Statistical learning of the menstrual cycle from noisy and missing hormone observations

Iñigo Urteaga

joint work with

Tristan Bertin, Theresa M. Hardy, David J. Albers and Noémie Elhadad

Applied Physics and Applied Mathematics
Data Science Institute
Biomedical informatics

Columbia University
in the City of New York

June 2, 2022

The menstrual cycle and hormonal dynamics

The female reproductive endocrine system drives the hormonal cycle and its phases

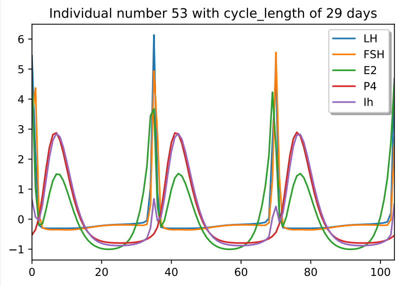

Figure: An individual's hormonal cycle over time

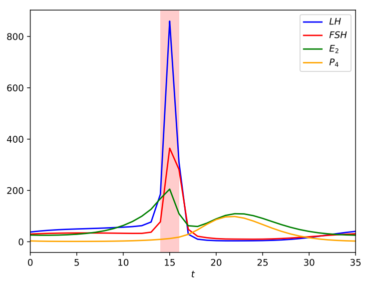

Figure: Ovulation separates the follicular and luteal phases.

The menstrual cycle and hormonal dynamics

The female reproductive endocrine system drives the hormonal cycle and its phases

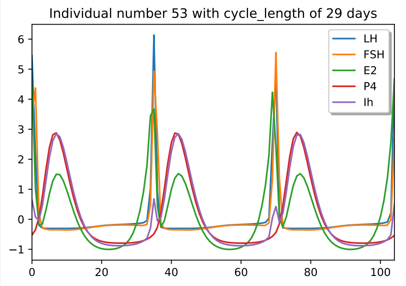

Figure: An individual's hormonal cycle over time

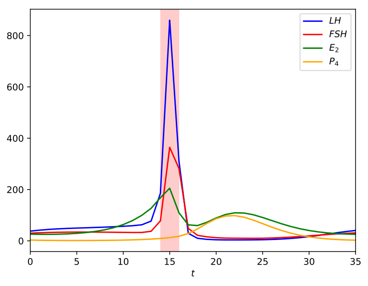

Figure: Ovulation separates the follicular and luteal phases.

The characterization of hormone levels over time

Full understanding of menstrual cycle

The menstrual cycle and hormonal dynamics

A complete characterization remains an open research question

Due (in part) to
a lack of direct and reliable measurements
over time and across individuals

The menstrual cycle and hormonal dynamics

A complete characterization remains an open research question

Due (in part) to
a lack of direct and reliable measurements
over time and across individuals

Current studies of female hormone levels over time

- Daily measurements of hormones
- Small-scale (30 individuals)
- Mechanistic models validated against these datasets

Challenges

Intrinsic challenges

Variability from one individual to another

- In hormone levels over-time
- In the timing of phases
- In the length of the menstrual cycle

Challenges

Intrinsic challenges

Variability from one individual to another

- In hormone levels over-time
- In the timing of phases
- In the length of the menstrual cycle

Extrinsic challenges

Over time, granular hormone sample collection

- The cost of repeated sampling
- Throughout many cycles, across many individuals

Challenges

Intrinsic challenges

Variability from one individual to another

- In hormone levels over-time
- In the timing of phases
- In the length of the menstrual cycle

Extrinsic challenges

Over time, granular hormone sample collection

- The cost of repeated sampling
- Throughout many cycles, across many individuals

Characterization of hormone levels over time

Very challenging and costly in practice!

