of the neuron?

The electric function of the heart: the ECG

The electric signal travels:

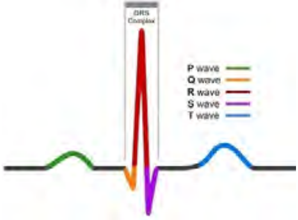
Sinus node $\rightarrow$ atria $\rightarrow$ ventricles

The ECG signal is recorded with the aid of electrodes as the projection of a vector onto a fixed axis.



The trace of a typical ECG signal forms different waves for a given cardiac cycle (heartbeat):

- P wave (atrial polarization).
- QRS complex: Q wave + R wave + S wave (ventricular depolarization).
- T wave (ventricular repolarization).



Successive recording of heartbeats in a time interval generates a continuous signal.

ECG Signal Analysis challenges and goals

The ECG signal is one of the most important tools for elucidate information about the heart diagnostic.

Challenges

- The morphology exhibits several non sinusoidal oscillations.
- The morphology changes from one cycle to another.
- The morphology changes across patients and conditions.
- The signal is contaminated with noise artefacts.

Goals

- To describe the signal with a physiological interpretable model.
- To extract parameters from the data that provide useful information for clinical decision-making.

Two widespread approaches to the analysis of ECG signals

Theoretical modelling

- Mathematical models are physiologically interpretable as describe the electric function of the heart.
- The models are too complex, most based on systems of differential equations, what make more difficult the identification of parameters. They fail to generate all the variety of real signals.

Data driven

- Data driven models have proven useful only for specific tasks as filtering, data compression or detection of the main waves.
- Deep and Machine learning approaches appear useful for the diagnosis of elected pathologies. They are highly dependent on the training dataset. The black-box nature of the solutions does not help the physicians beyond just the diagnosis. There is not any explanation or any clue about the damaged part of the heart and neither a way to verify the results.

most researchers agree to approach the analysis *beat-by-beat* .

The ECG analysis up-to-day

Is a relevant topic that has attracted the interest of researchers from different fields as mathematics, physics, bioengineering or medicine.

Despite all the research done, there is not an efficient mathematical formulation for ECG modelling, being one of the major problems in biomedical research.

There is a very large room for improvement related to automatic diagnosis.

The FMM: a theoretical model and a data-driven approach

- The mathematical formulation is simple enough to be easily parametrized and rich enough to provide realistic signals.
- The parameters are physiologically interpretable in a similar way to what a physician would look in the signals thus quite useful in diagnosis.

The FMM approach is a revolution in electrocardiography research, useful for many questions in the clinic practice.

Table of Contents

Signal Analysis : A topic to establish bridges

The analysis of signals have been addressed by researchers from very different disciplines with diverse objectives and a variety of applications.

- Physicists: Design dynamic models defined in terms of differential equations.
- Mathematicians: Develop theoretically sound decomposition approaches.
- Engineers: Develop physically interpretable decomposition approaches.
- Statisticians: Design data-driven models.

Data scientists design technologies and algorithms to extract information and discover patterns.

PROBLEM: The terminology, and even some important concepts, are different depending on the researcher's field of specialization.

Time space, Period and Phase

Time space

"t" moves in the unit circle instead of the real line.

In other case, transform time points in $[t_0, T + t_0]$ to $(0, 2\pi]$ using $\frac{(t-t_0)2\pi}{T}$.

Period

The interval of time between consecutive events (of the same type) in a cyclic system.

The period is often assumed known in biomedical oscillatory signals. It is a fixed number, such as "24 hours" for signals associated to the circadian clock or, stable under standard conditions, as the heart rate for the ECG signal.

Phase

Is a number in $(0, 2\pi]$ that indicates the position of the system within the cycle.

It is an elusive concept but more important than the period for oscillatory signals.

The complex and the Analytic signal

Let $\mu(t)$ the real signal under study and $S(t)$ the complex-valued signal associated to $\mu(t)$:

Complex Signal

$S(t) = \mu(t) + i\nu(t)$.

WHY a real signal may be represented by a complex signal?

