

# Model globokih gaussovskih procesov za identifikacijo procesa Bouc-Wen

Tadej Krivec<sup>1,2</sup>, Juš Kocijan<sup>1,3</sup>

<sup>1</sup>Institut Jožef Stefan, Jamova cesta 39, Ljubljana, Slovenija

<sup>2</sup>Mednarodna podiplomska šola Jožefa Stefana, Jamova cesta 39, Ljubljana, Slovenija

<sup>3</sup>Univerza v Novi Gorici, Vipavska 13, Nova Gorica, Slovenija

E-pošta: tadej.krivec@ijs.si, jus.kocijan@ijs.si

## Deep Gaussian process model for the identification of the Bouc-Wen process

*Gaussian process regression is a well know probabilistic machine learning method that can be used for dynamic system identification where the goal is to find a simulation model. Full Gaussian process regression has a cubic complexity with respect to the input data. Different approximations are used which make a compromise between the flexibility of the model and the computational demand. This article presents the use of a deep Gaussian process, a hierarchical Gaussian process that provides a very flexible structure on the cost of increasing the computational complexity, to model a dynamic system. Since this kind of architecture is analytically intractable, doubly stochastic variational inference is used for approximation. Prediction and simulation results are compared between different number of hidden layers for modeling the Bouc-Wen process.*

## 1 Uvod

Idejo o modeliranju z gaussovskimi procesi je področju strojnega učenja približal C.E. Rasmussen s svojo doktorsko dizertacijo [1]. Gaussovski procesi ponujajo alternativo bolj priljubljenemu pristopu k modeliranju z nevronskimi mrežami. Monografija s tega področja je [2], modeliranje dinamičnih sistemov z gaussovskimi procesi pa je opisano v [3]. Glavni problem metode predstavlja kubična računski kompleksnost glede na število vhodnih podatkov. Pregled približkov za znižanje računski kompleksnosti lahko najdemo v [4]. Med približki, ki se lahko zelo dobro skalirajo na veliko količino podatkov je stohastični variacijski gaussovski proces [5]. Ta približek je osnova za širjenje ideje modeliranja na hierarhično strukturo, podobno nevronskim mrežam, kjer so nivoji nevronov zamenjani z gaussovskimi procesi. Takemu modelu pravimo globok gaussovski proces in je predstavljen v [6]. V tem prispevku predstavljamo uporabo globokih gaussovskih procesov za identifikacijo dinamičnih sistemov na primeru procesa Bouc-Wen v kombinaciji z avtoregresijskim modelom.

## 2 Nelinearna identifikacija dinamičnih sistemov

S hitrim razvojem informacijske tehnologije, predvsem z znižanjem stroškov shranjevanja podatkov in računalniškega obdelovanja podatkov, se popularnost in uporabnost metod za identifikacijo dinamičnih sistemov hitro zvišuje. V tem poglavju je predstavljen možen podatkovni način kako pristopiti k reševanju identifikacije dinamičnih sistemov z metodo gaussovskih procesov.

### 2.1 Avtoregresijski model

Nelinearni dinamični sistem lahko modeliramo kot

$$\mathbf{y}(k) = f(\mathbf{z}(k), \boldsymbol{\theta}) + \boldsymbol{\nu}(k), \quad (1)$$

kjer  $k$  predstavlja zaporedno število vzorca,  $f$  nelinearno preslikavo,  $\boldsymbol{\theta}$  vektor parametrov,  $\boldsymbol{\nu}(k)$  gaussovski beli šum in  $\mathbf{z}(k)$ ,  $\mathbf{y}(k)$  predstavljata vektor vhodov in izhodov v odvisnosti od  $k$ . Predpostavimo, da interne strukture sistema, ki ga identificiramo, ne poznamo (ang. black-box). Manipuliramo lahko z vhodnim signalom in merimo izhod. V našem primeru predpostavimo, da ima model avtoregresijsko strukturo (ang. Nonlinear Autoregressive Exogenous Model). V takem primeru je vhodni vektor definiran kot

$$\mathbf{z}(k) = [\mathbf{y}(k - n_p), \dots, \mathbf{y}(k - n_p - n_a + 1), \mathbf{u}(k - n_k), \dots, \mathbf{u}(k - n_k - n_b + 1)], \quad (2)$$

