Characterization of Spatio-Temporal Cardiac Action Potential Variability at Baseline and Under β-Adrenergic Stimulation by Combined Unscented Kalman Filter and Double Greedy Dimension Reduction

David Adolfo Sampedro-Puente ⓘ, Fabien Raphel ⓘ, Jesus Fernandez-Bes ⓘ,
Pablo Laguna ⓘ, Damiano Lombardi ⓘ, and Esther Pueyo ⓘ

Abstract—*Objective:* Elevated spatio-temporal variability of human ventricular repolarization has been related to increased risk for ventricular arrhythmias and sudden cardiac death, particularly under β-adrenergic stimulation (β-AS). This work presents a methodology for theoretical characterization of temporal and spatial repolarization variability at baseline conditions and in response to β-AS. For any measured voltage trace, the proposed methodology estimates the parameters and state variables of an underlying human ventricular action potential (AP) model by combining Double Greedy Dimension Reduction (DGDR) with automatic selection of biomarkers and the Unscented Kalman Filter (UKF). Such theoretical characterization can facilitate subsequent characterization of underlying variability mechanisms. *Material and Methods:* Given an AP trace, initial estimates for the ionic conductances in a stochastic version of the baseline human ventricular O'Hara *et al.* model were obtained by DGDR. Those estimates served to initialize and update model parameter estimates by the UKF method based on formulation of an associated nonlinear state-space representation and joint estimation of model parameters and state variables. Similarly, β-AS-induced phosphorylation levels of cellular substrates were estimated by the DGDR-UKF methodology. Performance was tested by building an experimentally-calibrated population of virtual cells, from which synthetic AP traces were generated for baseline and β-AS conditions. *Results:* The combined DGDR-UKF methodology led to 25% reduction in the error associated with estimation of ionic current conductances at baseline conditions and phosphorylation levels under β-AS with respect to individual DGDR and UKF methods. This improvement was not at the expense of higher computational load, which was diminished by 90% with respect to the individual UKF method. Both temporal and spatial AP variability of repolarization were accurately characterized by the DGDR-UKF methodology. *Conclusions:* A combined DGDR-UKF methodology is proposed for parameter and state variable estimation of human ventricular cell models from available AP traces at baseline and under β-AS. This methodology improves the estimation performance and reduces the convergence time with respect to individual DGDR and UKF methods and renders a suitable approach for computational characterization of spatio-temporal repolarization variability to be used for ascertainment of variability mechanisms and its relation to arrhythmogenesis.

Index Terms—Cardiac electrophysiological models, spatio-temporal variability, parameter estimation, joint estimation, unscented kalman filter, double greedy dimension reduction.

I. INTRODUCTION

CLINICAL, experimental and computational studies have demonstrated the important role of cardiac spatio-temporal variability in electrical function at a whole range of scales from the cellular action potential (AP) to the body surface electrocardiogram [1]–[7]. Spatial variability refers to electrophysiological differences between cardiac cells or regions of cells and has been to some extent attributed to distinct ionic current contributions to individual APs [8]–[12]. Temporal variability refers to AP differences between cardiac beats and has been suggested to arise from random fluctuations in ionic currents as well as variations in intracellular calcium handling [8], [10], [13]–[16]. Particularly regarding variability in ventricular repolarization, i.e. in the return of ventricular cells to

their resting state after a depolarization, numerous investigations have associated elevated temporal and/or spatial variability with pro-arrhythmicity and sudden cardiac death [17]–[21].

β-adrenergic stimulation (β-AS) has been shown to produce exaggerated increases in beat-to-beat variability of repolarization (BVR), particularly under conditions of reduced repolarization reserve [2], [22], [23]. *In vitro* experiments in isolated cardiomyocytes have suggested that this elevation in BVR by β-AS is a relevant contributor to arrhythmogenesis by the development of afterdepolarizations and triggered activity [2], [9], [22], [24], [25]. In an *in vivo* animal model of long-QT1 syndrome, β-AS has been shown to induce increments in both temporal and spatial dispersion of repolarization and to facilitate the development of early afterdepolarizations (EADs) and left ventricular aftercontractions, altogether providing the substrate and triggers for the ignition of Torsade de Pointes, a life-threatening ventricular arrhythmia [23]. Computational investigations have further contributed to shed light into the mechanisms underlying the relationship between β-AS-induced elevation in BVR and pro-arrhythmic risk [25]–[27]. Nevertheless, most of the computational approaches employed so far in the literature do not concomitantly account for realistic modeling of both cell-to-cell and beat-to-beat AP differences [21], [26], [28]–[32], which should be fundamental to better understand the relationship between BVR and arrhythmogenesis and its modulation by β-AS.

For the above reasons, the development of stochastic cardiac computational AP models fed with information acquired from human cells or tissues becomes of major interest. In recent years, different methodologies have been proposed to integrate information from cardiac AP signals, or from a set of markers derived from them, by identifying the values of parameters and/or state variables of an underlying electrophysiological model. This allows obtaining a population of virtual AP models representative of a set of experimental data of interest, with the advantage of facilitating assessment of the causes and consequences of BVR by simultaneous assessment of voltage and ionic currents/concentrations. In [28]–[30], methodologies based on Genetic Algorithms, Moment-Matching and Gaussian Process Emulators were designed for parameter identification at a population level, thus allowing to reproduce the overall AP characteristics in the investigated cell population but hampering individual identification of the parameters associated with each cellular AP trace. In [31], [32] ionic parameters were estimated from voltage signals by using Markov Chain Monte Carlo (MCMC)-based methods, which enable parameter estimation for each individual cell. However, on top of the high computational load associated with these methods, they do not account for beat-to-beat variability and do not provide an estimation for other non-measurable state variables of the model, such as ionic concentrations or channel open probabilities, as neither do the methods proposed in [28]–[30]. In a work of ours [33], a methodology based on nonlinear state-space representations [34] and the Unscented Kalman Filter (UKF) [35] was proposed to identify the parameters and state variables of stochastic human ventricular AP models. This methodology provided robust one-to-one model parameter and state estimation for each AP

trace individually, but the computational load was high and it required a long AP signal for accurate estimation.

On the basis of the above described limitations, a methodology for AP model parameter and state estimation that combines fast methodologies based on biomarkers' information with other more complex methodologies based on AP traces' information could be most useful. When condensing AP data into a set of biomarkers, it is important to keep a sufficient amount of information to avoid any risk of degradation in the estimation. To ensure this, the number of biomarkers, also called dictionary entries, can be potentially substantial (hundreds, thousands, ...) and even higher than the sample size. In this regime, various phenomena can appear, referred to as curse of dimensionality [36], which require data processing to improve classification or regression. Recent studies have addressed this by using Double Greedy Dimension Reduction (DGDR) [37]. In DGDR, the parameter estimations from AP signals are obtained by building a low-dimensional classifier input, which is generated by projecting the dictionary entries into a low-dimensional linear subspace to improve the success rate of a given classification problem. This linear subspace is automatically built by a sparse linear combination of the dictionary entries to prevent any over-fitting risk [38].

The present study proposes the combined use of DGDR- and UKF-based methodologies to extract information from AP signals at baseline and under β-AS. Initial DGDR parameter estimates are used to initialize and/or update subsequent UKF estimates so as to facilitate that these remain close to their actual values. To assess the performance of our proposed methodology, a population of stochastic human ventricular cell models is constructed and used to run simulations at baseline conditions and following β-AS. Methodological performance is first tested over the synthetic AP signals generated for baseline conditions, from which a set of ionic current conductances are inferred for each virtual cell. In a second step, the methodology is tested over synthetic AP signals of the same population following β-AS, from which the phosphorylation levels of a set of cellular substrates are inferred, considering the previously identified ionic conductances. The ability of our methodology to characterize spatial and temporal variability in human ventricular repolarization is demonstrated, showing remarkable improvement with respect to the individual use of DGDR- or UKF-based methods while keeping the computational load at affordable levels.

To the best of our knowledge, this is the first work where a biomarker-based estimation method, like DGDR, and an AP-driven method, like the one grounded on state-space representations and UKF, are combined to obtain a more robust and faster parameter and state variable identification for cardiac AP models. Another major novelty of the present work lies on the fact that parameters and state variables of an underlying cardiac AP model are not only identified for baseline conditions, but, importantly, for β-AS conditions. This is particularly relevant provided the role of β-AS in modulating spatio-temporal ventricular repolarization variability and facilitating the development of arrhythmias at tissue and whole-heart levels. Our work thus provides an important tool to assess mechanisms

underlying cardiac spatio-temporal variability and identify those with pro-arrhythmic potential.

