Gaussian process imputation of multiple financial series

ICASSP 2020

Authors
Taco de Wolff
Alejandro Cuevas
Felipe Tobar

Supported by Fondecyt-Iniciación 11171165 and Conicyt-PIA AFB 170001 Center for Mathematical Modeling

UNIVERSIDAD DE CHILE

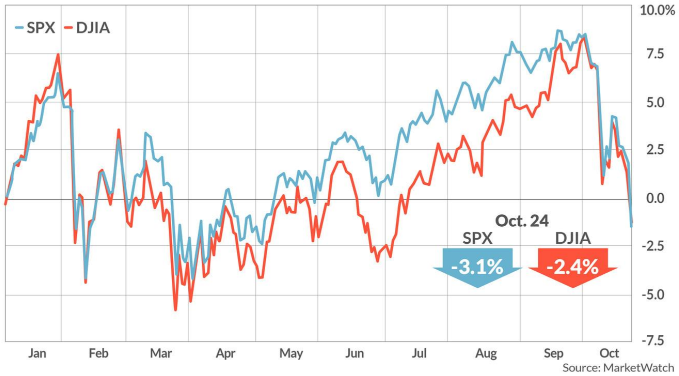
CMM Center for Mathematical Modeling

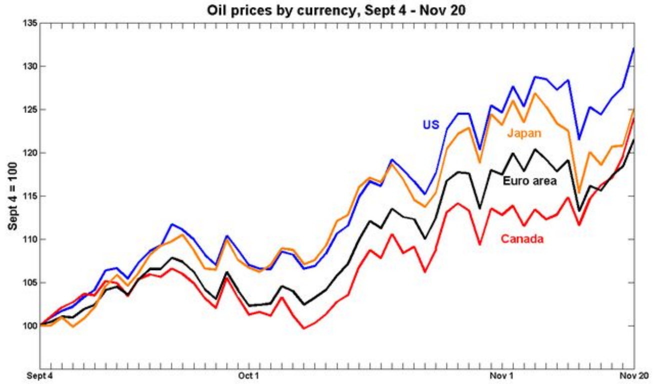
Training multiple financial indicators simultaneously to analyse market dependencies and to recover missing data

Presentation layout

1. Financial data series
2. Background
    a. Gaussian processes (GPs)
    b. Spectral mixture kernel (SM)
    c. Multi-output Gaussian processes (MOGPs)
    d. Multi-output spectral mixture kernel (MOSM)
3. Experimental setup: data, methods, and optimization
4. Results and comparison
    a. Gold, Oil, NASDAQ, USD-index
    b. Currency exchanges with the USD
5. Discussion and conclusion

Financial data series





Relevant why?

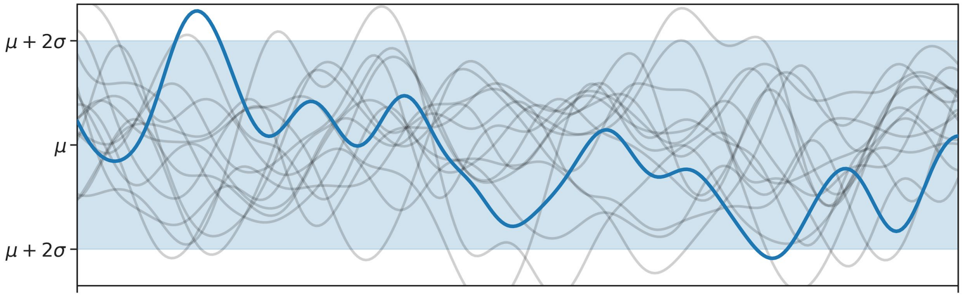
- Understanding market behaviour and dynamics
- Exploring market interdependencies between stocks, currencies, ...
- Predicting stock markets (ambitious!)
- Predict alternative market trajectory when omitting key (governmental) policy

Figures: https://www.marketwatch.com/story/dow-futures-drop-220-points-as-stock-market-extends-rout-2018-10-24
https://worthwhile.typepad.com/worthwhile_canadian_initi/2007/11/recent-oil-pric.html