kjer  $k$  predstavlja zaporedno število vzorca,  $n_p, n_a, n_k, n_b$  pa parametre zakasnitev vzorcev. Identifikacija dinamičnih sistemov je proces, kjer v našem primeru iščemo ustrezne parametre avtoregresijskega modela in modeliramo nelinearno preslikavo med vhodnim in izhodnim vektorjem. Tak model lahko uporabljamo za napovedovanje časovnih vrst, vodenje dinamičnih sistemov, detekcijo napak itd. Uporablja se lahko za napovedovanje ali simulacijo, kjer pri napovedi napovedujemo za en časovni korak vnaprej, pri simulaciji pa iterativno izračunamo bodoče vrednosti izhoda za poljubno število korakov vnaprej po enačbi

$$\hat{\mathbf{y}}(k) = f(\hat{\mathbf{z}}(k), \boldsymbol{\theta}), \quad (3)$$

kjer velja

$$\hat{\mathbf{z}}(k) = [\hat{\mathbf{y}}(k - n_p), \dots, \hat{\mathbf{y}}(k - n_p - n_a + 1), \mathbf{u}(k - n_k), \dots, \mathbf{u}(k - n_k - n_b + 1)] \quad (4)$$

in  $\hat{z}$  in  $\hat{y}$  predstavljata ocenjene (napovedane) vrednosti iz prejšnjih korakov.

## 2.2 Model gaussovskih procesov za dinamične sisteme

Model gaussovskih procesov je bayesovska metoda strojnega učenja, ki posplošuje gaussovsko porazdelitev naključnih spremenljivk ali vektorjev funkcijskih vrednosti. Predpostavimo, da imamo znane meritve vhodov in izhodov v obliki

$$\mathbf{y} = f(\mathbf{z}) + \boldsymbol{\epsilon}, \quad (5)$$

kjer  $\mathbf{z}$  predstavlja vhodni vektor v obliki avtoregresijskega modela,  $f$  latentno funkcijo,  $\mathbf{y}$  opazovane izhodne podatke popačene s šumom  $\boldsymbol{\epsilon}$ . Predpostavimo, da je šum neodvisno in identično porazdeljen z gaussovsko porazdelitvijo

$$\epsilon_n \sim \mathcal{N}(0, \sigma_n^2). \quad (6)$$

Vektor  $\mathbf{f} = [f_1, f_2, \dots, f_n]^T$  predstavlja vrednosti latentne funkcije kjer velja  $f_i = f(\mathbf{z}_i)$ . Stolpični vektor vhodov dolžine  $n$  sestavlja matriko  $\mathbf{Z} \in \mathcal{R}^{D \times n}$ , kjer je poljubna vrstica predstavljena z vektorjem  $\mathbf{z}$ . Množica podatkov je potem označena kot  $\mathcal{D} = (\mathbf{Z}, \mathbf{y})$ . Cilj je poiskati porazdelitev napovedi neznanega vektorja latentnih funkcijskih vrednosti  $\mathbf{f}_*$  pri znanih testnih lokacijah  $\mathbf{z}_*$ .

Gaussovski proces definira porazdelitev funkcij, bolj natančno porazdelitev vektorjev funkcijskih vrednosti latentne funkcije  $f$ . Definiran je s funkcijo srednje vrednosti  $m(\mathbf{z})$  in s kovariančno funkcijo  $k(\mathbf{z}, \mathbf{z}')$

$$m(\mathbf{z}) = \mathbb{E}[f(\mathbf{z})], \quad (7a)$$

$$k(\mathbf{z}, \mathbf{z}') = \mathbb{E}[(f(\mathbf{z}) - m(\mathbf{z}))(f(\mathbf{z}') - m(\mathbf{z}'))] \quad (7b)$$

in je označen z  $f(\mathbf{z}) \sim \mathcal{GP}(m(\mathbf{z}), k(\mathbf{z}, \mathbf{z}'))$ . Parametri kovariančne funkcije so poimenovani kot hiperparametri [3]. V modelu gaussovskih procesov je apriori porazdelitev nad vektorjem latentnih funkcijskih vrednosti transformirana z Bayesovim izrekom v aposteriorno porazdelitev

$$p(\mathbf{f}|\mathbf{y}, \boldsymbol{\theta}) = \frac{p(\mathbf{y}|\mathbf{f}, \boldsymbol{\theta})p(\mathbf{f}|\boldsymbol{\theta})}{p(\mathbf{y}|\boldsymbol{\theta})}, \quad (8)$$