Learning task and goal

How to infer the reproductive hormonal dynamics over time

In a minimally-invasive, pragmatic measurement setting

Learning task and goal

How to infer the reproductive hormonal dynamics over time

In a minimally-invasive, pragmatic measurement setting

Model and predict female reproductive hormonal patterns

Personalized and accurate forecasting of daily hormone levels

Female reproductive hormone dataset

Mechanistically simulated, diverse and realistic dataset

- Use mechanistic models to simulate hormone trajectories
- Simulations grounded on real-world cycle data
  e.g., self-tracked cycle lengths and ovulation through BBT
- Realistic simulated reproductive hormones over many cycles
  e.g., **inter-person variability**

Female reproductive hormone dataset

Mechanistically simulated, diverse and realistic dataset

- Use mechanistic models to simulate hormone trajectories
- Simulations grounded on real-world cycle data
  e.g., self-tracked cycle lengths and ovulation through BBT
- Realistic simulated reproductive hormones over many cycles
  e.g., **inter-person variability**

Reconstruct the hormone patterns based on few samples

Sampling rate and noise requirements to reconstruct the cycle

The learning task at hand

Reconstruction and prediction with minimally invasive sampling

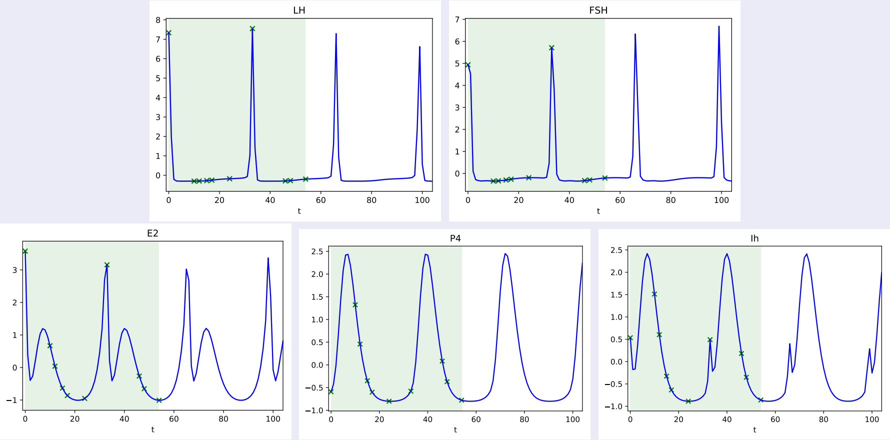

Figure: Simulated hormone levels over time, with limited measurements.

Learning female reproductive hormones

End-to-end statistical framework [1]

- Combines probabilistic generative models
  - Individualized Multi-task Gaussian processes (MGPs)
- With neural networks
  - Population level dilated convolutional architecture

Reference

[1] *Multi-Task Gaussian Processes and Dilated Convolutional Networks for Reconstruction of Reproductive Hormonal Dynamics*. I. Urteaga, T. Bertin, T.M. Hardy, D.J. Albers, N. Elhadad. In Proceedings of the 4th Machine Learning for Healthcare, volume 106, of Proceedings of Machine Learning Research, pages 66–90, 09–10 Aug 2019. PMLR

MGPs for female reproductive hormones

Individual Multi-task Gaussian Process (MGP)

Given a few **non-uniformly sampled measurements** per individual, learn a personalized posterior distribution of hormone levels over time

MGPs for female reproductive hormones

Individual Multi-task Gaussian Process (MGP)

Given a few **non-uniformly sampled measurements** per individual, learn a personalized posterior distribution of hormone levels over time

MGP to deal with uncertainty

Irregular and sparse sampling
Noisy and indirect measurements

MGPs for female reproductive hormones

Per-individual MGP model

- Covariance function (expert prior knowledge)

$$k\left(y(t), y(t')|\theta\right) = k(h, h'|\theta_h) \otimes k(t, t'|\theta_t)$$

  - $k(h, h')$ is a PSD matrix for inter-hormone dependencies
  - $k(t, t')$ the covariance function over time

MGPs for female reproductive hormones

Per-individual MGP model

- Covariance function (expert prior knowledge)

$$k\left(y(t), y(t')|\theta\right) = k(h, h'|\theta_h) \otimes k(t, t'|\theta_t)$$