II. METHODS

A. Stochastic AP Models at Baseline and Under β-AS

1) Stochastic Human Ventricular ORd Model: A stochastic version of the O'Hara-Virág-Varró-Rudy (ORd) human ventricular epicardial AP model [39] was developed to account for BVR. Following the subunit-based approach described in [8], the set of ordinary differential equations (ODEs) describing ion channel gating for the four principal currents active during AP repolarization, namely I_{Ks} (slow delayed rectifier potassium current), I_{Kr} (rapid delayed rectifier potassium current), I_{to} (transient outward potassium current) and I_{CaL} (L-type calcium current), were transformed into stochastic differential equations (SDEs) by adding a stochastic term of the form shown in Eq. 1 for a generic ionic gate x_g, where x_{g_∞} is the steady-state value and τ_{x_g} is the time constant of this generic ionic gate x_g:

$$dx_g = \frac{x_{g_\infty} - x_g}{\tau_{x_g}}dt + \frac{\sqrt{x_{g_\infty} + (1 - 2x_{g_\infty})x_g}}{\sqrt{\tau_{x_g} N_g}}dw. \quad (1)$$

The added stochastic term containing the increments of a Wiener process (dw) multiplied by a factor inversely proportional to the number of ion channels (N_g) of the corresponding type was added to the deterministic term defining x_g gating. By including this stochastic term with an accurately estimated number of channels, realistic fluctuations in the ionic gates and the whole-cell ionic currents are reproduced, which are the source for BVR in cellular AP. The number of channels N_g associated with I_{Ks}, I_{Kr}, I_{to} and I_{CaL} were calculated by dividing the default ionic conductance values in the ORd model by the corresponding single channel conductances experimentally reported in the literature, as described in [33].

2) β-Adrenergic Signaling Model: β-AS effects were modeled following the approach described in [26], where a modified version of the Xie *et al.* model [40], with definition of graded and dynamic phosphorylation levels of cellular protein kinase A (PKA) substrates, was used. The Xie *et al.* model was updated from the original β-adrenergic signaling formulation proposed in [41] to slow down the I_{Ks} phosphorylation and dephosphorylation rate constants to fit experimental observations. PKA-mediated phosphorylation of phospholemman was accounted for in [40] by increasing the Na$^+$-K$^+$ (NaK) pump affinity for intracellular Na$^+$ concentration.

B. Synthetic Data

A population of stochastic AP models was constructed to reproduce the experimentally reported inter-individual variability in human ventricular electrophysiological properties. An initial population of virtual cells was generated by using a Monte-Carlo method in which the conductances of eight main ionic conductances were varied in the range $\pm100\%$ of their nominal values in the ORd model, with those currents being: I_{Ks}; I_{Kr}; I_{to}; I_{CaL}; inward rectifier potassium current, I_{K1}; sodium current,

TABLE I
CALIBRATION CRITERIA FOR HUMAN VENTRICULAR CELL MODELS

AP characteristic	Min. accept. value	Max. accept. value
Under baseline conditions ([39], [42], [43])		
APD$_{90}$ (ms)	178.1	442.7
APD$_{50}$ (ms)	106.6	349.4
RMP (mV)	-94.4	-78.5
V_{peak} (mV)	7.3	-
Under 90% I_{Ks} block ([39])		
ΔAPD$_{90}$ (%)	-54.4	62
Under 70% I_{Kr} block ([44])		
ΔAPD$_{90}$ (%)	34.25	91.94
Under 50% I_{K1} block ([45])		
ΔAPD$_{90}$ (%)	-5.26	14.86
Na$_i^+$ concentration under baseline conditions ([46])		
Max. conc. (μM)	-	39.27
Ca$_i^{2+}$ concentration under baseline conditions ([47])		
Max. systolic (μM)	-	2.23
Max. diastolic (μM)	-	0.40

I_{Na}; sodium-calcium exchanger current, I_{NaCa}; and sodium-potassium pump current, I_{NaK}. This corresponded to definition of eight multiplying conductance factors, namely θ_{Ks}, θ_{Kr}, θ_{to}, θ_{CaL}, θ_{K1}, θ_{Na}, θ_{NaCa}, θ_{NaK}, varying between 0 and 2. From the 8000 initially generated models, only 2373 models presenting electrophysiological properties within physiologically plausible limits were retained, with those limits shown in Table 1 as determined based on [39], [42]–[47]. The quantified properties at baseline included AP duration (APD) at 90% (APD$_{90}$) and 50% (APD$_{50}$) repolarization, resting membrane potential (RMP), peak membrane potential (V_{peak}) and maximal concentrations of intracellular sodium (Na$_i^+$) and calcium (Ca$_i^{2+}$). Also, the percentage of change in APD$_{90}$ (ΔAPD$_{90}$) after potassium current inhibitions were computed for each individual cell and used as part of the calibration process. The retained models represent virtual cells with distinct ionic properties.

To simulate a range of potentially different β-AS effects in the constructed population of stochastic AP models, multiplying factors θ_{fCal}, θ_{fKs} and θ_{fNaK} for the PKA phosphorylation levels f_{CaL}, f_{Ks} and f_{NaK} were varied so that these phosphorylation levels ranged between the values at baseline (i.e. without Isoproterenol (ISO)) and the values after application of an ISO dose of 1 μM associated with maximal effects. This population of phosphorylation levels, generated by using a Monte-Carlo method, was combined with the above described population of stochastic AP models to obtain a global population of 2373 models with 11 simultaneously varying parameters. This population was divided into training and validation subpopulations with 2000 and 373 models, respectively.

AP traces of 1100 beats were simulated at baseline and under β-AS, respectively, by applying 1-ms rectangular stimulus pulses of 52 pA/pF amplitude delivered at 1 Hz pacing frequency. The Euler-Maruyama scheme was used to solve the SDEs with an integration time step of dt = 0.02 ms that ensured numerical convergence. The last 100 beats of each condition (baseline, β-AS) were used for further analysis to ensure convergence had been reached.

Independent standard Gaussian noise was added to the synthetically generated AP data, as described in [33], to simulate recording noise as in experimentally acquired data. These noisy APs were input to the estimation methodologies tested in this study.

C. State-Space Formulation and Augmented States

1) State-Space Formulation: The stochastic version of the ORd model with unknown ionic conductance factors (for baseline conditions) or phosphorylation levels (for β-AS conditions) was formulated as a non-linear discrete-time state-space model [34] following the approach described in [33]. In these state-space models the only measured variable was considered to be transmembrane voltage, while there were a number of hidden variables, including ionic concentrations and opening probabilities of ionic gates.

For baseline conditions, model parameters to be estimated were the factors multiplying the nominal conductances of I_{Ks}, I_{Kr}, I_{to}, I_{CaL}, I_{K1}, I_{Na}, I_{NaCa} and I_{NaK}. Hence, the vector of static model parameters was $\boldsymbol{\theta} = \{\theta_{Ks}, \theta_{Kr}, \theta_{to}, \theta_{CaL}, \theta_{K1}, \theta_{Na}, \theta_{NaCa}, \theta_{NaK}\}$, representing variations in the ionic conductances relative to the default values in the ORd model, $I_j = I_{j,ORd}\theta_j$, where $j \in \{Ks, Kr, to, CaL, K1, Na, NaCa, NaK\}$. Note that the same factor θ_j applies to the number of ion channels of each species: $N_j = N_{j,ORd}\theta_j$, as the unitary conductance of each ionic species was assumed to be constant based on reported experimental findings [48].

For β-AS conditions, model parameters to be estimated were the factors multiplying the phosphorylation levels of the PKA substrates whose phosphorylation had a remarkably higher impact on the AP as evaluated in performed computational tests, which were I_{Ks}, I_{CaL} and I_{NaK} currents [49], in agreement with findings reported for other β-adrenergic signaling models [50]. Consequently, the vector of static model parameters was $\boldsymbol{\theta} = \{\theta_{fKs}, \theta_{fCaL}, \theta_{fNaK}\}$, representing variations in the phosphorylation levels f_{Ks}, f_{CaL} and f_{NaK} relative to the default values in the modified Xie model, $f_j = f_{j,Xie}\theta_j$, where $j \in \{Ks, CaL, NaK\}$. For both baseline and β-AS conditions, the vector $\boldsymbol{\theta}$ of model parameters was estimated for each given input AP trace.

The state-space representations were of the form:

$$\mathbf{x}(k) = f(\mathbf{x}(k-1), \mathbf{q}(k-1), \boldsymbol{\theta}) \tag{2}$$

$$y(k) = h(\mathbf{x}(k)) + r(k), \tag{3}$$

where the process equation (Eq. 2) was defined by a non-linear function $f(\cdot)$ with three different input vectors: $\mathbf{x}(k)$, containing the state variables of the stochastic AP model; $\mathbf{q}(k)$ representing non-additive process noises related to Wiener increments; and $\boldsymbol{\theta}$ containing the model parameters to be estimated. On the other hand, the measurement equation (Eq. 3) was defined by the function $h(\cdot)$ relating the measured variable (transmembrane voltage) with the vector of the model state variables. In this study, $y(k) = v(k) + r(k)$, where $v(k)$ represents the noiseless AP and $r(k)$ was assumed to be an additive white Gaussian noise.