Background: Gaussian processes

Gaussian processes: a Gaussian distribution over functions [1]

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



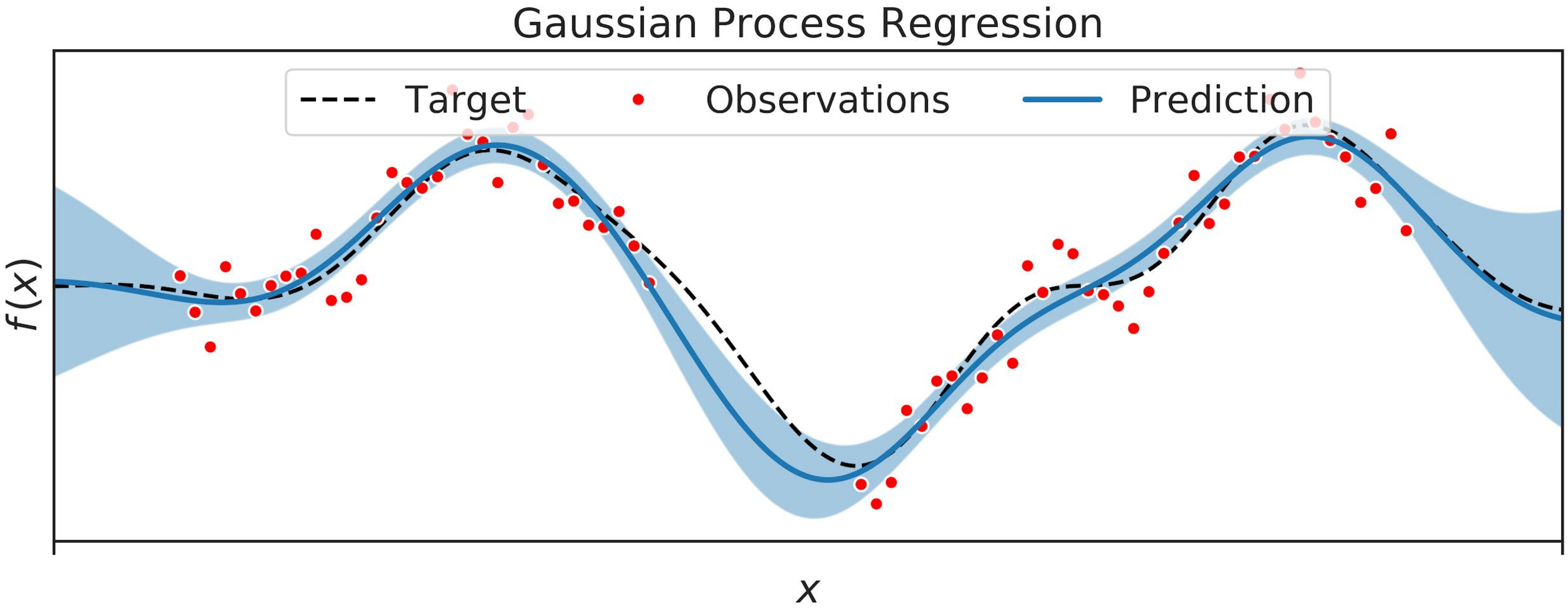
where m(x) is the mean function,
and k(x,x') the covariance function

$$\begin{aligned} m(x) &= \mathbb{E}[f(x)] \\ k(x, x') &= \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))] \\ &= \text{cov}(f(x), f(x')) \,. \end{aligned}$$

in general we assume stationarity, i.e. $k(x, x') = k(x - x')$,

[1] Rasmussen and Williams, "Gaussian processes for machine learning", MIT Press, 2006