  - $k(h, h')$ is a PSD matrix for inter-hormone dependencies
  - $k(t, t')$ the covariance function over time

Per-individual MGP fitting

- Hormone levels $y_i(t_i)$, sampled at few time-instants $t_i \subset \{1, \cdots, T\}$

MGPs for female reproductive hormones

Per-individual MGP model

- Covariance function (expert prior knowledge)

$$k\left(y(t), y(t')|\theta\right) = k(h, h'|\theta_h) \otimes k(t, t'|\theta_t)$$

  - $k(h, h')$ is a PSD matrix for inter-hormone dependencies
  - $k(t, t')$ the covariance function over time

Per-individual MGP fitting

- Hormone levels $y_i(t_i)$, sampled at few time-instants $t_i \subset \{1, \cdots, T\}$
- Learn parameters θ_i per individual,
  maximizing the marginal likelihood of their observed data

MGPs for female reproductive hormones

Per-individual MGP model

- Covariance function (expert prior knowledge)

$$k\left(y(t), y(t')|\theta\right) = k(h, h'|\theta_h) \otimes k(t, t'|\theta_t)$$

  - $k(h, h')$ is a PSD matrix for inter-hormone dependencies
  - $k(t, t')$ the covariance function over time

Per-individual MGP fitting

- Hormone levels $y_i(t_i)$, sampled at few time-instants $t_i \subset \{1, \cdots, T\}$
- Learn parameters θ_i per individual,
  maximizing the marginal likelihood of their observed data
- Individual MGP posterior over any arbitrary time point

$$z_i(t) \sim MGP(m_i(\cdot), k_i(\cdot, \cdot)|\theta_i), t = \{1, \cdots, T\}$$

MGPs for female reproductive hormones

MGP posteriors

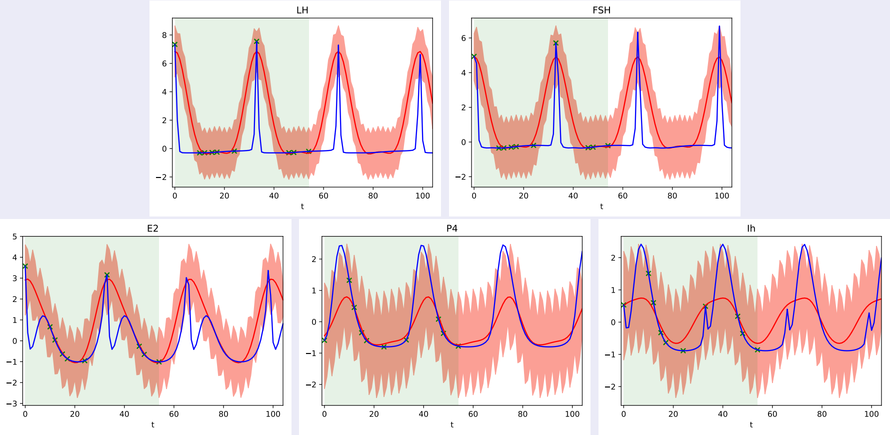

Figure: True hormone levels (blue), training data points (green crosses) and GP posterior (in red).

Neural networks for female reproductive hormones

Dilated Convolutional Neural Network (DCNN)

Given MGP posterior samples over **equally spaced time instants**
learn to correct mispredictions from population data

Neural networks for female reproductive hormones

Dilated Convolutional Neural Network (DCNN)

Given MGP posterior samples over **equally spaced time instants**
learn to correct mispredictions from population data

DCNN to improve accuracy

MGP posterior samples to learn from
at regularly sampled measurements

DCNNs for female reproductive hormones

Dilated Convolutional Neural Network (DCNN)

- Draw S hormone level samples $z_i^{(s)}(t), t = \{1, \cdots, T\}$

DCNNs for female reproductive hormones

Dilated Convolutional Neural Network (DCNN)

- Draw S hormone level samples $z_i^{(s)}(t), t = \{1, \cdots, T\}$
- Learn a population level mapping from $z_i^{(s)}$ to $y_i(s)$, $\forall i, s$