2) Augmented State-Space: To perform joint estimation of model parameters and state variables for a given input noisy AP, the state-space representation of Eq. (2)–(3) was reformulated as described in [33]. In brief, *state augmentation* [34] was applied to convert the static parameter vector $\boldsymbol{\theta}$ into a time-varying parameter vector $\widetilde{\boldsymbol{\theta}}(k)$ using a random walk model with drift:

$$\widetilde{\boldsymbol{\theta}}(k) = \widetilde{\boldsymbol{\theta}}(k-1) + \boldsymbol{\delta}(k), \tag{4}$$

where $\boldsymbol{\delta}(k)$ represents an artificial noise whose components were defined by i.i.d. zero-mean Gaussian processes with very small variance. An augmented state vector $\mathbf{z}(k)$ was built by joining the state variable vector $\mathbf{x}(k)$ with the new parameter vector $\widetilde{\boldsymbol{\theta}}(k)$ and the process noise vector $\mathbf{q}(k)$:

$$\mathbf{z}(k) = \left[\mathbf{x}(k), \mathbf{q}(k), \widetilde{\boldsymbol{\theta}}(k)\right]^T. \tag{5}$$

The previous process (Eq. 2) and measurement equations (Eq. 3) were replaced with:

$$\mathbf{z}(k) = f_a(\mathbf{z}(k-1)) + \boldsymbol{\epsilon}(k) \tag{6}$$

$$y(k) = h_a(\mathbf{z}(k)) + r(k), \tag{7}$$

where f_a and h_a are the augmented versions of f and h, respectively, and $\boldsymbol{\epsilon}(k)$ contains noises related to the Wiener increments of the stochastic AP model represented by $\mathbf{q}(k)$ and to the new parameter vector $\widetilde{\boldsymbol{\theta}}(k)$ represented by $\boldsymbol{\delta}(k)$.

D. Individual and Combined DGDR- and UKF-Based Methods

1) DGDR: The DGDR method was used to estimate the parameters of the stochastic AP model, which represent part of the components of the augmented state vector $\mathbf{z}(k)$. DGDR is based on high-dimensional data analysis and aims at mitigating the curse of dimensionality [36] by projecting data into a low subspace through a sparse linear combination of the dictionary entries. In this work, the dictionary entries comprised a set of biomarkers and combinations of them, with these biomarkers related to AP amplitude and duration, such as V_{peak}, RMP and APD at different repolarization levels, biomarkers related to BVR, such as short and long-term variability, as well as wavelet decomposition of the input AP signal. In [37], data projection is performed such that a classification success rate is maximal, which can be achieved by maximizing a score function based on the distributions of the projected data of each class. To apply the DGDR method to regression problems, the cost function was replaced by an ℓ^2 norm that minimizes the error between the actual values of the ionic conductances or phosphorylation levels and a sparse linear combination of the dictionary entries in a training set:

$$\omega_* = \arg\min_{\omega_i} \left\| \sum_{i=1}^{n_g} \omega_i g_i - \theta_c \right\|_{\ell^2} \tag{8}$$

where ω_i are the weights to be determined, n_g is the number of dictionary entries, g_i is the ith dictionary entry of the training set and θ_c are the known values of the parameters in the training set.
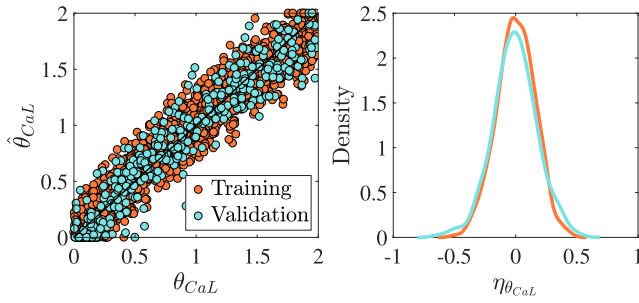
Fig. 1.   Left panel: Estimated ($\hat{\theta}_{CaL}$) vs actual (θ_{CaL}) values of the factor multiplying maximal I_{CaL} in the training and validation populations. Right panel: Density of the absolute error in the estimation of θ_{CaL} for the training and validation sets.

As in [37], the early stopping criterion was applied on a validation set to avoid over-fitting risk, which leads to a sparse combination of the dictionary entries and the weight vector ($\|\omega\|_{\ell^0} \ll n_g$). Thus, given a new AP trace in the validation set, the learned linear combination was applied to estimate the model parameters. For this study the number of extracted dictionary entries was 889, of which 100 were selected for the linear combination, as this already led to improvements in all estimation errors below 10^{-3} when adding a new dictionary entry over the training set. The linear combination of 100 entries was a good choice to minimize the cost function in the training set while avoiding over-fitting in the validation set. As expected, the dictionary entries selected for the estimation of each model parameter were strongly related to the AP phase where the ionic conductances or phosphorylation factors have a more dominant role. As an example, the most relevant biomarkers for estimation of θ_{Na} and θ_{K1} were related to the AP upstroke velocity and RMP, respectively.

A learning phase was separately performed for each of the model parameters to be estimated. The selected dictionary entries were not the same, which is a direct consequence of the goal-oriented concept of the DGDR method and ensures a certain explanation of the selected entries. The full process for the training step took around 3 hours on about 50 processors for the estimation of the eight ionic current conductances at baseline and proportionally less for the estimation of the three phosphorylation levels under β-AS. Once the learning phase was performed, the estimation of a new sample was immediate (scalar product between two vectors).

This training process was performed over a population of 2000 models while evaluation was carried out over 373 models, leading to adequate levels of accuracy. Fig. 1 (left panel) illustrates an example of θ_{CaL} estimation by DGDR, showing the uniform dispersion of the point cloud that provides a measure of the uncertainty in the estimation. In addition, the DGDR method led to similar accuracy levels for training and evaluation populations as can be observed in Fig. 1 (right panel) where the distribution of the absolute error between the actual (θ_{CaL}) and estimated ($\hat{\theta}_{CaL}$) parameter values is shown. Similar results were obtained in the estimation of the other model parameters. These results

served to support the adequacy of separating the population into a training set of 2000 models and a validation set of 373 models.

2) UKF: The UKF [35] was used to estimate the states of the nonlinear state-space formulation described by Eq. (6)–(7), which provides estimates for the parameters and state variables of a stochastic human ventricular cell model for any given AP trace. The values of three UKF setting parameters, commonly denoted by α, β and κ, were set to define the spreading of Sigma-Points around the mean state estimates (controlled by α and κ) and to reflect prior knowledge of states' statistical distributions (controlled by β). In this work, $\alpha = 1$, $\beta = 0$ and $\kappa = 3$-L were set [51], being L the number of states (L = 71 for baseline and L = 68 for β-AS conditions). This led to a value for the spread of the state covariance matrix corresponding to $\sqrt{\gamma} = 1.7321$, in accordance with feasible values [52], and to sums of weights of means and covariances equal to one: $\sum_{i=0}^{2L} W_i^{(c)} = 1$, $\sum_{i=0}^{2L} W_i^{(m)} = 1$.

Two additional hyper-parameters were set in the UKF implementation, which determine the process noise variance σ_θ^2 (the same for all components of the model parameter vector) and the measurement noise variance σ_r^2. A range of values for σ_θ^2 were tested and the one rendering best performance was selected. The value for σ_r^2 was set to 1 mV [33].

The initialization of the mean and covariance matrix of the state vector was obtained from the training population. The state variables related to stochastic AP model parameters (representing multiplying factors for ionic conductances at baseline and for phosphorylation levels under β-AS) were constrained to remain in the interval [0, 2].

3) Combined UKF-DGDR: DGDR and UKF methods were combined to enhance their individual characteristics in terms of estimation accuracy and computational cost. In particular, DGDR was used for **initialization** and **updating** of UKF estimation to take parameter estimates closer to their actual values and to avoid local minima in the estimation:

Initialization (INI): The model parameter estimates obtained by DGDR were used to initialize the corresponding elements of the state vector, which was subsequently estimated by UKF. DGDR provided estimates for both the mean of the parameter vector, $\hat{\theta}^{DGDR}$, and its covariance matrix, P^{DGDR}.

Updating (UP): The model parameter estimates obtained by DGDR were used to update the UKF-based parameter estimation in each cardiac cycle. At the end of each cycle, the corresponding elements of the state vector estimated by UKF (mean $\hat{\mathbf{z}}_k$ and covariance matrix $\mathbf{P}_k$) were updated according to the estimates for the mean $\hat{\theta}^{\mathbf{DGDR}}$ and covariance matrix $\mathbf{P}^{\mathbf{DGDR}}$ obtained by DGDR as follows:

$$\mathbf{d} = \hat{\mathbf{z}}^{\mathbf{DGDR}} - \mathbf{H}\hat{\mathbf{z}}_k \tag{9}$$

$$\mathcal{S} = \mathbf{H}\mathbf{P}_k\mathbf{H}^{\mathbf{T}} + \mathbf{P}^{\mathbf{DGDR}} \tag{10}$$

$$\mathcal{K}_{up} = \mathbf{P}_k\mathbf{H}^{\mathbf{T}}\mathbf{S}^{-1}, \tag{11}$$

with $\hat{\mathbf{z}}^{\mathbf{DGDR}} = [\mathbf{O}_{\mathbf{N}_x} \ \hat{\theta}^{\mathbf{DGDR}} \ \mathbf{O}_{\mathbf{N}_q}]$, where $\mathbf{O}_{\mathbf{N}_x}$ is a $N_x \times 1$ zero vector and $\mathbf{O}_{\mathbf{N}_q}$ is a $N_q \times 1$ zero vector, and $\mathbf{H}$ is a $(N_x + N_\theta + N_q) \times (N_x + N_\theta + N_q)$ matrix of 0 values everywhere

except for the last $N_\theta \times N_\theta$ submatrix occupied by an identity matrix. In the above, N_θ is the number of model parameters, N_x is the number of model state variables and N_q is the number of Wiener processes.