- Often in application there are two signals in the system; not always both are visible.
- Differential equations describing a dynamic system have complex solutions.
- A complex signal carries information about two real signals that leads to *"an economical algebra and evocative geometry"* and gives new insights by using the amplitude and phase modulation: $S(t) = A(t)e^{i\phi(t)}$

  $A(t) = \sqrt{\mu(t)^2 + \nu(t)^2}$ is *Instantaneous* ***Amplitude***.
  $\phi(t) = atan\left(\frac{\nu(t)}{\mu(t)}\right)$ is the *Instantaneous* ***Phase***.

The most widespread method to derive $\nu(t)$ from $\mu(t)$:

Analytic signal

$S(t) = \mu(t) + iHT(\mu(t))$, where HT() is the Hilbert Transform.

A basic family of signals

Monocomponent Signal

$$S(t) = A(t)e^{i\phi(t)}; \frac{\partial \phi(t)}{\partial t} \geq 0.$$

The real signal: $\mu(t) = A(t)cos(\phi(t))$

$\frac{\partial \phi(t)}{\partial t}$ is the *Instantaneous* ***Frequency*** (IF).

$IF \geq 0$ guarantees physical interpretability.

The most simple monocomponent signal is the circular signal (single oscillation):

$\phi(t)$ is an increasing function describing a complete trajectory around the unit circle only once in a cycle:

The most simple circular signal is $S(t) = cos(t) + i\ sin(t)$.

Most of the real signals, as the ECG, exhibit several oscillations. A superposition of components is needed to correctly represent the signal.

The AM-FM representation of a multicomponent signal

$$\mu(t) = \sum_{J=1}^{k} A_J(t) cos(\phi_J(t)); \ \frac{\partial \phi_J(t)}{\partial t} \geq 0$$

AM refers to the modulation of $A(t)$ and FM refers to the modulation of $\phi(t)$

There are multiple decompositions/AM-FM representations for a given signal.

Each problem or application has a suitable solution.

The one with the least number of components, physiologically interpretable, is often the most efficient and useful in practice.

Two well known decomposition methods

The Fourier decomposition: FD

Each component J is defined as $A_J cos(\frac{2\pi t}{P_J})$.

In addition to the amplitude parameter,

The components only differ in period, P_J. [1df]

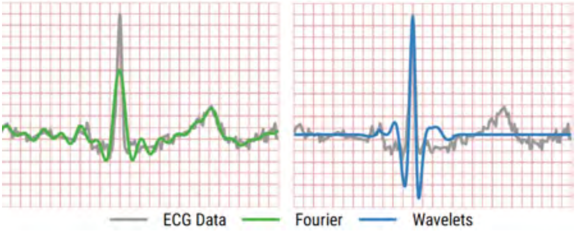
The wavelet decomposition: WD

The components come from a set of orthonormal series generated by a wavelet.

In addition to the amplitude parameter,

The components differ in period and location. [2df]

Neither FD nor WD are efficient methods for oscillatory signals defined in $(0, 2\pi]$, as the ECG.



FD (15 components); WD(frequency resolution=1/20,Period = [1,50])

FD and WD fail reconstructing this signal and use a large number of non identifiable coefficients.

The ECG signal is a combination of components with the same period but much different in location and shape.

The ECG signal is corrupted by multiple noise sources, such as muscle artefact or instrumentation noise.

Other decomposition methods/signal representation approaches such as, Wigner-Ville distribution, EMD (empirical model decomposition) also fails in reconstructing and giving interpretable components for this signal.