Background: a trained RBF kernel



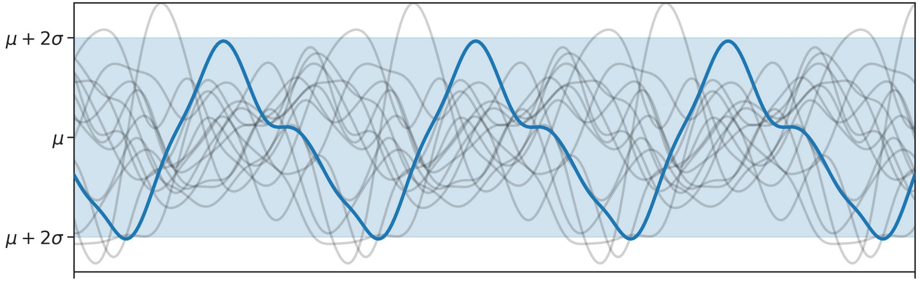
A trained RBF kernel: the mean represents the predicted time series where the confidence intervals collapse near our data but widen when further away

Background: spectral mixture kernel

Spectral mixture kernel: define the covariance function in frequency space [2]

$$k(\tau) = \int S(\omega) \exp\left(2\pi i \omega \cdot \tau\right) d\omega$$

$$S(\omega) = \int k(\tau) \exp\left(-2\pi i \omega \cdot \tau\right) d\tau$$



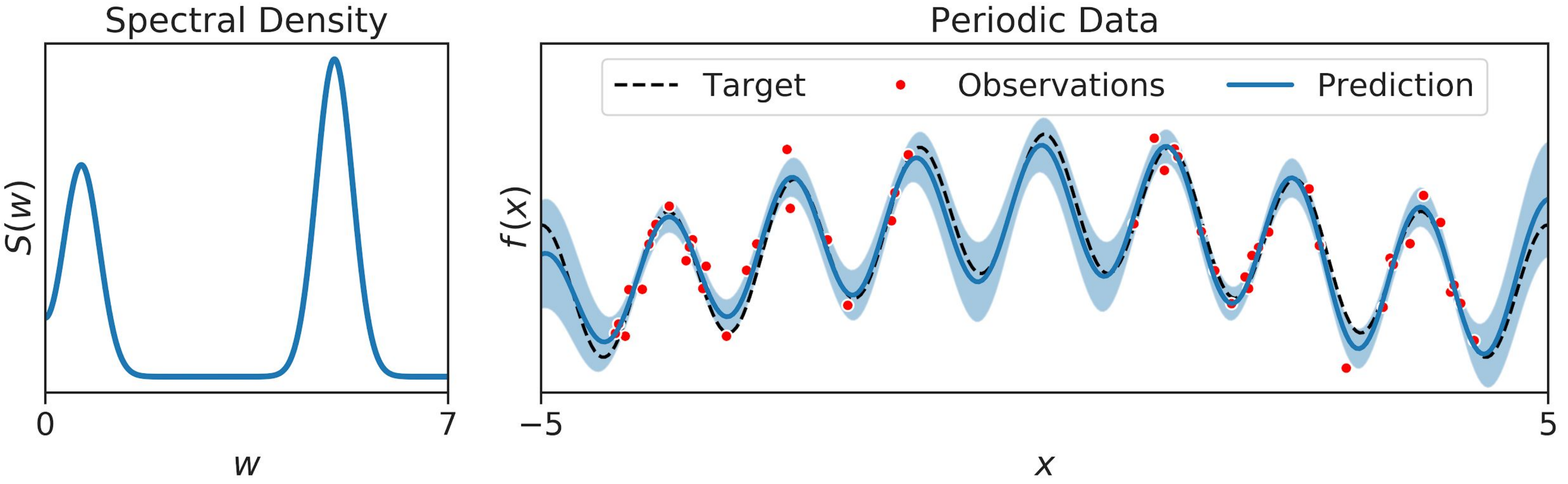
Using the Fourier pair above, we define Q Gaussian distributions in frequency space, and obtain the following kernel:

$$k(\tau) = \sum_{q=1}^{Q} w_q \exp\left(-\frac{1}{2}\tau^\top \Sigma_q \tau\right) \cos(\mu_q^\top \tau)$$