The UKF-based updated estimates for the mean and the covariance matrix of the state vector were:

$$\hat{\mathbf{z}}_{k_{up}} = \hat{\mathbf{z}}_k^- + \mathcal{K}_{up}\mathbf{d} \tag{12}$$

$$\mathbf{P}_{k_{up}} = (\mathbf{I}_{\mathbf{N}_\theta} - \mathcal{K}_{up}\mathbf{H})\mathbf{P}_k. \tag{13}$$

E. Performance Evaluation

The performance of DGDR, UKF and their combination was evaluated for estimation of eight ionic current conductances at baseline conditions and for estimation of three phosphorylation levels under β-AS conditions. In the latter case, the values for the eight ionic conductances were set at those estimated at baseline. The estimation performance was evaluated by [33]:

1) AP Estimation: The root mean square error between the original noiseless AP trace and the estimated AP trace was calculated over the last 5 cycles,

$$\xi_v = \sqrt{\frac{1}{K_N}\sum_{k=0}^{K_N-1}|v(k)-\hat{v}(k)|^2}, \tag{14}$$

where K_N is the number of samples contained within the last $N = 5$ cycles.

2) State and Parameter Estimation: The mean absolute error between the actual and estimated values of each state was calculated over the last 5 cycles,

$$\eta_{z_j} = \frac{1}{K_N}\sum_{k=0}^{K_N-1}\left\{|z_j(k)-\hat{z}_j(k)|\right\}, \tag{15}$$

where z_j is the actual value of the state variable j and $\hat{z}_j$ is the estimated value, with $j = 1, \ldots, L$, being L the length of the augmented state vector $\mathbf{z}(k)$.

A global accuracy measurement $\bar{\eta}_\theta$ of model parameter estimation was defined as the average of the mean absolute errors $\eta_{\theta_i}, i = 1, \ldots, N_\theta$, corresponding to all estimated model parameters:

$$\bar{\eta}_{\boldsymbol{\theta}} = \frac{1}{M}\sum_{\theta' \in \boldsymbol{\theta}}\eta_{\theta'}, \tag{16}$$

where $\eta_{\theta'}$ is the mean relative error for model parameter $\theta' \in \boldsymbol{\theta}$ and $M = 8$ (for conductance factors) or $M = 3$ (for phosphorylation factors).

3) Reproducibility of AP Markers: AP-derived biomarkers were calculated from the noiseless AP trace and from the estimated AP trace calculated according to the estimated values for the model parameters. AP biomarkers comprised mean and standard deviation of APD_{90} (in the following denoted by APD) and short-term variability (STV) of APD calculated as the average distance perpendicular to the identity line in the Poincaré plot [33], in both cases evaluated over the last
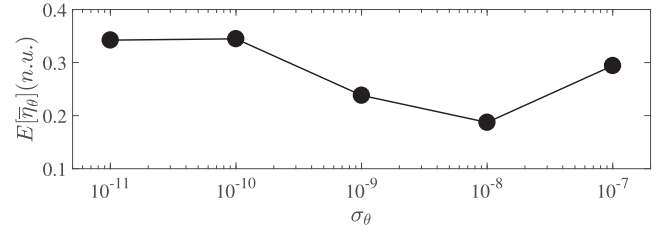

Fig. 2. Average of mean absolute parameter estimation error $E[\bar{\eta}_\theta]$ in the ORd model as a function of the standard deviation of the process noise σ_θ.

30 cycles:

$$m_{APD_{90}} = \frac{1}{N}\sum_{n=1}^{N}APD_{90}(n), \tag{17}$$

$$s_{APD_{90}} = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(APD_{90}(n) - m_{APD_{90}})^2}, \tag{18}$$

$$STV = \sum_{n=1}^{N-1}\frac{|APD_{90}(n+1) - APD_{90}(n)|}{(N-1)\sqrt{2}}. \tag{19}$$

III. RESULTS

A. Implementation of UKF Method

The performance of the UKF method as a function of the process noise standard deviation σ_θ is illustrated in Fig. 2, which shows the mean parameter estimation error in the ORd model when varying σ_θ by several orders of magnitude. The minimal average error $E[\bar{\eta}_\theta]$ was achieved for $\sigma_\theta = 10^{-8}$, which was used for all subsequent analyses. In the case of the root mean square error in AP estimation, ξ_v, its values were minimally affected by the choice of σ_θ for all tested σ_θ values.

In the following sections the estimation performances of the DGDR and UKF methods individually and in combination are presented.

B. Combined DGDR and UKF Methods: Initialization Effects

The use of the estimates obtained by DGDR for the mean, $\hat{\theta}^{DGDR}$, and the covariance matrix, P^{DGDR}, of the model parameter vector as initialization for the UKF method led to two important benefits. On the one hand, it reduced the time required for the estimates to reach convergence, in turn diminishing the computational cost. On the other hand, it led to more accurate estimates, as shown in Fig. 3 for the estimation of θ_{Na} in one of the models of the population at baseline conditions. While for the individual UKF method more than 40 beats were required for the estimation error to be below 0.04, when the combined UKF+INI method was used the number of required beats was 5 for that particular example.
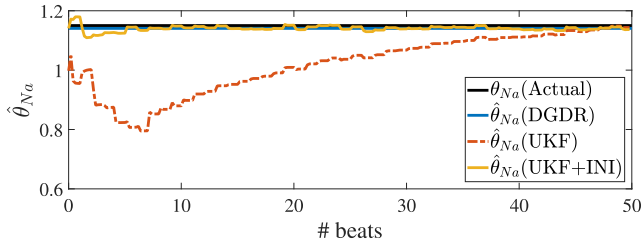
Fig. 3.   Example of actual θ_{Na} value and time course of $\hat{\theta}_{Na}$ as estimated by DGDR, UKF and UKF+INI methods for a virtual cell at baseline.
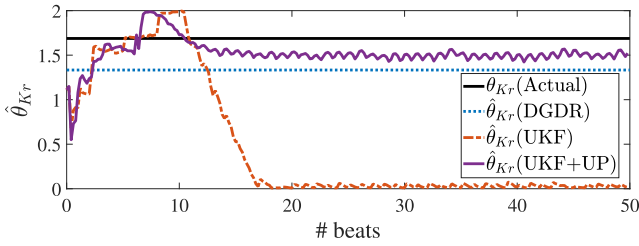

Fig. 4.   Example of actual θ_{Kr} value and time course of $\hat{\theta}_{Kr}$ as estimated by DGDR, UKF and UKF+UP methods for a virtual cell at baseline.

C. Combined DGDR and UKF Methods: Updating Effects

The use of the estimates obtained by DGDR for the mean, $\hat{\theta}^{DGDR}$, and the covariance matrix, P^{DGDR}, of the model parameter vector to update the UKF estimation at the end of each beat helped to retain the parameter estimates close to the actual parameter values and to reduce the uncertainty in the estimation, as confirmed by a reduction in the estimation covariance matrix. Additionally, this UKF+UP approach diminished the convergence time and, thus, the associated computational cost. The benefit of using the DGDR-derived estimates for UKF updating is illustrated in Fig. 4 for the estimation of θ_{Kr} in one of the models of the population at baseline conditions. When only UKF is employed, the parameter estimates may fall in a local minimum and may never reach a value close to the actual one. As can be observed from the figure, the UKF and UKF+UP estimates were the same for the first beat whereas the updating subsequently led to remarkably enhanced results.

D. Performance Comparison

The performances of the individual DGDR and UKF methods and their combinations, either by initialization and/or updating, were assessed in terms of the average mean $E[\bar{\eta}_\theta]$ and average standard deviation $E[\bar{\sigma}_{\eta_\theta}]$ of the absolute error. Top panel of Fig. 5 illustrates $E[\bar{\eta}_\theta]$ for the five evaluated methods at baseline conditions. As can be seen from the figure, the individual DGDR and UKF methods led to approximately the same level of error ($E[\bar{\sigma}_{\eta_\theta}]$ values of 0.1806 and 0.1775, respectively), with a larger associated computational cost in the case of the UKF method. The combination of DGDR and UKF remarkably
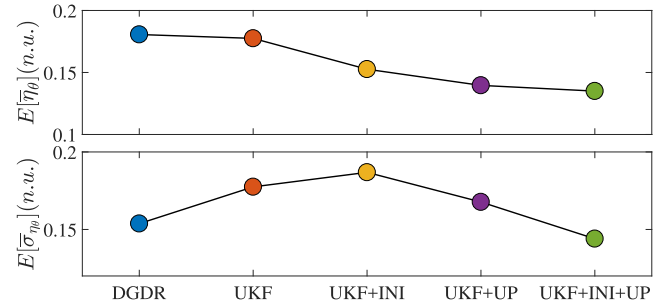

Fig. 5.   Average over the validation population at baseline of mean (top panel) and standard deviation (bottom panel) of absolute parameter estimation error $\bar{\eta}_\theta$ for the five evaluated methods.

improved the estimation performance, either when combined through initialization or through update and, particularly, when combined through both ($E[\bar{\sigma}_{\eta_\theta}]$ values of 0.1526 for UKF+INI, 0.1396 for UKF+UP and 0.1350 for UKF+INI+UP). Bottom panel of Fig. 5 presents the estimation uncertainty for the five evaluated methods. As can be observed, initialization and updating by DGDR contributed to reduce the parameter estimation uncertainty of the UKF method.