Table of Contents

Möbius waves

$$W(t,\alpha,\beta,\omega) = cos(\phi(t)) = cos(\beta + 2atan(\omega tan(\frac{t-\alpha}{2}))$$

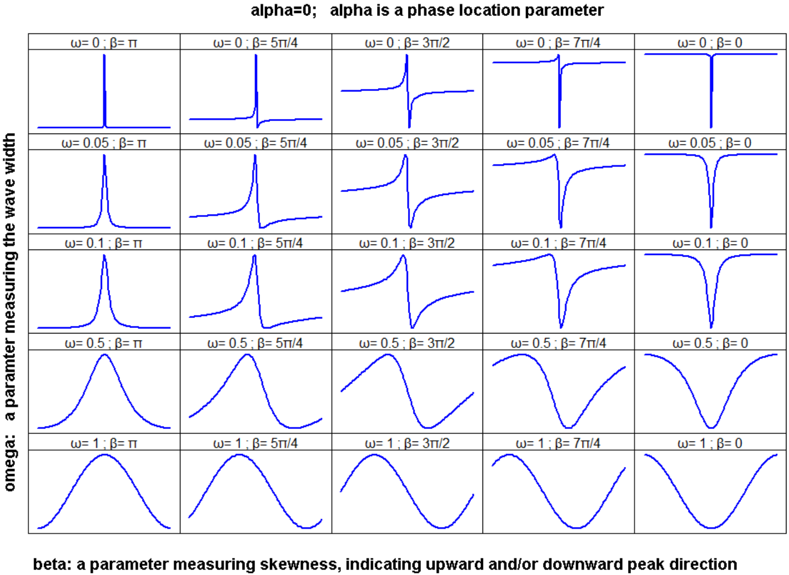
The waves differ in location (α), width (ω), and direction (β). [3df]

Where the Möbius waves come from?

A Möbius wave is the real part of a complex signal defined by the Möbius transform, which is a monocomponent signal (IF>0) with $A(t) = 1$.

Two equivalent formulations for the Möbius complex signal

- $S(t) = e^{i\phi(t)} = \tau_{ab}(z), \quad t \in (0, 2\pi], z = e^{it} \in \mathbb{C}$, and $\tau_{ab}(z)$ is the Möbius transform :
  For $\quad a, b \in \mathbb{C}, |a| < 1, |b| = 1, \tau_{ab}(z) = b\frac{z-a}{1-\bar{a}z}; a = re^{i\alpha}$ and, $b = e^{i\varphi}$.
- $S(t) = e^{i\phi(t)}; \phi(t) = \beta + 2atan(\omega tan(\frac{t-\alpha}{2}))$. $\varphi = \beta - \alpha$ and $r = \frac{1-\omega}{1+\omega}; \varphi, r \in \Re$ and, $\alpha \in (0, 2\pi]$.


alpha=0; alpha is a phase location parameter
ω= 0 ; β= π
ω= 0 ; β= 5π/4
ω= 0 ; β= 3π/2
ω= 0 ; β= 7π/4
ω= 0 ; β= 0
ω= 0.05 ; β= π
ω= 0.05 ; β= 5π/4
ω= 0.05 ; β= 3π/2
ω= 0.05 ; β= 7π/4
ω= 0.05 ; β= 0
ω= 0.1 ; β= π
ω= 0.1 ; β= 5π/4
ω= 0.1 ; β= 3π/2
ω= 0.1 ; β= 7π/4
ω= 0.1 ; β= 0
ω= 0.5 ; β= π
ω= 0.5 ; β= 5π/4
ω= 0.5 ; β= 3π/2
ω= 0.5 ; β= 7π/4
ω= 0.5 ; β= 0
ω= 1 ; β= π
ω= 1 ; β= 5π/4
ω= 1 ; β= 3π/2
ω= 1 ; β= 7π/4
ω= 1 ; β= 0
omega: a paramter measuring the wave width
beta: a parameter measuring skewness, indicating upward and/or downward peak direction

The FMM as a decomposition approach

FMM is the acronym of **Frequency Modulate Möbius**. It is an FM decomposition approach that represents a signal as the sum of k individual components, each one being a scaled Möbius wave.

The FMM$_k$ real signal

$\mu(t) = \sum_{J=1}^{k} A_J W(t, \alpha_J, \beta_J, \omega_J)$.

Using simple properties of the HT, the Analytic Signal expression is derived:

The FMM$_k$ Complex, Analytic Signal

$S(t) = \sum_{J=1}^{k} A_J e^{i\phi_J(t)}$; $\phi_J(t) = \beta_J + 2atan(\omega_J tan(\frac{t-\alpha_J}{2}))$.