τ is the distance between two points in the time domain
μ is the frequency of the pattern
Σ is a covariance matrix of the inverse length scales
w is the weight for each Q

[2] Wilson and Adams, “Gaussian process kernels for pattern discovery and extrapolation”, ICML 30, 2013

Background: a trained spectral mixture kernel



A spectral mixture kernel with Q=2, fitting two Gaussians in the frequency domain to fit our data

Background: MOGPs and the MOSM

Multi-output Gaussian processes: learn inter-channel dependencies by using a matrix of covariance functions between channel i and channel j

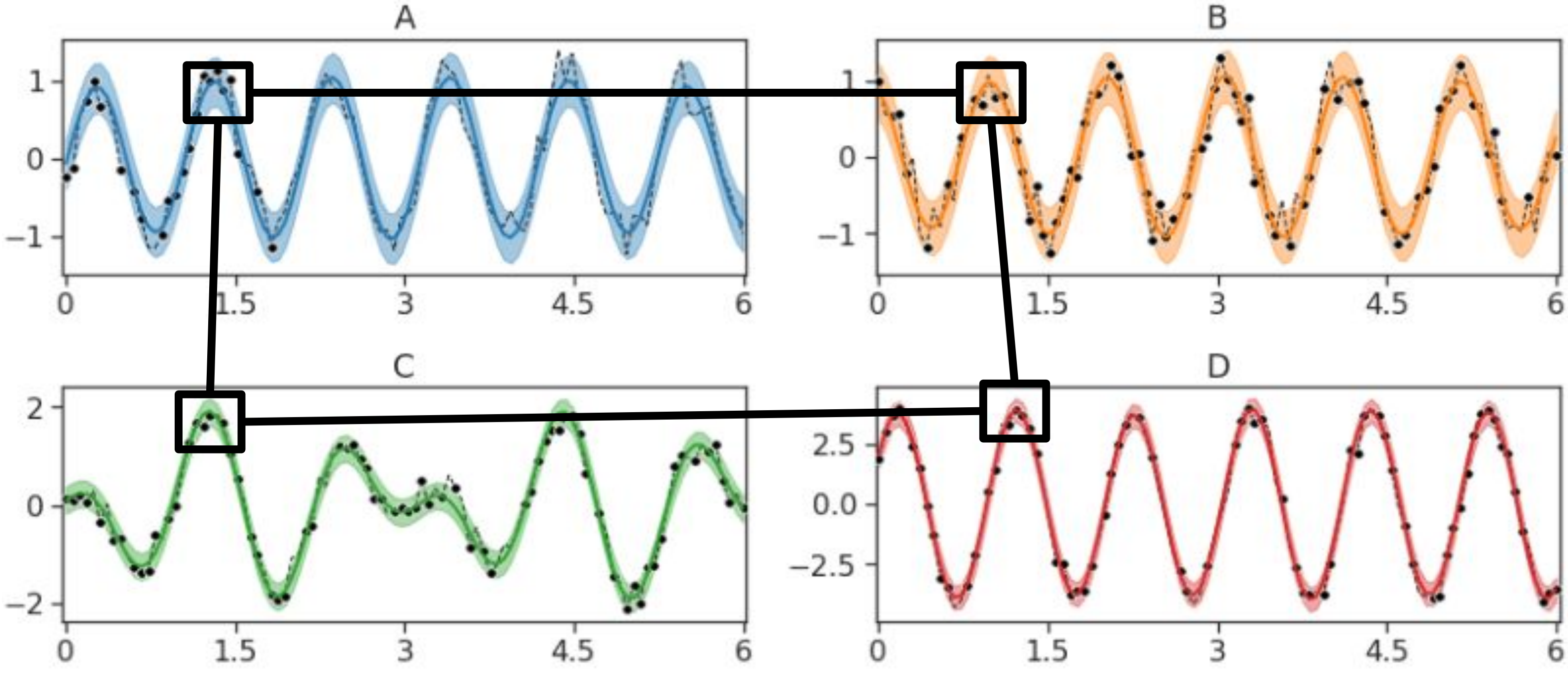
We specify the multi-output spectral mixture kernel (MOSM) [3]