DCNNs for female reproductive hormones

Dilated Convolutional Neural Network (DCNN)

- Draw S hormone level samples $z_i^{(s)}(t), t = \{1, \cdots, T\}$
- Learn a population level mapping from $z_i^{(s)}$ to $y_i(s)$, $\forall i, s$
- Back-propagating the loss to DCNN parameters

$$\begin{aligned} w^* &= \underset{w}{\operatorname{argmin}} \sum_{i=1}^{I} \mathbb{E}_{z_i \sim MGP(\cdot|\theta_i)} \left\{ \mathcal{L}_2 \left[g\left(z_i, w\right), y_i \right] \right\} \\ &= \underset{w}{\operatorname{argmin}} \sum_{i=1}^{I} \sum_{s=1}^{S} \left[g\left(z_i^{(s)}, w\right) - y_i \right]^2 \end{aligned}$$

MGP+DCNNs for female reproductive hormones

DCNN expected output

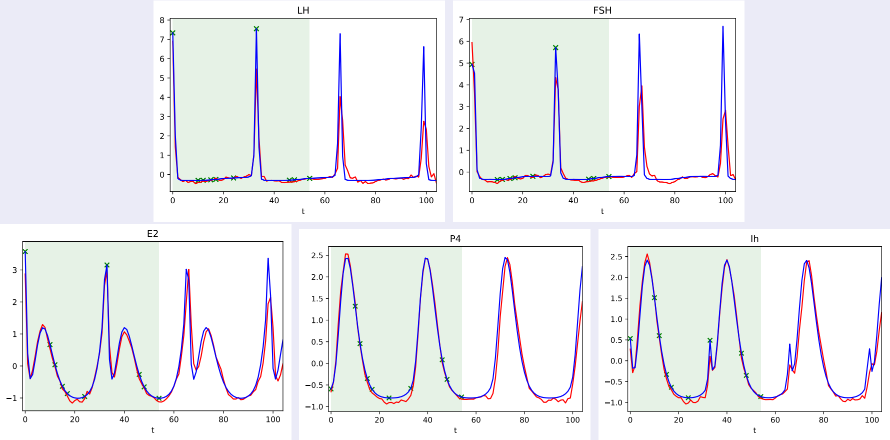

Figure: True hormone levels (blue), training data points (green crosses) and DCNN output (in red).

MGP+DCNNs for female reproductive hormones

MGP and DCNN output

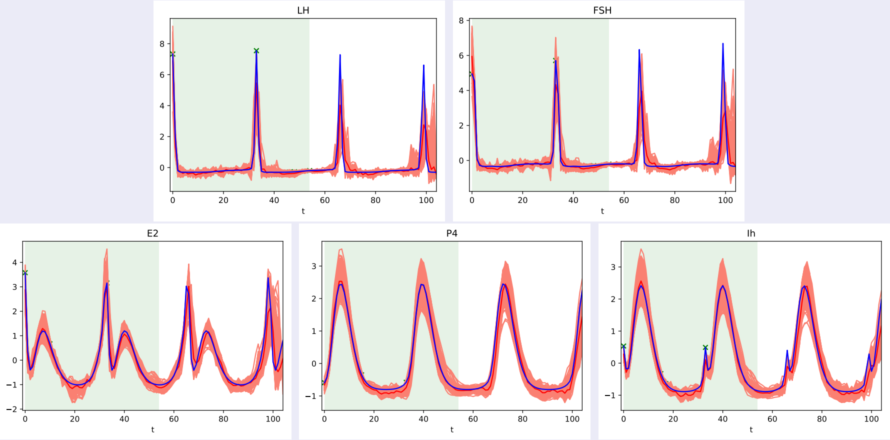

Figure: True hormone levels (blue), training data points (green crosses) and MGP + DCNN output (in red).