kjer  $p(\mathbf{y}|\mathbf{f}, \boldsymbol{\theta})$  predstavlja pogojno verjetnost pri definiranih hiperparametrih in opazovanim vektorjem  $\mathbf{f}$ ,  $p(\mathbf{f}|\boldsymbol{\theta})$  apriori porazdelitev nad vektorjem latentnih funkcijskih vrednosti,  $p(\mathbf{y}|\boldsymbol{\theta})$  marginalno pogojno porazdelitev pri znanih hiperparametrih in  $p(\mathbf{f}|\mathbf{y}, \boldsymbol{\theta})$  predstavlja aposteriorno porazdelitev vektorja latentnih funkcijskih vrednosti. V nadaljevanju predpostavimo pogojenost porazdelitve glede na hiperparametre  $\boldsymbol{\theta}$  in jih v notaciji izpustimo.

Kovariančna funkcija definira apriori porazdelitev vektorja latentnih funkcijskih vrednosti. Kovariančna funkcija je lahko katerakoli funkcija, ki tvori simetrično, semipozitivno in kvadratno matriko. Popularna izbira je kovariančna funkcija RBF (ang. Radial Basis Function)

$$k(\mathbf{z}, \mathbf{z}') = \sigma_f^2 e^{-\frac{\|\mathbf{z} - \mathbf{z}'\|^2}{2l^2}}. \quad (9)$$

Po definiciji gaussovskega procesa ima skupna porazdelitev med vektorjem latentnih funkcijskih vrednosti  $\mathbf{f}$  in testnim vektorjem latentnih funkcijskih vrednosti  $\mathbf{f}_*$

tudi skupno gaussovsko porazdelitev. Apriori skupna porazdelitev je potem transformirana z Bayesovim izrekom ob znanem vektorju vrednosti izhodov na aposteriorno skupno porazdelitev definirano z

$$p(\mathbf{f}, \mathbf{f}_*|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{f})p(\mathbf{f}, \mathbf{f}_*)}{p(\mathbf{y})}, \quad (10)$$

kjer je apriori skupna porazdelitev definirana z

$$p(\mathbf{f}, \mathbf{f}_*) \sim \mathcal{N}(\mathbf{0}, \begin{bmatrix} \mathbf{K}_{nn} & \mathbf{K}_{n*} \\ \mathbf{K}_{*n} & \mathbf{K}_{**} \end{bmatrix}) \quad (11)$$

in  $\mathbf{K}_{nn}$  predstavlja kovariančno matriko opazovanih učnih podatkov,  $\mathbf{K}_*$  kovariančno matriko učnih in testnih podatkov and  $\mathbf{K}_{**}$  kovariančno matriko testnih podatkov. Pogojna porazdelitev  $p(\mathbf{y}|\mathbf{f})$  je definirana kot

$$p(\mathbf{y}|\mathbf{f}) = \mathcal{N}(\mathbf{f}, \sigma_n^2 \mathbf{I}), \quad (12)$$

kjer  $\mathbf{I}$  predstavlja matriko identitete in  $\sigma_n^2$  standardni odklon šuma. Hiperparametre učimo z maksimizacijo logaritma marginalne pogojne porazdelitve  $p(\mathbf{y})$ . Porazdelitev napovedi je izračunana z marginalizacijo  $\mathbf{f}$  iz skupne pogojne porazdelitve

$$p(\mathbf{f}_*|\mathbf{y}) = \int p(\mathbf{f}, \mathbf{f}_*|\mathbf{y})d\mathbf{f}. \quad (13)$$

Pri iterativni simulaciji je vhodni vektor v poljubnem koraku stohastičen in odziv nima analitične rešitve. V tem primeru se odziv simulacije v poljubnem koraku izračuna kot

$$p(\mathbf{f}_*|\mathbf{y}) \approx \frac{1}{S} \sum_{i=1}^S \int p(\mathbf{f}, \mathbf{f}_*|\mathbf{y}, \mathbf{z}_*)d\mathbf{f}, \quad (14)$$

kjer  $S$  predstavlja število vzorcev iz stohastičnega vhodnega vektorja  $\mathbf{z}_*$ . Model gaussovskih procesov ima kubično računsko kompleksnost, ki se lahko izboljša z aproksimacijskimi metodami.