$S(t) = \mu(t) + i\nu(t)$; where, $\nu(t) = \sum_{J=1}^{k} A_J sin(\phi_J(t))$.

with some mathematics a closed expression of the IF can be found

Compared with FD or WD the FMM is much more flexible for oscillatory signals like the ECG. The components have [3df], while the wavelets components have [2df] and Fourier [1df].

In real practice we have n observations: $X(t_i)$, $t_1 < ... < t_n$, $t_i \in (0, 2\pi]$.

The problem is to reconstruct the underlying signal $\mu(t)$ and finding identifiable and interpretable parameters, taking into account the presence of noise.

The FMM$_k$ model

$$X(t_i) = M + \sum_{J=1}^{k} A_J W(t_i, \alpha_J, \beta_J, \omega_J) + \epsilon(t_i)$$

Where, for $J = 1, ...k$,

- Parameter Space: $M, A \in \Re$; $\omega_J \in [0, 1]$; $\alpha_J, \beta_J \in (0, 2\pi]$.
- Restrictions: $A_1 = max_{1 \leq J \leq k} A_J$; $A_J \geq 0$; $\alpha_1 \leq \alpha_2 \leq .... \leq \alpha_k \leq \alpha_1$.
- Error term: $(\epsilon(t_1), ..., \epsilon(t_n))' \sim N_n(0, \sigma^2 I)$.

The restrictions guarantee the identifiability of the model parameters.

The inclusion of more restrictions, specific to each application, makes possible to find biologically interpretable solutions.

MLE of the FMM_k model parameters

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{i=1}^{n} (X(t_i) - \mu(t_i, \theta))^2,$$

where, θ is the vector of paramters, Θ refers to the parameter space.

Well known results in nonlinear normal regression guarantee the consistency and asymptotic normality of the MLE estimators, as standard regularity conditions on the response function are verified when the true parameter configuration verifies $\alpha_J \in (0, 2\pi], \beta_J \in (0, 2\pi], w_J > 0; J = 1, ...k$.

The optimization problem is computationally intensive.

The backfitting algorithm

Algorithm

- **Initialize** for $J = 1, \ldots, k$:
  $M = \frac{1}{n}\sum_{i=1}^{n} X(t_i); A_J = 0, \alpha_J = 5, \beta_J = \pi, \omega_J = 1; J = 1, \ldots, k$

- **Do until** R^2 **increases less than** C ,
  - $\rightarrow$ For each $J; J = 1, \ldots k$:
    $$\underset{M, A_J, \alpha_J, \beta_J, \omega_J}{\operatorname{argmin}} \sum_{i=1}^{n} (X(t_i) - \sum_{I \neq J} \hat{A}_I W(t_i, \hat{\alpha}_I \hat{\beta}_I, \hat{\omega}_I) - \mathrm{M} - \mathrm{A_J W}(\mathrm{t_i}, \boldsymbol{\alpha_J}, \boldsymbol{\beta_J}, \boldsymbol{\omega_J}))^2 =$$
    $$= \hat{M}, \hat{\alpha}_J, \hat{\beta}_J, \hat{\omega}_J$$
  - $\rightarrow$ Order the components using $A_1 = \max_{1 \leq J \leq k} A_J$ and $\alpha_1 \leq \ldots \leq \alpha_k \leq \alpha_1$
  - $\rightarrow$ $\mu(t_i, \hat{\theta}) = \hat{M} + \sum_{J=1}^{k} \hat{A}_J W(t_i, \hat{\alpha}_j \hat{\beta}_J, \hat{\omega}_J)$
  - $\rightarrow$ R^2

In the internal loop, at each step, a single FMM component is fitted to the residue with a *grid search plus Nelder-Mead* algorithm. The output being a preliminary solution and a value R^2, a measure of the explained variance.

The output of the internal loop are the initial values for the subsequent step of the external loop. The stop criteria for the external loop relies in R^2.

The solution converges in probability to a local minimum.

Our experience returns an excellent performance in different scenarios.