Fig. 6 shows boxplots for the mean absolute error in the estimation of each ionic conductance factor by each of the five evaluated methods at baseline conditions. As can be observed from the figure, the combined UKF+INI+UP method presents better performance than the individual DGDR and UKF methods for practically all estimated factors. The most accurate results were obtained for θ_{Na}, with median estimation errors $\hat{\theta}_{Na}$ being lower than 0.05. On the other hand, the least accurate results were obtained for θ_{Ks}, θ_{NaCa} and θ_{NaK}.

Fig. 7 presents results related to estimation uncertainty. Fig. 7, left panel, illustrates the time course of the estimation uncertainty quantified by the square root of the covariance matrix $\sqrt{P_{NaK}}$ in the estimation of θ_{NaK} for one virtual cell in the population at baseline conditions. As can be observed from the figure, the combination of DGDR and UKF presented lower uncertainty than the individual DGDR and UKF methods, with the impact of updating being notably larger than that of initialization. Fig. 7, right panel, provides an additional characterization of the estimation uncertainty quantified by the number of beats required by each UKF-based method to reach the same value of the averaged standard deviation of the absolute estimation errors as the individual UKF. The impact of updating on the reduction of the estimation uncertainty is clear from this figure too.

E. Replication of AP Traces and Biomarkers at Baseline

The performance of the five proposed methods to replicate AP traces at baseline conditions was assessed by generating APs from the ORd model with the different sets of estimated parameters and by comparing them with the input AP traces. Also, the comparison was established in terms of AP-derived biomarkers like APD and STV. Fig. 8, left panel, shows the probability density function of the differences between the APD
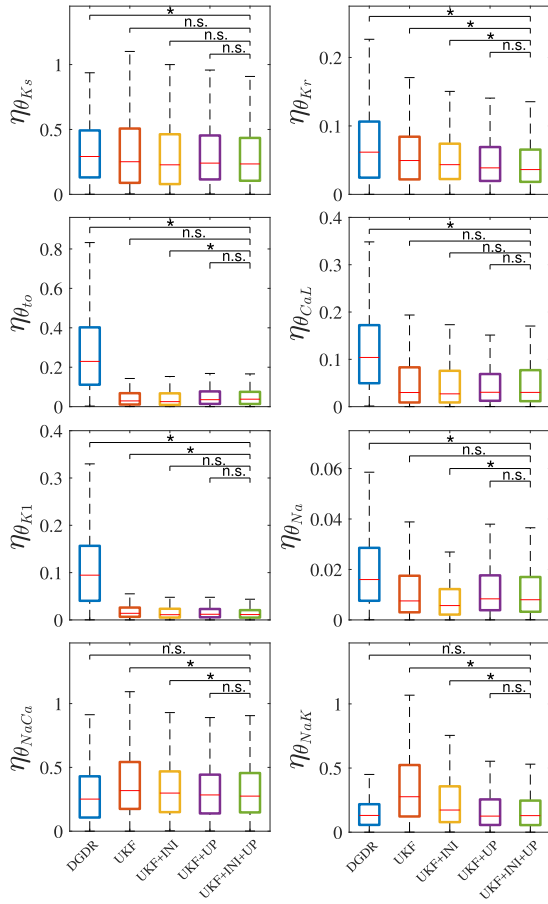
Fig. 6. Boxplots of absolute estimation errors η_θ for the factors multiplying ionic current conductances calculated for the five evaluated methods. Statistically significant differences by Wilcoxon signed-rank test (p-value < 0.05) are denoted by $*$, while non-significant differences are denoted by $n.s$, for a number of cells equal to 373.
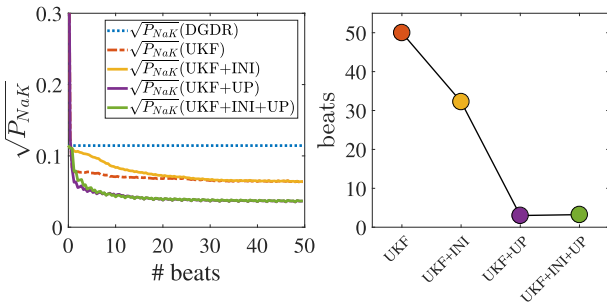

Fig. 7. Left panel: Time course of estimation uncertainty in terms of square root of covariance matrix $\sqrt{P_{NaK}}$ for each of the five evaluated methods. Right panel: Number of beats required by each evaluated method to reach the same level of accuracy as the UKF method, as quantified by the averaged covariance over all estimated model parameters.

from the input AP trace and the APD calculated from the estimated AP trace for DGDR, UKF and UKF+INI+UP. Similarly, Fig. 8, right panel, shows results for STV. As can be observed from the figure, the combined UKF+INI+UP method provides the best fitting to the actual data, as confirmed by the fact that the distributions of ΔAPD and ΔSTV are more concentrated
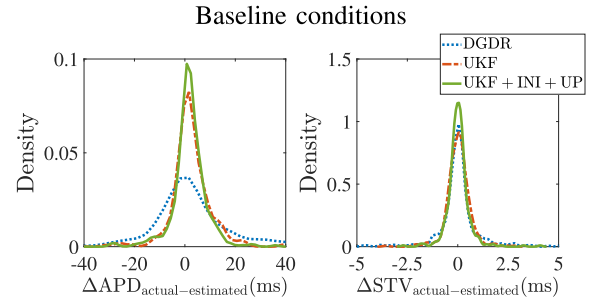

Fig. 8. Probability density function of ΔAPD (left panel) and ΔSTV (right panel) for the validation population, with ΔAPD (ΔSTV, respectively) calculated as the difference between APD (STV, respectively) from the input AP trace and APD (STV, respectively) from the estimated AP trace for each evaluated method under baseline conditions.
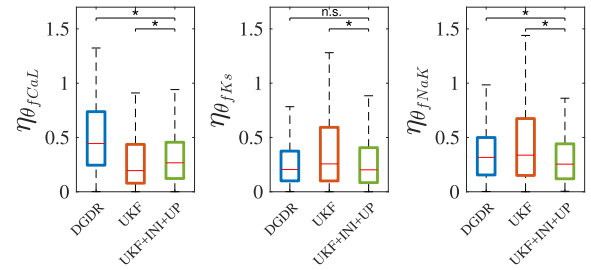

Fig. 9. Boxplots of absolute estimation errors η_θ for the factors multiplying ISO-induced phosphorylation levels calculated for three evaluated methods. Statistically significant differences by Wilcoxon signed-rank test (p-value < 0.05) are denoted by $*$, while non-significant differences are denoted by $n.s$, for a number of cells equal to 373.

around 0. On the other hand, the DGDR method presents reduced accuracy for APD estimation, although very similar to UKF and UKF+INI+UP for STV estimation.

F. Estimation of Phosphorylation Factors, AP Traces and Biomarkers Under β-AS

Considering the ionic conductance estimates obtained for baseline conditions, the next step was to test the performance of DGDR, UKF and UKF+INI+UP to estimate the phosphorylation levels for the validation population of models under β-AS effects. Fig. 9 shows boxplots of the mean absolute errors $\overline{\eta}_\theta$ for the estimation of the three ISO-induced phosphorylation levels. As can be observed from the figure, the UKF+INI+UP method increased the accuracy in the estimation of θ_{fKs} and θ_{fNaK} with respect to the individual DGDR and UKF methods, whereas for θ_{fCaL} UKF was slightly better in terms of median absolute error, but not in terms of averaged absolute error ($\overline{\eta}_\theta = 0.34$ for both methodologies). Taking together the three estimated factors for the phosphorylation levels and results over the whole validation population, the combined UKF+INI+UP method led to a reduction in the averaged mean absolute error, $E[\overline{\eta}_\theta]$, of 15.29% and 20.01% with respect to the individual use of DGDR and UKF, respectively. The average mean absolute errors, $E[\overline{\eta}_\theta]$, for ISO-induced phosphorylation level factors were higher (0.38, 0.40 and 0.32 for DGDR, UKF and combination respectively) than those obtained for ionic conductance factors due to the fact
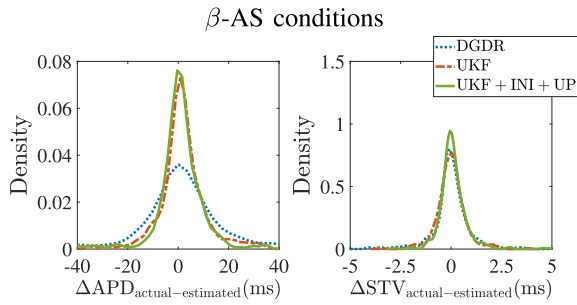
Fig. 10. Probability density function of ΔAPD (left panel) and ΔSTV (right panel) for the validation population, with ΔAPD (ΔSTV, respectively) calculated as the difference between APD (STV, respectively) from the input AP trace and APD (STV, respectively) from the estimated AP trace for each evaluated method under β-AS conditions.
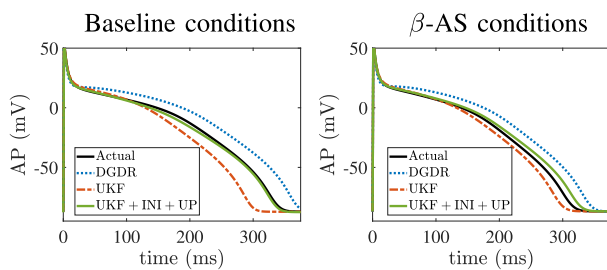

Fig. 11. Actual and estimated APs (mean over 100 beats) calculated from the set of estimated parameters by each evaluated method at baseline (left panel) and under β-AS (right panel) for one of the virtual cells in the validation population.

that the error in the ionic conductance estimation was propagated into the phosphorylation level estimation.

