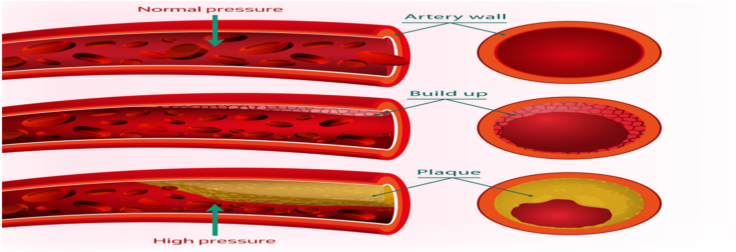
Physics–informed Statistical Machine Learning for sensor-acquired big data

Michail Spitieris

NTNU

March 18, 2019

Hypertension (High Blood Pressure)



- One of the most important sources of mortality.
- Usually does not cause symptoms.
- Population based diagnosis and treatment are ineffective.
- Haemodynamic pattern may be critical in understanding, diagnosis and treatment of hypertensive patients.
- Physical activity–Positive effect on arterial compliance subsequently to hypertension.

My Medical Digital Twin (MyMDT)

Aim

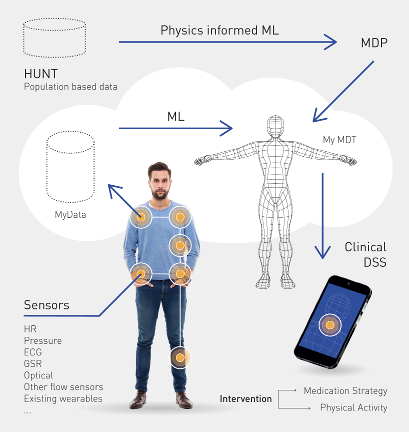
Develop a digital platform for personalized diagnosis, prognosis and treatment of essential hypertension.

Medical Digital Prototype

Generic Model of the Cardiovascular system based on population data (e.g. HUNT study)

Individual Sensor Data

Data Streams from wearable sensors



Minimal Cardiovascular Model

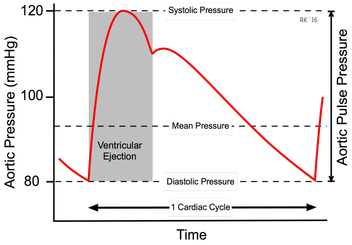
The following model gives the arterial blood pressure waveform

$$I(t) = \frac{P(t)}{R} + C\frac{dP(t)}{dt}$$

$$\boxed{\textbf{I}\text{–Blood Flow}} \xrightarrow[\textbf{R}\text{–total peripheral Resistance}]{\textbf{C}\text{–arterial Compliance}} \boxed{\textbf{P}\text{–Blood Pressure}}$$

Unknown Model Parameters

C and R



Uncertainties in Computer Modelling

Computer Models

Complex deterministic mathematical models–implemented in computer code.

- Cardiovascular model – Physical exercise model.
- Parameters with physical meaning should be specified. (e.g. Resistance over the aortic valve)
- Observations of the real process are used to estimate these pareameters Calibration

Uncertainties

Parameter Uncertainty, Parametric variability, Observation Error, **Model Discrepancy**

Bayesian Calibration of Computer Models[1]

Suppose that we have n observations of the physical system $z_1, \ldots, z_n$

$$\mathbf{z_i} = \underbrace{\boldsymbol{\zeta(x_i)}}_{\textbf{\textit{True value}}} + \boldsymbol{\varepsilon_i} \tag{1}$$

$$\mathbf{z_i} = \underbrace{\boldsymbol{\eta(x_i, \theta)}}_{\textbf{\textit{Physical Model}}} + \underbrace{\boldsymbol{\delta(x_i)}}_{\textbf{\textit{Model Discrepancy}}} + \boldsymbol{\varepsilon_i} \tag{2}$$