Table of Contents

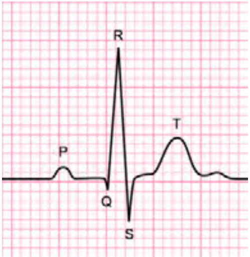
A specific FMM model designed for the beat-by-beat analysis of ECG signals.

The FMM_{ecg} model

$$X(t_i) = M + \sum_{J \in \{P,Q,R;S,T\}} A_J W(t_i, \alpha_J, \beta_J, \omega_J) + \epsilon(t_i)$$

Where, for $J \in \{P, Q, R, S, T\}$,

- $M \in \Re, A_J \in \Re^+; \beta_J \in (0, 2\pi];\omega_J \in [0, 1];\alpha_J \in (0, 2\pi]$.
- $\pi \leq \alpha_P \leq \alpha_Q \leq \alpha_R \leq \alpha_S \leq \alpha_T \leq \pi$.
- $(\epsilon(t_1), ..., \epsilon(t_n))' \sim N_n(0, \sigma^2 I)$.



A tricky identification algorithm using additional restrictions is designed to identification of the waves.

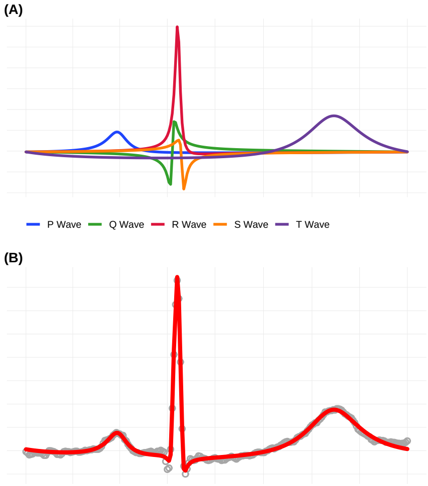
The new restrictions are based on previous knowledge, such as:

- $0.005 < \omega_J < 0.5, J \in \{P, Q, R; S, T\}$.
- $cos(\alpha_Q - \alpha_R) > 0.7;cos(\alpha_S - \alpha_R) > 0.5; cos(\alpha_T - \alpha_R) < 0.7$.

The hidden waves representing the heart's electric system, are uncovered by the FMM_{ecg} model.

The parameters are physiologically interpretable as they characterize the wave shapes, similar to what a cardiologist would look in the signals.

As a result, the model identifies which part of the system is working in a regular or failing way, quite useful for automatic diagnosis.



A typical heartbeat pattern. (A): the predicted five FMM_{ecg} waves. (B): the observed and predicted FMM_{ecg} signal.

Results I: some of the achievements

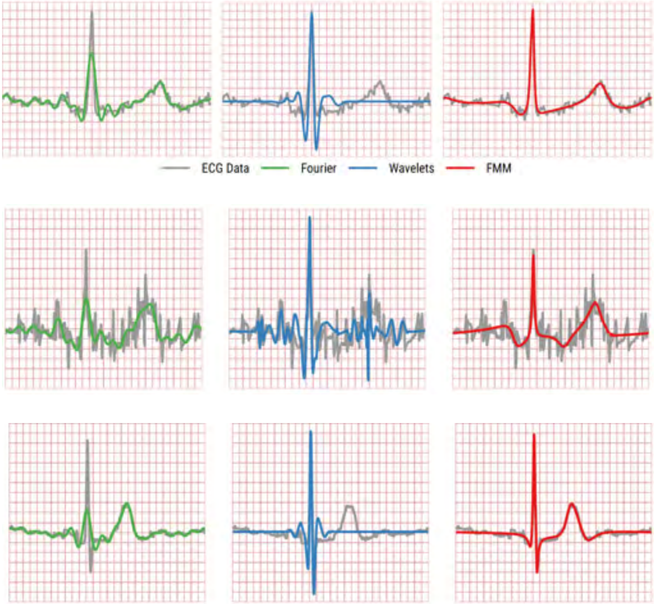
Until now: more than 1.000.000 heartbeats from around 36.000 patients have been analysed.