$$k_{ij}(\tau) = \sum_{q=1}^{Q} \alpha_{ij}^{(q)} \exp\left(-\frac{1}{2}\left(\tau + \theta_{ij}^{(q)}\right)^{\top} \Sigma_{ij}^{(q)} \left(\tau + \theta_{ij}^{(q)}\right)\right)$$
$$\cdot \cos\left(\left(\tau + \theta_{ij}^{(q)}\right)^{\top} \mu_{ij}^{(q)} + \phi_{ij}^{(q)}\right),$$

τ is the distance between two points in the time domain
α is the weight for each Q
θ the delay
ϕ the phase shift
μ is the frequency of the pattern
Σ is a covariance matrix of the inverse length scales

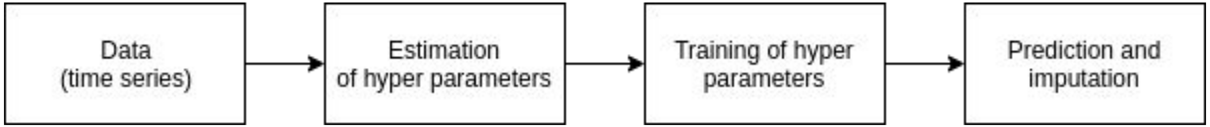
[3] Parra and Tobar, “Spectral mixture kernels for multi-output Gaussian processes”, NIPS 30, 2017

Background: learning across channels



Correlating four channels using the MOSM: train signal frequencies, amplitudes, delays and phases between channels

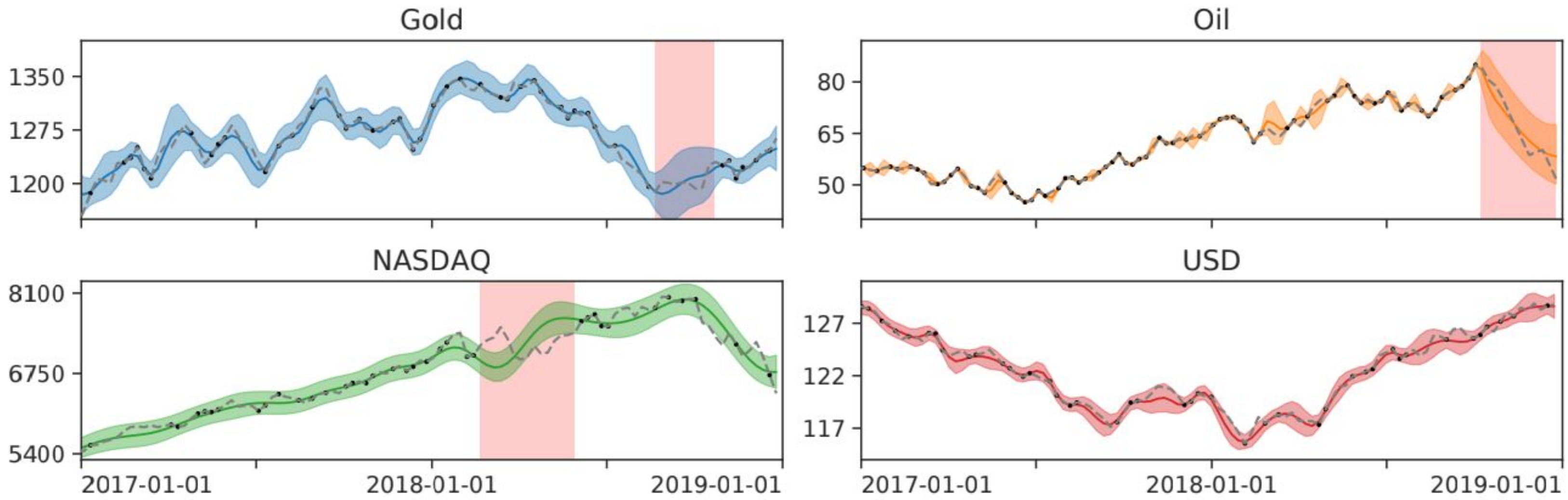
Experimental setup



- Data from multiple time series
- Using the MOSM kernel as our model
- Hyper parameter estimation for improved training using BNSE [4]
- Training the hyper parameters using L-BFGS-B and an Nvidia GeForce GTX 1080 video card
- Predict and imputate data with the trained model
- Analyze the trained hyper parameters to learn about cross-correlations