A Gaussian Process prior is used to model both calibration parameters and model discrepancy

$$\boldsymbol{\eta(\cdot,\cdot) \sim GP(m_1(\cdot,\cdot), c_1\{(\cdot,\cdot),(\cdot,\cdot)\})}$$

and

$$\boldsymbol{\delta(\cdot) \sim GP(m_2(\cdot), c_2(\cdot,\cdot))}.$$

[1] known as Kennedy–O'Hagan approach

Motivating Example

Simple Machine

The amount of work delivered is proportional to effort we put into it, that is

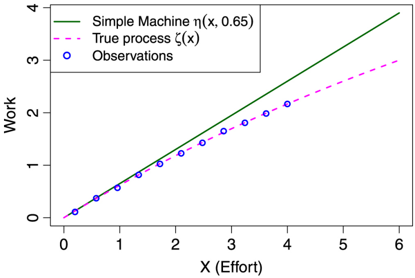
$$\eta(x,\theta) = \theta x$$

Synthetic Data

$z_i = \zeta(x_i) + \varepsilon_i, \quad i = 1, \ldots, n$
where ε_i are $i.i.d.$ $N(0, 0.01^2)$ and inputs x_i are evenly spread over the interval $[0.2, 4]$

True Process

$$\zeta(x) = \frac{\theta x}{1 + x/20}$$



The importance of Model Discrepancy

Figure: Posterior densities of θ

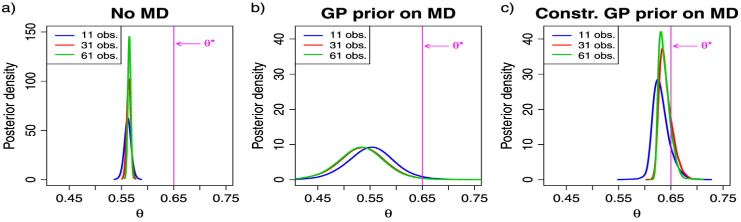
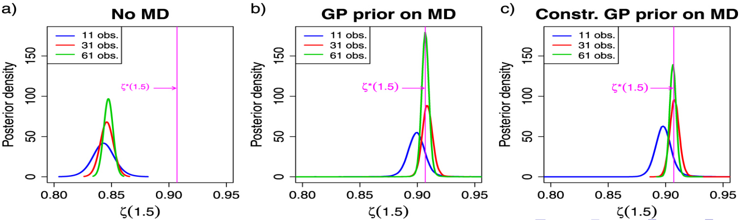


Figure: Posterior densities of true process at $x_0 = 1.5$



From MDT to MyMDT

- Develop a new methodology for the calibration problem for personalized computer models.
- Remain in the same Bayesian framework as proposed by O'Hagan, but we model the calibration parameters and the model discrepancy in a Mixed model Approach
- Personalize and at the same time utilize population based knowledge (HUNT data)
- Learning between Digital Twins

The formulation of the calibration problem will be the following

$$z_{ij} = \eta(x_{ij}, \theta_j) + \delta(x_{ij}) + \epsilon_{ij}, \tag{3}$$

where j is the indicator of the j^{th} person.

Research Plan

Personalized Approach

A new "personalized" approach will be developed for the calibration problem–Extend O'Hagan's approach in a Mixed Model approach.

Cardiovascular Model–Understanding

The Bayesian Calibration Approach will be applied to the cardiovascular Model.

Merge Approaches–Apply to Cardiovascular Model

Merge the approaches that mentioned before & apply it to the cardiovascular model.

Blood pressure and Physical Activity

Generalized Linear Mixed Models (GLMM) will be used to analyze the relation between Cardiovascular diseases (CVD), Blood Pressure (BP) and Personal Activity Intelligence (PAI).

Courses

1. MA8704 Probability Theory and Asymptotic Methods–Fall 2018
   (7.5 credits)
2. MA8701 General Statistical Methods–Spring 2019
   (7.5 credits)
3. MA8001 Methods for Model based personalized Medicine–Spring 2019
   (7.5 credits)
4. SMED8004 Research, ethics and society–Fall 2019
   (5.0 credits)
5. SMED8005 Communication of Science–Fall 2019
   (3.0 credits)

References

Kennedy, Marc C., & Anthony O'Hagan (2001)
Bayesian calibration of computer models.
Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63(3), 425-464.

Brynjarsdóttir, Jenný, & Anthony O'Hagan(2014)
The importance of model discrepancy.
Inverse Problems 30(11), 114007.

Smith, B. E., & Madigan, V. M. (2018)
Understanding the Haemodynamics of Hypertension.
Current hypertension reports 20(4), 29.

Krokstad, S., Langhammer, A., Hveem, K., Holmen, T. L., Midthjell, K., Stene, T. R., ... & Holmen, J. (2012)
Cohort profile: the HUNT study, Norway.
International journal of epidemiology 42(4), 968–977.