## 2.3 Stohastični variacijski gaussovski proces

Stohastični variacijski gaussovski proces je aproksimacijska metoda za znižanje računske kompleksnosti modela na podlagi gaussovskih procesov. Predpostavimo, da obstaja vektor vhodnih psevdoloških lokacij  $\mathbf{z}_u$  in pripadajoči psevdološki vektor latentnih spremenljivk  $\mathbf{u} = [u_1, \dots, u_m]$ , ki je skupno gaussovsko porazdeljen z naključnima spremenljivkama  $\mathbf{f}$  in  $\mathbf{f}_*$ . Predpostavimo, da sta  $\mathbf{f}$  in  $\mathbf{f}_*$  pogojno neodvisna ob znanem  $\mathbf{u}$ . Aposteriorna skupna porazdelitev je potem definirana z

$$p(\mathbf{f}, \mathbf{f}_*, \mathbf{u}, \mathbf{y}) = p(\mathbf{f}|\mathbf{u}, \mathbf{y})p(\mathbf{f}_*|\mathbf{u})p(\mathbf{u}). \quad (15)$$

Hensman [5] definira variacijski približek kot

$$q(\mathbf{f}, \mathbf{f}_*, \mathbf{u}) = p(\mathbf{f}|\mathbf{u})p(\mathbf{f}_*|\mathbf{u})q(\mathbf{u}), \quad (16)$$

kjer je  $q(\mathbf{u})$  gaussovsko porazdelitev definirana s  $q(\mathbf{u}) \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Lambda})$ . Spodnja meja marginalne pogojne porazdelitve je potem izračunana z minimizacijo divergence Kullback-Leibler (KL) med aposteriorno porazdelitvijo definirano z enačbo (15) in variacijskim približkom definiranim z enačbo (16) in je prikazana z enačbo (17).

$$\log p(\mathbf{y}) \geq \int [\log p(\mathbf{y}|\mathbf{f})]q(\mathbf{f})d\mathbf{f} - \text{KL}[q(\mathbf{u})||p(\mathbf{u})]. \quad (17)$$

Spodnja meja marginalne pogojne porazdelitve se lahko poenostavi. Izraz  $p(\mathbf{y}|\mathbf{f})$  se faktorizira,  $q(\mathbf{f}_i)$  pa je odvisen le od pripadajočih vhodov  $\mathbf{z}_i$ . Poenostavljena spodnja meja marginalne pogojne porazdelitve je potem definirana kot

$$\log p(\mathbf{y}) \geq \sum_{n=1}^N \int [\log p(y_n|f_n)] q(f_n) df - \text{KL}[q(\mathbf{u})||p(\mathbf{u})]. \quad (18)$$

Spodnja meja marginalne pogojne porazdelitve se lahko izračuna stohastično, kjer naključno vzamemo le podmnožico originalnih podatkov za izračun ene iteracije optimizacije (ang. batch optimisation). Porazdelitev vektorja napovedi testnih latentnih funkcijskih vrednosti ob znanem  $\mathbf{y}$  je izračunana kot

$$q(\mathbf{f}_*) = \int p(\mathbf{f}_*|\mathbf{u})q(\mathbf{u})d\mathbf{u}. \quad (19)$$

Simulacijski odziv se izračuna z enačbo (14).

### 3 Globoki gausovski proces

Fleksibilnost stohastičnega gaussovskega procesa je lahko omejena z izbiro kovarične funkcije. Kovarična funkcija definira družino funkcij, ki jih lahko modeliramo. Na ta način vnesemo apriori znanje o modelirani preslikavi, posledično pa dodamo tudi omejitve, če o procesu ne vemo nič. Globoki gausovski procesi podobno kot globoke nevronske mreže definirajo hierarhično strukturo in povečajo fleksibilnost gaussovskiega procesa. V [7] definirajo rekurzivno apriori porazdelitev nad matriko latentnih vrednosti funkcije  $\mathbf{F}$  in opazovano matriko meritev  $\mathbf{Y}$ , ki sta lahko poljubnih dimenzij. Tak proces ima skupno porazdelitev definirano kot

$$p(\mathbf{Y}, \{\mathbf{F}^l, \mathbf{U}^l\}_{l=1}^L) = p(\mathbf{y}_i|\mathbf{f}_i^L) \prod_{l=1}^L p(\mathbf{F}^l|\mathbf{F}^{l-1}, \mathbf{U}^l)p(\mathbf{U}^l), \quad (20)$$

kjer  $\mathbf{F}^l$  in  $\mathbf{U}^l$  predstavljata vektorja latentnih in psevdofunkcijskih vrednosti v nivoju  $l$  in velja  $\mathbf{F}^0 = \mathbf{Z}$ . Ker ta model nima analitične rešitve, v [6] uporabijo variacijski približek skupne porazdelitve definiran kot