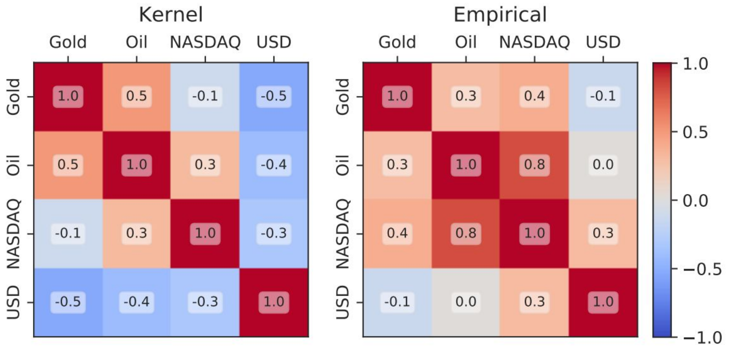
[4] Tobar, “Bayesian nonparametric spectral estimation”, NIPS 32, 2018

Results: gold, oil, NASDAQ, USD-index



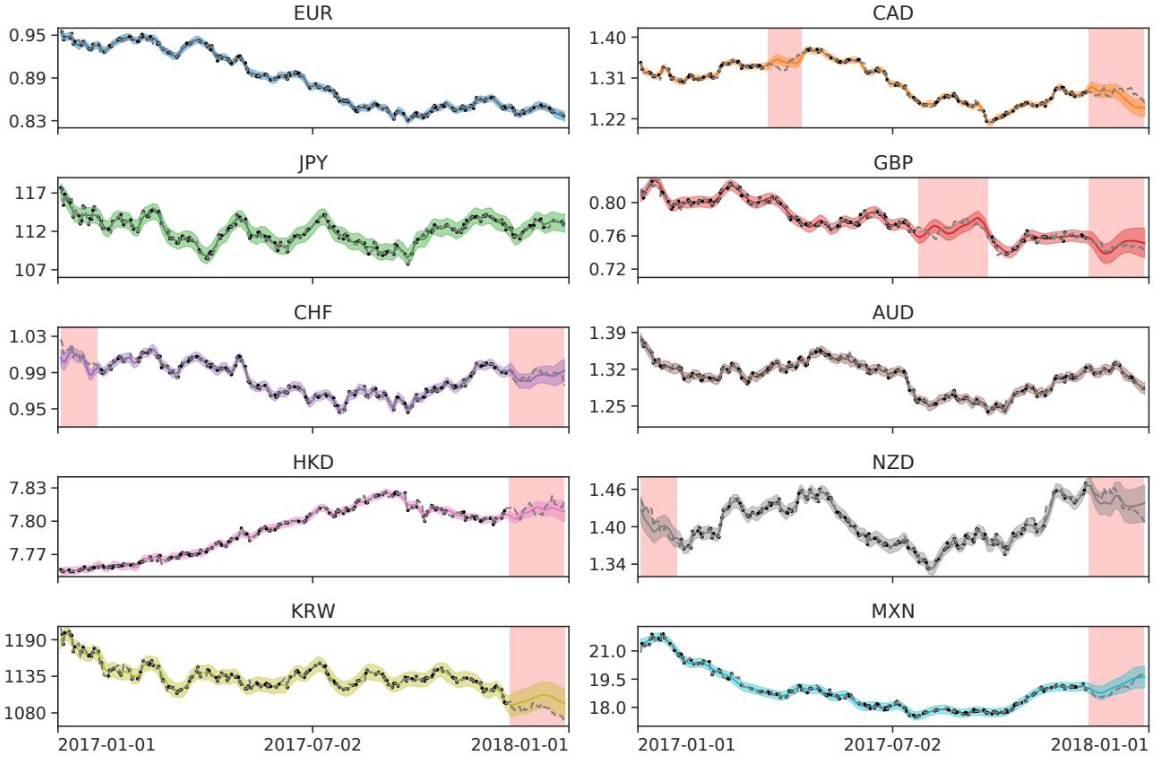
Data imputation

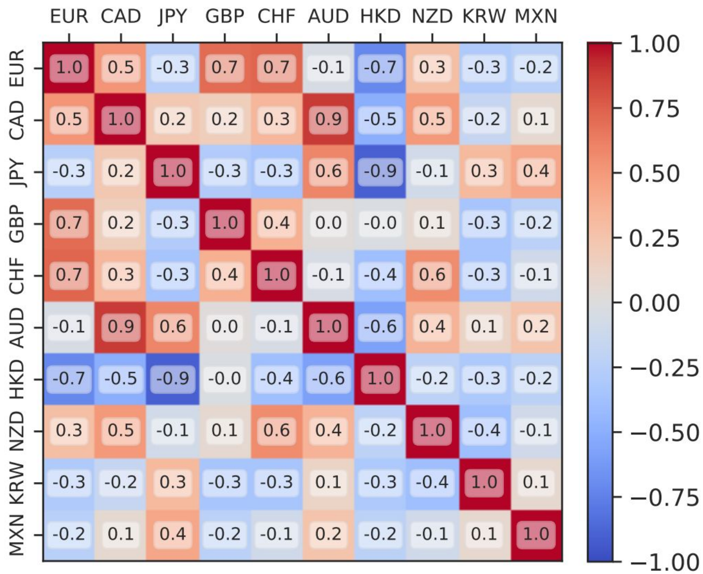
Cross-correlation matrix



Red shade: missing data
Black dots: data points used for training
Colored line: prediction mean and confidence interval

Results: currency exchanges with the USD





Red shade: missing data
Black dots: data points used for training
Colored line: prediction mean and confidence interval

Results and comparison

Model	Gold, Oil, NASDAQ, USD index nMAE (10^{-2})	nRMSE (10^{-2})	Currency exchange rates nMAE (10^{-3})	nRMSE (10^{-3})
SM-IGP [17]	2.817 ± 0.000	5.071 ± 0.000	5.478 ± 0.000	7.481 ± 0.000
SM-LMC [20, 21]	2.5 ± 0.4	3.4 ± 0.6	6.6 ± 0.5	8.9 ± 0.6
CSM [18]	1.88 ± 0.02	$\mathbf{2.44 \pm 0.06}$	8 ± 1	10 ± 2
MOSM [19]	$\mathbf{1.8 \pm 0.1}$	2.6 ± 0.4	$\mathbf{4.8 \pm 0.3}$	$\mathbf{6.5 \pm 0.4}$

Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for four kernels. The errors are normalized between the channels so that they are comparable.

We observe that the MOSM is a flexible kernel that is able to fit the data well.

[17] Wilson and Adams, “Gaussian process kernels for pattern discovery and extrapolation”, ICML 30, 2013
[18] Ulrich, “Gaussian process kernels for cross-spectrum analysis in electrophysiological time series”, 2016
[19] Parra and Tobar, “Spectral mixture kernels for multi-output Gaussian processes”, NIPS 30, 2017
[20] Goovaerts, “Geostatistics for natural resources evaluation”, Oxford University Press, 1997
[21] Wilson, “Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes”, 2014

Discussion and conclusion

- MOSM and other MOGPs are capable of capturing cross-correlation information from financial time series
- The MOSM is a more flexible kernel that can fit the data better, but is also harder to train
- Parameter estimation is crucial for all MOGPs to improve training results
- The interpretation of trained hyper parameter values can aid in increased understanding of market dependencies and dynamics
- Future work could improve parameter estimation and the use of non-Gaussian processes (such as Student-t processes)