Fig. 10, left panel, shows the probability density function of the differences between the APD from the input AP trace and the APD calculated from the estimated AP trace after estimation of the ionic conductances at baseline and phosphorylation factors under β-AS for DGDR, UKF and UKF-INI-UP. Fig. 10, right panel, shows analogous results for STV. Again, the combined UKF-INI-UP provided the best fitting for both ΔAPD and ΔSTV, whereas the DGDR method presented the highest differences between actual and estimated APD and comparable performance to UKF and UKF+INI+UP in the case of STV.

As an illustration of the above results, Fig. 11 shows the actual and estimated APs (mean over 100 beats) calculated from the set of estimated parameters by each of the evaluated methods for a cell in the validation population. Both at baseline and under β-AS, the AP estimated by DGDR+UKF remarkably better matched the actual AP as compared to those obtained by DGDR or UKF individually. Not only the mean AP, but also the variability over 100 beats was better reproduced by DGDR+UKF as compared to DGDR and UKF.

IV. DISCUSSION

A novel approach based on the combined use of the Double Greedy Dimension Reduction (DGDR), with automatic selection of biomarkers, and the Unscented Kalman Filter (UKF)

has been proposed as a method for joint estimation of parameters and state variables of computational human ventricular stochastic models from given input AP traces. By using this combined methodology, different sets of ionic parameters, namely ionic current conductances and phosphorylation levels of cellular substrates, are estimated for each given individual AP trace at baseline conditions and following β-AS. The use of complete AP traces and a collection of AP-based biomarkers contributes to the identifiability of the model parameters and state variables in the estimation problem. The proposed methodology outperforms individual DGDR and UKF methods and has an affordable computational cost. It allows realistic characterization of spatio-temporal variability at baseline and following β-AS, thus enabling improved investigation of variability mechanisms and arrhythmic risk prediction. This can prove fundamental to assess the role of β-AS in leading to exaggerated increases in BVR that facilitate the occurrence of arrhythmic events in certain cases but not in others [27]. In the following, relevant characteristics of the proposed methodology as well as major benefits and shortcomings associated with its use are discussed.

A. DGDR Method

The DGDR method was used to obtain estimates for the model parameters, which were subsequently fed to the UKF method to build the combined DGDR-UKF method. Methods similar to DGDR have been used for other applications, although using strategies different from those used here.

A key factor in the performance of the DGDR method involves a correct training phase. To obtain high levels of estimation accuracy, training should be performed over large populations, which in the case of this study corresponds to a large set of synthetic AP traces. Confirmation on the appropriateness of the training population dimension was provided by the fact that similar estimation errors were attained in both the training and validation populations. If training dimension had not been sufficient, estimation uncertainty in the validation population would have been much greater than that obtained in the training population. The time required to obtain the estimation dictionaries from the training population was just three hours, being subsequent calculation of parameter estimates immediate (scalar product of two vectors) when given a new AP trace of the validation population.

B. UKF Method

After formulating the estimation problem as a nonlinear state-space representation where a noisy voltage trace is considered as the observed variable and SDEs defining a human ventricular cell model are used to describe the process equations, the UKF method was applied for joint model parameter and state variables, providing not only mean estimates but also measurements of estimation uncertainty. UKF has been used in many other studies, but not with the same purpose as here.

When using the UKF, appropriate calibration of its hyperparameters σ_θ and σ_r, representing process and measurement variances, respectively, is a critical point to achieve high levels of accuracy. According to our results, an inadequate selection of

these hyperparameters may lead to an increase in the estimation error above 50% of the value attained for optimally adjusted σ_θ and σ_r values. Based on a previous work of ours [33], σ_r was set to 1 mV, equal to the variance of the measurement noise added to the clean synthetic AP signal. In the case of σ_θ, which is closely related to the convergence speed and potentially oscillatory behavior of the estimates, its value was set to $\sigma_\theta = 10^{-8}$, as this value led to a minimum average mean absolute error in parameter estimation, as shown in Section III-A. This value is in the range of feasible values shown in our previous work [33], with a slight difference in the optimal value justified by the fact that a higher number of model parameters were estimated in the present study as well as to the fact that a subunit-based formulation of SDEs for ionic gates, rather than the channel-based formulation used in [33], was here employed.

C. Combined DGDR-UKF Method by Initialization and Updating

The use of DGDR estimates for both initialization and updating of the UKF estimates has been demonstrated to play a very significant role in improving the estimation performance. To the best of our knowledge, this is the first time these two methods are combined. On the one hand, providing an initialization for the UKF method in terms of its mean and covariance matrix based on DGDR estimates allowed reducing the mean estimation error and the uncertainty around it. Also, the convergence time was remarkably diminished, as described in Section III-B. As a proof, the combined DGDR initialization + UKF approach required approximately 35% of the number of beats than the individual UKF method to reach the same level of estimation uncertainty.

On the other hand, updating the UKF estimates at the end of each cardiac beat by using the DGDR estimates allowed the solution of the combined method to remain within a relatively narrow range around the actual parameter values and avoided the estimation to fall into local minima. In addition, it contributed to accelerate estimation convergence, reducing by more than 95% the number of beats required by the UKF method to reach the same level of uncertainty. It is interesting to highlight that this updating process improved the estimation of not only the mean and covariance of the model parameters, but also of all other model state variables.

The combined DGDR-UKF method provides relevant advantages as compared to other methods used in the literature for similar purposes. The DGDR-UKF method renders a one-to-one matching between input AP traces and the sets of estimated parameters, whereas other methods based on Genetic Algorithms, Moment-Matching or Gaussian Process Emulators provide only parameter estimates at a population level [28]–[30]. In addition, when comparing the DGDR-UKF method with other methods rendering individual parameter estimates, such as Markov Chain Monte Carlo (MCMC)-based methods [31], [32], the DGDR-UKF method presents lower computational cost. Also, it is able to deal with beat-to-beat variability and to provide estimates of not only the model parameters but also of the hidden state variables, thus improving the global estimation accuracy.

The enhanced performance and reduced convergence time attained by the combined DGDR-UKF method are particulary relevant for subsequent studies aimed at investigating repolarization variability from human ventricular experimental voltage traces, which are commonly of short duration.

D. Estimation of Ionic Current Conductances at Baseline

Eight ionic current conductances were estimated at baseline conditions, as variations in those conductances have been postulated to be major factors for spatial (cell-to-cell) AP variability [8], [9], [11], [12], [28]. Other studies in the literature have addressed estimation of ionic current conductances, even if not in all cases for as many currents as in this work and not always considering temporal (beat-to-beat) AP changes but just focusing on a steady-state AP [29], [31], [33]. In the present study, stochastic human ventricular cell models accounting for temporal variability were developed to improve the estimation accuracy by considering dynamic information additional to the static information commonly considered in the literature. The eight estimated model parameters were multiplying factors for the conductances of six major ionic currents (I_{Ks}, I_{Kr}, I_{to}, I_{CaL}, I_{K1}, I_{Na}) and the maximal values of I_{NaCa} and I_{NaK} with respect to their nominal values in the ORd model.

The least accurate results with our DGDR-UKF method, as well as with all other tested methodologies (individual DGDR and UKF methods and UKF with only initialization or updating from DGDR), were obtained for θ_{Ks}, in line with results reported in [33]. This can be due to the intrinsic characteristics of the ORd model, in which the I_{Ks} current has little influence on the AP, and consequently on AP-derived biomarkers, at baseline conditions. Other experimental and computational studies support this outcome regarding the limited influence of I_{Ks} on the AP shape and duration at baseline [40], [53], [54]. Since a wide range of θ_{Ks} values generate little differences in the corresponding AP traces, accurate identification becomes challenging. This issue is framed within the context of identifiability and observability and may be solved in future studies by complementing the estimation process with signals obtained while stimulating the cells at other pacing frequencies or under ionic current blocks. Similarly, the estimation errors associated with θ_{NaCa} and θ_{NaK} were among the highest for all tested methodologies, which can in this case be due to the longer time scale required for I_{NaCa} and I_{NaK} variations to impact the AP.

Of note, estimation of θ_{to} rendered much higher errors when the DGDR method was used as compared with any of the other methods involving UKF. This can be attributed to the fact that none of the defined AP-derived biomarkers may be closely related to the AP notch, which is the AP phase where this current has the largest influence. Similarly happened with θ_{K1}, for which estimation errors where higher for DGDR than for any UKF-based method. In this case, despite considering biomarkers in the DGDR method like the resting membrane potential, which are expected to contribute to θ_{K1} identification, the UKF-based methods can deliver more accurate results because they use all samples of the AP trace, both during the AP as well as during

the resting phase, and thus have more information to adjust θ_{K1} estimation.