MGP+DCNNs for female reproductive hormones

Overall results: realistic sampling budgets with accuracy

Test-set overall MSE	Subsampling budget				
Model	$\lvert t_i \rvert = 10$	$\lvert t_i \rvert = 15$	$\lvert t_i \rvert = 25$	$\lvert t_i \rvert = 35$	$\lvert t_i \rvert = 70$
LSTM	0.358	0.247	0.203	0.186	0.168
Independent GPs	1.245	1.066	0.833	0.140	0.109
MGP	0.823	1.085	0.708	0.127	0.057
MGP-DCNN	0.302	0.120	0.189	0.041	0.071

Table: Test-set overall average MSE for all hormones

Summary

Statistical learning for female reproductive hormones

- Proposed a personalized hormonal dynamic modeling framework
- Accommodate realistic (non-uniform, reduced) sampling budgets
- Accurate hormone level reconstruction is possible
- Reduce burden for better understanding of the cycle

Summary

Statistical learning for female reproductive hormones

- Proposed a personalized hormonal dynamic modeling framework
- Accommodate realistic (non-uniform, reduced) sampling budgets
- Accurate hormone level reconstruction is possible
- Reduce burden for better understanding of the cycle

Limitations

- Performance subject to evidence of phase
  e.g., sample peak/valley
- Limited within-person variability
- Somehow restrictive training set-up

Mental nuggets (i.e., future work)

Hybridized modeling

Merge mechanistic models & prior knowledge with data-driven learning

Mental nuggets (i.e., future work)

Hybridized modeling

Merge mechanistic models & prior knowledge with data-driven learning

Other data modalities

Self-tracking data of ovulation and/or period events

Mental nuggets (i.e., future work)

Hybridized modeling

Merge mechanistic models & prior knowledge with data-driven learning

Other data modalities

Self-tracking data of ovulation and/or period events

Beyond hormone prediction and clinical impact

Real hormone measurements and health outcomes

Thank you

Questions?

Gaussian Processes

Overview of GPs

- A GP is a Bayesian nonparametric approach
  useful for smoothing and forecasting
- A GP is a distribution over functions

$$f(x) \sim GP(m(x), k(x,x'))$$

$$\text{with} \begin{cases} \text{mean function } m(x) = \mathbb{E}\left\{f(x)\right\} \\ \text{covariance (or kernel) function } k(x,x') = \mathrm{Cov}\left\{f(x), f(x')\right\} \end{cases}$$

Gaussian Processes

Bayesian view of GPs

- A GP imposes a prior over functions
  with expected properties, e.g, smoothness, periodicity, etc.
- The properties of the functions induced by a GP
  are determined by the mean and kernel function
- As data is observed,
  the GP prior is updated to the posterior

Dilated Convolutional Neural networks

Overview of DCNNs

A Dilated Convolution

$$F[t] = (X *_d f)[t] = \sum_{k=0}^{K-1} f[k] X[t - dk]$$

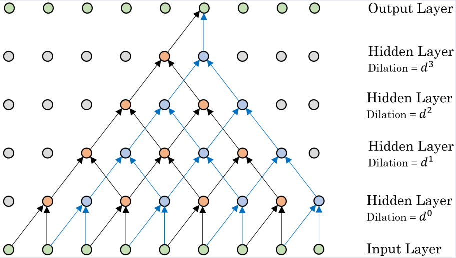

Figure: Dilated Convolutional Neural Network: 4 layers, $d = 2$ and $K = 2$.

Dilated Convolutional Neural Networks

Benefits of DCNNs

- Captures long dependencies in the input sequence
  via the dilated layers

Dilated Convolutional Neural Networks

Benefits of DCNNs

- Captures long dependencies in the input sequence
  via the dilated layers
- Reduced gradient instability
  exploding and vanishing gradients common with recurrence

Dilated Convolutional Neural Networks

Benefits of DCNNs

- Captures long dependencies in the input sequence
  via the dilated layers
- Reduced gradient instability
  exploding and vanishing gradients common with recurrence
- Lower memory requirements in training
  convolution filters are shared across layers
  backpropagation path depends only on the network depth