$$q(\{\mathbf{F}^l, \mathbf{U}^l\}_{l=1}^L) = \prod_{l=1}^L p(\mathbf{F}^l|\mathbf{F}^{l-1}, \mathbf{U}^l)q(\mathbf{U}^l), \quad (21)$$

kjer  $q(\mathbf{U}^l) \sim \mathcal{N}(\mathbf{m}^l, \mathbf{\Lambda}^l)$  predstavlja gaussovsko porazdelitev v nivoju  $l$ . Spodnja meja marginalne pogojne porazdelitve je izračunana z minimizacijo divergencije KL med  $p(\mathbf{Y}, \{\mathbf{F}^l, \mathbf{U}^l\}_{l=1}^L)$  in  $q(\{\mathbf{F}^l, \mathbf{U}^l\}_{l=1}^L)$  in je definirana z

$$\log p(\mathbf{Y}) \geq \sum_{i=1}^N \int [\log p(y_i|\mathbf{f}_i^L)] q(\mathbf{f}_i^L) - \sum_{l=1}^L \text{KL}[q(\mathbf{U}^l)||p(\mathbf{U}^l)]. \quad (22)$$

Spodnja meja ima podobne lastnosti kot meja v enačbi (18), kjer so  $i$ -te marginalne porazdelitve v zadnjem nivoju odvisne le od  $i$ -tih marginalnih porazdelitev v spodnjih nivojih. Zaradi tega lahko pri optimizaciji uporabimo le naključno podmnožico celotnega učnega vzorca v vsaki iteraciji. Napoved se izračuna z gaussovsko mešanico in je definirana z

$$q(\mathbf{f}_*^L) = \frac{1}{S} \sum_{s=1}^S q(\mathbf{f}_*^L|\mathbf{f}_*^{(s)L-1}), \quad (23)$$

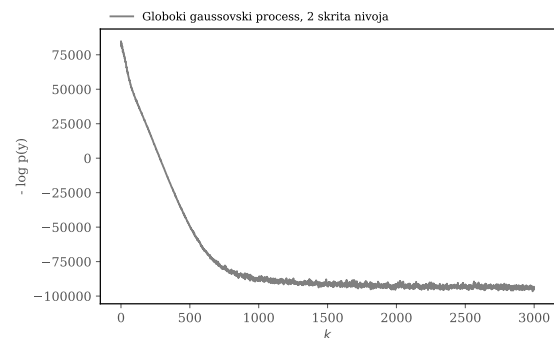
kjer vzamemo  $S$  vzorcev naključne spremenljivke  $\mathbf{f}_*^{L-1}$ . Simulacijski odziv se podobno kot pri navadnem gaussovskem procesu izračuna z enačbo (14).

## 4 Globok gausovski proces za identifikacijo procesa Bouc-Wen

V tem poglavju so predstavljeni rezultati modeliranja procesa Bouc-Wen. Primerjali smo delovanje globokih gaussovskih procesov pri različnih vrednostih števila skritih nivojev in preizkusili ali hierarhična struktura izboljša rezultate napovedovanja in simulacije na testnih podatkih.

### 4.1 Proces Bouc-Wen

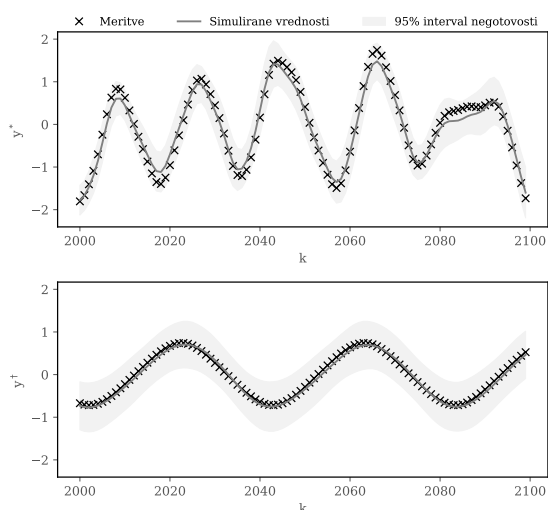
Podatkovni set procesa Bouc-Wen predstavlja vibracije Bouc-Wen sistema, ki vsebuje histerezo. Podrobnejše informacije o problemu so predstavljene v [8]. Učna množica je sestavljena iz vzburjanja s sinusnim vektorjem različnih frekvenc (ang. multi-sine)  $u(k)$  in izhodnim vektorjem  $y(k)$ , kjer je  $k = 1, \dots, N$  in  $N = 40,960$ . Prva testna množica je dobljena iz vzburjanja s sinusnim vektorjem različnih frekvenc (ang. multi-sine) in vsebuje 8,192 vzorcev. Druga testna množica je dobljena iz sinusnega signala s spreminjajočo se frekvenco (ang. sine-sweep) in vsebuje 153,000 vzorcev. Parametri avtoregresijskega modela so bili izbrani kot  $n_k = 0, n_b = 5, n_p = 1, n_a = 5$ . Za kovarično funkcijo smo izbrali funkcijo RBF in 100 psevdovhodov, ki so bili inicializirani z rojenjem kot je predlagano v [9]. V vsakem koraku optimizacije smo uporabili samo naključnih 1000 točk iz celotne množice podatkov.