E. Estimation of Phosphorylation Levels of Cellular Substrates Under β-AS Conditions

The phosphorylation levels corresponding to the three cellular substrates most significantly contributing to AP changes under β-AS, as determined from performed computational tests, were estimated using our proposed DGDR-UKF method and compared with other tested methods. To the best of our knowledge, this is the first study where the phosphorylation levels of a β-adrenergic signaling model have been estimated, together with other state variables, based on the static and dynamic AP changes induced by β-AS. The results obtained with our proposed combined method were generally better than those of individual DGDR and UKF methods. Nevertheless, it should be noted that the average mean absolute errors obtained for phosphorylation levels under β-AS were higher than those obtained for ionic conductances at baseline. This can be partly explained because the errors in the estimated baseline conductances were propagated to the estimation of the phosphorylation levels, as the latter were calculated based on the corresponding APs estimated at baseline.

Although simultaneous estimation of ionic conductances and phosphorylation levels under β-AS could be thought of, this turns out to be unfeasible due to the multiplicative relation of ionic conductances and phosphorylation levels in the coupled electrophysiological-adrenergic signaling model. On the basis of such multiplicative relation, many combinations of conductance and phosphorylation level values could lead to the same estimation results even if the estimated parameter values were in fact far from their actual values.

F. Characterization of Spatio-Temporal AP Variability From Parameter Estimates

It is a main purpose of this study to propose a method suitable for investigation of temporal and spatial variability in human ventricular repolarization, with one-to-one identification of an underlying computational AP model for each experimentally available voltage trace. Provided data is available at baseline and under β-AS conditions, our proposed DGDR-UKF method can identify the specific electrophysiological and adrenergic signaling characteristics at those two conditions. Our method was indeed able to precisely reproduce the AP shape, duration and variability of individual AP traces, rendering statistical distributions of the errors in the estimation of APD and STV remarkably more concentrated around 0 than those obtained with other tested methods, particularly when comparing with the DGDR method.

On top of the DGDR-UKF method rendering better match between actual and estimated AP-derived biomarkers than other methods, it led to improved match between actual and estimated voltage traces, as illustrated in Section III-F. This can be justified on the basis that our methodology provides estimates of not only the parameter values but of the complete vector of model state variables, which allows for more accurate AP reconstruction.

G. Limitations and Future Studies

In this work a total of 11 different human ventricular cell model parameters have been identified, corresponding to 8 ionic current conductances at baseline and 3 phosphorylation levels under β-AS. Future studies could include estimation of additional ionic currents conductances (e.g. for I_{Cab}, I_{Nab}, I_{Kb} or I_{pCa}), phosphorylation levels (e.g. for ryanodine receptors, phospholamban or troponin I) or time constants of ionic gates (e.g τ_{xrs}, τ_{xs1} or τ_{xk1}). Also, stochasticity could be added to other ionic currents like the late sodium current, which can have a relevant contribution to BVR.

To test the performance of our proposed methodology for estimation of model parameters and one-to-one replication of AP traces and AP-derived biomarkers, synthetic voltage traces were generated at 1 Hz stimulation frequency. Future studies could test the extent to which the estimation performance is improved by applying the proposed DGDR-UKF method onto voltage traces obtained at different stimulation frequencies. In addition, voltage traces could be generated under different ionic blocks to offer additional information to be used for parameter identification, which could prove particularly useful for identification of θ_{Ks}, θ_{NaCa}, θ_{NaK}, whose estimation was the most challenging in the present work.

A set of AP-derived biomarkers were used in the DGDR method and, consequently, in the DGDR-UKF method. Those biomarkers reflect AP characteristics related to its upstroke, repolarization and resting potential as well as temporal APD variability. Novel AP-derived biomarkers reflecting additional information from the AP notch and plateau phases could help in the identification of model parameters, like θ_{to} and θ_{CaL}, thus globally improving the performance of the DGDR method and of the combined DGDR-UKF method.

This study has presented the combined DGDR-UKF method and has assessed its performance over a large set of synthetically generated AP traces. As a next step, the proposed method could be tested over experimental AP traces recorded from human ventricular cardiomyocytes or even extend the method to be applied onto voltage traces measured from human ventricular tissues. This would allow identification of underlying computational tissue models with representation of cell-to-cell electrical coupling.

V. CONCLUSION

A novel methodology based on the combined use of Double Greedy Dimension Reduction (DGDR), with automatic selection of biomarkers, and the Unscented Kalman Filter (UKF) has been proposed to estimate parameters and state variables of an underlying human ventricular action potential (AP) model for any given input voltage trace. The proposed methodology is tested over synthetic voltage traces generated from an experimentally-calibrated population of stochastic human ventricular cell models at baseline and under β-adrenergic stimulation. The combined methodology remarkably improves the estimation performance of individual DGDR and UKF methods while reducing the computational cost. The estimated ionic current conductances at baseline conditions and phosphorylation

levels of cellular substrates under β-adrenergic stimulation allow for computational characterization of spatio-temporal ventricular repolarization, which can prove very useful to investigate variability changes induced by disease or drugs, uncover its underlying ionic mechanisms and establish a relationship with arrhythmic risk.

REFERENCES

[1] J. Magyar *et al.*, "Concept of relative variability of cardiac action potential duration and its test under various experimental conditions," *General Physiol. Biophys.*, vol. 35, no. 1, pp. 55–62, 2016.

[2] D. M. Johnson *et al.*, "Diastolic spontaneous calcium release from the sarcoplasmic reticulum increases beat-to-beat variability of repolarization in canine ventricular myocytes after β-adrenergic stimulation," *Circulation Res.*, vol. 112, no. 2, pp. 246–256, 2013.

[3] M. Zaniboni, F. Cacciani, and N. Salvarani, "Temporal variability of repolarization in rat ventricular myocytes paced with time-varying frequencies," *Exp. Physiol.*, vol. 92, no. 5, pp. 859–869, 2007.

[4] E. Pueyo *et al.*, "Experimentally-based computational investigation into beat-to-beat variability in ventricular repolarization and its response to ionic current inhibition," *PloS One*, vol. 11, no. 3, 2016, Art. no. e0151461.

[5] C. Antzelevitch *et al.*, "Heterogeneity within the ventricular wall. Electrophysiology and pharmacology of epicardial, endocardial, and M cells," *Circulation Res.*, vol. 69, no. 6, pp. 1427–1449, 1991.

[6] J. Ramírez *et al.*, "T-wave morphology restitution predicts sudden cardiac death in patients with chronic heart failure," *J. Am. Heart Assoc.*, vol. 6, no. 5, 2017, Art. no. e005310.

[7] M. Baumert *et al.*, "QT interval variability in body surface ECG: Measurement, physiological basis, and clinical value: Position statement and consensus guidance endorsed by the European Heart Rhythm Association jointly with the ESC Working Group on Cardiac Cellular Electrophysiology," *Europace*, vol. 18, no. 6, pp. 925–944, 2016.

[8] E. Pueyo *et al.*, "A multiscale investigation of repolarization variability and its role in cardiac arrhythmogenesis," *Biophys. J.*, vol. 101, no. 12, pp. 2892–2902, 2011.

[9] N. Szentandrássy *et al.*, "Contribution of ion currents to beat-to-beat variability of action potential duration in canine ventricular myocytes," *Pfluegers Arch.-Eur. J. Physiol.*, vol. 467, no. 7, pp. 1431–1443, 2015.

[10] C. Lerma, T. Krogh-Madsen, M. Guevara, and L. Glass, "Stochastic aspects of cardiac arrhythmias," *J. Statistical Phys.*, vol. 128, no. 1-2, pp. 347–374, 2007.

[11] G.-R. Li, J. Feng, L. Yue, and M. Carrier, "Transmural heterogeneity of action potentials and Ito1 in myocytes isolated from the human right ventricle," *Am. J. Physiol.: Heart Circulatory Physiol.*, vol. 275, no. 2, pp. H369–H377, 1998.

[12] C. Sánchez *et al.*, "Inter-subject variability in human atrial action potential in sinus rhythm versus chronic atrial fibrillation," *PLOS ONE*, vol. 9, no. 8, 2014, Art. no. e105897.

[13] K. Kistamas *et al.*, "Changes in intracellular calcium concentration influence beat-to-beat variability of action potential duration in canine ventricular myocytes," *J. Physiol. Pharmacol.*, vol. 66, no. 1, pp. 73–81, 2015.

[14] M. Lemay, E. de Lange, and J. P. Kucera, "Effects of stochastic channel gating and distribution on the cardiac action potential," *J. Theor. Biol.*, vol. 281, no. 1, pp. 84–96, 2011.

[15] G. Antoons *et al.*, "Calcium release near L-type calcium channels promotes beat-to-beat variability in ventricular myocytes from the chronic AV block dog," *J. Mol. Cellular Cardiol.*, vol. 89, pp. 326–334, 2015.

[16] A. J. Tanskanen, J. L. Greenstein, B. O'Rourke, and R. L. Winslow, "The role of stochastic and modal gating of cardiac L-type Ca^{2+} channels on early after-depolarizations," *Biophys. J.*, vol. 88, no. 1, pp. 85–95, 2005.

[17] M. Hinterseer *et al.*, "Usefulness of short-term variability of QT intervals as a predictor for electrical remodeling and proarrhythmia in patients with nonischemic heart failure," *Am. J. Cardiol.*, vol. 106, no. 2, pp. 216–220, 2010.

[18] R. Varkevisser, S. C. Wijers, M. A. van der Heyden, J. D. Beekman, M. Meine, and M. A. Vos, "Beat-to-beat variability of repolarization as a new biomarker for proarrhythmia in vivo," *Heart Rhythm*, vol. 9, no. 10, pp. 1718–1726, 2012.