The model accurately predicts most ECG signals

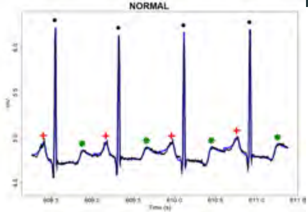
Percentiles of the R^2 values : $P_5 = 92.3\%$ and $P_{50} = 98.3\%$.

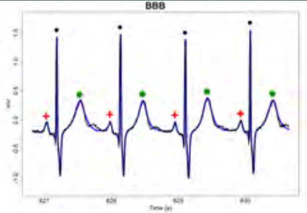
- Outperforms FD or WD and also an approach that uses a combination of Gaussian components.
- Outperforms several recent machine learning approaches in the task of detecting fiducial marks of the main waves.
- The capability to discriminate subjects has checked with signals from 105 patients from a well known data base (QT database).
- Simulation experiments conducted using different healthy and pathological patterns, highlight the good statistical properties of estimators.

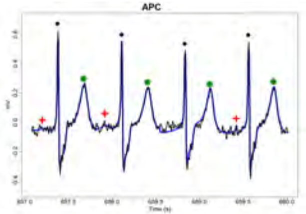
Results I: FMM against FD and WD

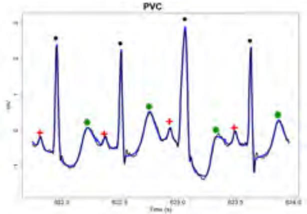


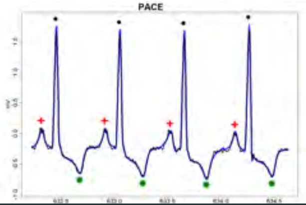
Results I: Examples QT database

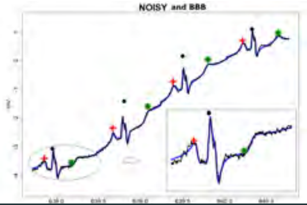




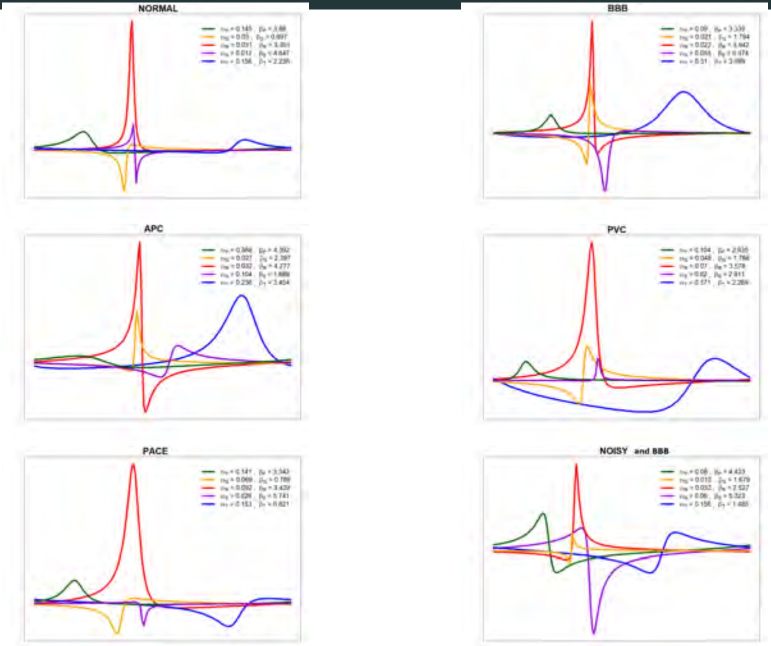








Results I: Examples QT database



Blundle Brunch Block (BBB)

A defect of the electrical conduction of the heart that implies ventricular enlargement or hypertrophy.

Two main forms, LBBB and RBBB, for left or right ventricle blocks, respectively.

These defects are associated with a higher risk of different cardiovascular diseases including acute myocardial infarction. LBBB is much more serious than RBBB.