[19] Y. Sobue *et al.*, "Beat-to-beat variability of T-wave amplitude for the risk assessment of ventricular tachyarrhythmia in patients without structural heart disease," *Europace*, vol. 13, no. 11, pp. 1612–1618, 2011.

[20] J. W. Waks *et al.*, "Beat-to-beat spatiotemporal variability in the T-vector is associated with sudden cardiac death in participants without left ventricular hypertrophy: The atherosclerosis risk in communities (ARIC) Study," *J. Am. Heart Assoc.*, vol. 4, no. 1, 2015, Art. no. e001357.

[21] D. A. Sampedro-Puente, J. Fernandez-Bes, N. Szentandrássy, P. Nánási, P. Taggart, and E. Pueyo, "Time course of low-frequency oscillatory behavior in human ventricular repolarization following enhanced sympathetic activity and relation to arrhythmogenesis," *Frontiers Physiol.*, vol. 10, p. 1547, 2019.

[22] D. M. Johnson *et al.*, "IKs restricts excessive beat-to-beat variability of repolarization during beta-adrenergic receptor stimulation," *J. Mol. Cellular Cardiol.*, vol. 48, no. 1, pp. 122–130, 2010.

[23] D. J. Gallacher *et al.*, "In vivo mechanisms precipitating Torsades de Pointes in a canine model of drug-induced long-QT1 syndrome," *Cardiovascular Res.*, vol. 76, no. 2, pp. 247–256, 2007.

[24] B. Hegyi, T. Bányász, L. T. Izu, L. Belardinelli, D. M. Bers, and Y. Chen-Izu, "β-adrenergic regulation of late Na^+ current during cardiac action potential is mediated by both PKA and CaMKII," *J. Mol. Cellular Cardiol.*, vol. 123, pp. 168–179, 2018.

[25] J. Heijman *et al.*, "Determinants of beat-to-beat variability of repolarization duration in the canine ventricular myocyte: A computational analysis," *PLOS Comput. Biol.*, vol. 9, no. 8, 2013, Art. no. e1003202.

[26] E. Pueyo, M. Orini, J. F. Rodríguez, and P. Taggart, "Interactive effect of beta-adrenergic stimulation and mechanical stretch on low-frequency oscillations of ventricular action potential duration in humans," *J. Mol. Cellular Cardiol.*, vol. 97, pp. 93–105, 2016.

[27] D. A. Sampedro-Puente, J. Fernandez-Bes, B. Porter, S. van Duijvenboden, P. Taggart, and E. Pueyo, "Mechanisms underlying interactions between low-frequency oscillations and beat-to-beat variability of cellular ventricular repolarization in response to sympathetic stimulation: Implications for arrhythmogenesis," *Frontiers Physiol.*, vol. 10, p. 916, 2019.

[28] B. A. J. Lawson, C. C. Drovandi, N. Cusimano, P. Burrage, B. Rodriguez, and K. Burrage, "Unlocking data sets by calibrating populations of models to data density: A study in atrial electrophysiology," *Sci. Adv.*, vol. 4, no. 1, 2018, Art. no. e1701676.

[29] E. Tixier, D. Lombardi, B. Rodriguez, and J.-F. Gerbeau, "Modelling variability in cardiac electrophysiology: A moment-matching approach," *J. Roy. Soc., Interface*, vol. 14, no. 133, 2017, Art. no. 20170238.

[30] E. T. Chang, M. Strong, and R. H. Clayton, "Bayesian sensitivity analysis of a cardiac cell model using a Gaussian Process emulator," *PLOS One*, vol. 10, no. 6, 2015, Art. no. e0130252.

[31] R. H. Johnstone *et al.*, "Uncertainty and variability in models of the cardiac action potential: Can we build trustworthy models?" *J. Mol. Cellular Cardiol.*, vol. 96, pp. 49–62, 2016.

[32] R. H. Johnstone, R. Bardenet, D. J. Gavaghan, L. Polonchuk, M. R. Davies, and G. R. Mirams, "Hierarchical Bayesian modeling of variability and uncertainty in synthetic action potential traces," *Comput. Cardiol.*, vol. 43, pp. 1089–1092, 2017.

[33] D. A. Sampedro-Puente, J. Fernandez-Bes, and E. Pueyo, "Data-driven identification of stochastic model parameters and state variables: Application to the study of cardiac beat-to-beat variability," *IEEE J. Biomed. Health Informat.*, no. 5, pp. 1697–1722, 2019.

[34] S. Särkkä, *Bayesian Filtering and Smoothing*. vol. 3. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[35] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[36] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*. vol. 2045. Princeton, NJ, USA: Princeton Univ. Press, 2015.

[37] D. Lombardi and F. Raphel, "A greedy dimension reduction method for classification problems," *HAL*, 2019. [Online]. Avalable: http://hal.inria.fr/hal-02280502/2019

[38] D. M. Hawkins, "The problem of overfitting," *J. Chem. Inf. Comput. Sci.*, vol. 44, no. 1, pp. 1–12, 2004.

[39] T. O'Hara, L. Virág, A. Varró, and Y. Rudy, "Simulation of the undiseased human cardiac ventricular action potential: Model formulation and experimental validation," *PLOS Comput. Biol.*, vol. 7, no. 5, 2011, Art. no. e1002061.

[40] Y. Xie, E. Grandi, J. L. Puglisi, D. Sato, and D. M. Bers, "β-adrenergic stimulation activates early afterdepolarizations transiently via kinetic mismatch of PKA targets," *J. Mol. Cellular Cardiol.*, vol. 58, pp. 153–161, 2013.

[41] A. R. Soltis and J. J. Saucerman, "Synergy between CaMKII substrates and β-adrenergic signaling in regulation of cardiac myocyte Ca^{2+} handling," *Biophys. J.*, vol. 99, no. 7, pp. 2038–2047, 2010.

[42] D. Guo *et al.*, "Electrophysiological properties of HBI-3000: A new antiarrhythmic agent with multiple-channel blocking properties in human ventricular myocytes," *J. Cardiovascular Pharmacology*, vol. 57, no. 1, pp. 79–85, 2011.

[43] O. J. Britton, A. Bueno-Orovio, L. Virág, A. Varró, and B. Rodriguez, "The electrogenic Na+/K+ pump is a key determinant of repolarization abnormality susceptibility in human ventricular cardiomyocytes: A population-based simulation study," *Frontiers Physiol.*, vol. 8, 2017. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphys.2017.00278/full

[44] E. Grandi, F. S. Pasqualini, and D. M. Bers, "A novel computational model of the human ventricular action potential and Ca transient," *J. Mol. Cellular Cardiology*, vol. 48, no. 1, pp. 112–121, 2010.

[45] N. Jost *et al.*, "Molecular basis of repolarization reserve differences between dogs and man," *Circulation*, vol. 118, no. 18, p. S342, 2008.

[46] B. Pieske, L. S. Maier, V. Piacentino, J. Weisser, G. Hasenfuss, and S. Houser, "Rate dependence of $[Na_i^+]$ and contractility in nonfailing and failing human myocardium," *Circulation*, vol. 106, no. 4, pp. 447–453, 2002.

[47] U. Schmidt, R. J. Hajjar, P. A. Helm, C. S. Kim, A. A. Doye, and J. K. Gwathmey, "Contribution of abnormal sarcoplasmic reticulum ATPase activity to systolic and diastolic dysfunction in human heart failure," *J. Mol. Cellular Cardiology*, vol. 30, pp. 1929–1937, 1998.

[48] T. Furukawa, S. Kimura, N. Furukawa, A. L. Bassett, and R. J. Myerburg, "Potassium rectifier currents differ in myocytes of endocardial and epicardial origin," *Circulation Res.*, vol. 70, pp. 91–103, 1992.

[49] J. A. Negroni *et al.*, "β-adrenergic effects on cardiac myofilaments and contraction in an integrated rabbit ventricular myocyte model," *J. Mol. Cellular Cardiology*, vol. 81, pp. 162–175, 2015.

[50] J. Heijman, P. G. Volders, R. L. Westra, and Y. Rudy, "Local control of β-adrenergic stimulation: Effects on ventricular myocyte electrophysiology and Ca^{2+}-transient," *J. Mol. Cellular Cardiology*, vol. 50, no. 5, pp. 863–871, 2011.

[51] E. A. Wan and R. van der Merwe, "The Unscented Kalman Filter," in *Kalman Filtering and Neural Networks*. Hoboken, NJ, USA: Wiley, 2001, pp. 221–280.

[52] C. Bertoglio, P. Moireau, and J.-F. Gerbeau, "Sequential parameter estimation for fluid-structure problems: Application to hemodynamics," *Int. J. Numer. Methods Biomed. Eng.*, vol. 28, no. 4, pp. 434–455, 2012.

[53] N. Jost *et al.*, "Restricting excessive cardiac action potential and QT prolongation: A vital role for I_{Ks} in human ventricular muscle," *Circulation*, vol. 112, no. 10, pp. 1392–1399, 2005.

[54] T. O'Hara and Y. Rudy, "Quantitative comparison of cardiac ventricular myocyte electrophysiology and response to drugs in human and nonhuman species." *Am. J. Physiol.: Heart Circulatory Physiol.*, vol. 302, no. 5, pp. H1023–30, 2012.