LBBB rule

$\omega_R > 0.06$

BBB rule

$\omega_R > 0.06$ or [$\omega_S > 0.05, \omega_R > 0.025$]

	Nº patients	SE(LBBB)	SP(NORM)	SP(ALL)
PTB-XL	18282	98%	99%	88%
Georgia	10170	96%	96%	84%
CPCS	6611	93%	98%	86%

SE(BBB)	SP(NORM)	SP(ALL)
96%	90%	79%
96%	87%	72%
95%	92%	77%

Tables: Sensitivity (SE) and specificity (SP) from three data bases. NORM is the label for Normal patients. Category ALL include all the patients except those with LBBB/BBB labels.

- **The FMM is a multi-purpose approach** that solves questions, such as the extraction of interpretable features, the detection of marks of the fundamental waves, the generation of synthetic data or signal compression.
- **The method is statistically and mathematically sound.**
- **Outperforms data-driven and model-based methods**, simultaneously, for the ECG analysis.
- **The greatest benefit from this new discovery is its potential as automatic interpretation method.**
- The user can try the model and the diagnostic rules in: **the R package FMM** and a pilot app: **https://fmmmodel.shinyapps.io/fmmEcg/**.

Table of Contents

Colaborators

All this has been possible, and continue to progress, thank to people in the FMM project: **http://www.eio.uva.es/fmm-project/**

Itziar Fernandez

Yolanda Larriba

Alejandro Rodriguez-Collado

Christian Canedo

Challenges for the future

- **Theoretical development** of 3D models and mathematical and statistical tools for the study of associations between signals.
- **Research on machine learning approaches** for supervised and unsupervised classification of waveforms/signals and application to different disciplines.
- **Design of ad hoc FMM models** to the analysis of many other biological such as ocular electrophysiology or brain signals.
- **Medical advances contributions** among the most important ones are the detection of cardiology pathologies or the determination of factors that influence the course of neurological diseases as Parkinson.
  Many exciting questions in chronobiology remain open, such as the relation of rhythmicity patterns with diseases as cancer or how the hormone patterns are related to physiological processes.

Recent Publications

- Rueda, C., Larriba, Y., and Peddada, S. D. (2019). Frequency Modulated Möbius Model Accurately Predicts Rhythmic Signals in Biological and Physical Sciences. Scientific reports, 9(1), 18701. https://doi.org/10.1038/s41598-019-54569-1
- **Rueda, C., Larriba, Y. and Lamela, A. (2021). The Hidden Wave in the ECG Uncovered Revealing a Sound Automated Interpretation Method. Scientific Reports, 11, 3724.** https://doi.org/10.1038/s41598-021-82520-w
- Rueda, C., Rodríguez-Collado, A. and Larriba, Y. (2021). A Novel Wave Decomposition for Oscillatory Signals. IEEE Transactions on Signal Processing, 69, 960-972. https://doi.org/10.1109/TSP.2021.3051428
- Rodríguez-Collado, A. and Rueda, C. (2021). Simple Parametric Representation of the Hodgkin-Huxley Model. PLOS ONE, 16(7), e0254152. https://doi.org/10.1371/journal.pone.0254152
- Rodríguez-Collado, A. and Rueda, C. (2021). Electrophysiological and Transcriptomic Features Reveal a Circular Taxonomy of Cortical Neurons. Frontiers in Human Neuroscience, 15, 410. https://doi.org/10.3389/fnhum.2021.684950
- Rueda, C., Fernández, I., Larriba, Y. and Rodríguez-Collado, A. The FMM Approach to Analyze Biomedical Signals: Theory, Software, Applications and Future. (2021). Mathematics, 9, 1145. https://doi.org/10.3390/math9101145
- Fernández, I., Rodríguez-Collado, A., Larriba, Y., Lamela, A., Canedo, C. and Rueda, C. FMM: An R Package for Modelling Rhythmic Patterns in Oscillatory Systems (2021). https://arxiv.org/abs/2105.10168
- Rueda, C., Fernández, I., Larriba, Y., Rodríguez-Collado, A. and Canedo, C. (2021) Compelling new electrocardiographic markers for automatic diagnosis. Under revision.
- Larriba, Y., Rueda, C., Scheer (2021). A Circadian Gene Expression Atlas in Humans. Preprint.