Slika 1: Potek marginalne pogojne porazdelitve v odvisnosti od koraka optimizacije za globok gausovski proces z dvema skritima nivojema.

Tabela 1: Koren srednje vrednosti kvadrata napake za napoved (ang. rooted mean squared error [RMSE]) in simulacijo za proces Bouc-Wen, kjer  $RMSE^*$  predstavlja napako pri sinusnem vzbujanju različnih frekvenc,  $RMSE^\dagger$  pa napako pri sinusnem vzbujanju spreminjajočih se frekvenc. Model z enim skritim nivojem je ekvivalenten stohastičnemu variacijskemu gaussovskemu procesu. Rezultati za primerjavo z drugimi metodami identifikacije procesa Bouc-Wen so na voljo v [11].

Št. nivojev	Napovedni $RMSE^* [\times 10^{-6}]$	Simulacijski $RMSE^* [\times 10^{-5}]$	Napovedni $RMSE^\dagger [\times 10^{-6}]$	Simulacijski $RMSE^\dagger [\times 10^{-5}]$
1	$7.06 \pm 0.18$	$13.38 \pm 0.67$	$4.66 \pm 0.27$	$8.75 \pm 0.91$
2	<b><math>6.10 \pm 0.15</math></b>	<b><math>12.18 \pm 0.53</math></b>	<b><math>3.23 \pm 0.23</math></b>	<b><math>6.77 \pm 1.16</math></b>
3	$6.46 \pm 0.33$	$16.1 \pm 1.82$	$4.44 \pm 0.58$	$15.21 \pm 2.94$
4	$6.99 \pm 0.32$	$19.5 \pm 2.13$	$5.41 \pm 0.63$	$19.54 \pm 3.08$



Slika 2: Simulacija procesa Bouc-Wen za prvo testno množico, ki je vzbujana s sinusnim signalom različnih frekvenc ( $y^*$ ) in simulacija procesa Bouc-Wen za drugo testno množico, ki je vzbujana s sinusnim signalom s spreminjajočo se frekvenco ( $y^\dagger$ ).

### Rezultati

Slika 1 prikazuje logaritem marginalne pogojne porazdelitve v odvisnosti od koraka optimizacije za najboljšo rešitev, kjer smo za optimizacijski algoritem uporabili Adam [10]. Tabela 1 predstavlja rezultate korena srednje vrednosti kvadrata napake in standardni odklon dotične mere napake za obe testni množici podatkov pri čemer so bili rezultati pridobljeni z učenjem modela s 30 ključnimi ponovitvami. Vidimo lahko, da globoki gausovski proces z 2 skritima nivojema deluje najbolje tako za napoved kot za simulacijo na prvi testni množici podatkov s sinusnim vzbujanjem različnih frekvenc. Tudi za drugo testno množico podatkov (vzbujane s sinusnim signalom s spreminjajočo se frekvenco) najbolje deluje globoki gausovski proces z dvema skritima nivojema. Časovna odziva simulacije sta prikazana na sliki 2. Zanimiva opazka je, da se s povečevanjem števila skritih nivojev (večjega od 2) napaka povečuje. Razlog za to je lahko velika kompleksnost modela, kjer zaradi približka analitične rešitve potencialno ne dosežemo teoretične zmoglosti modela. Drugi problem lahko predstavlja občutljivost modela na parametre optimizacije zaradi katerih lahko pristanemo v loknem minimumu cenilke, ki jo optimiziramo.

## 5 Zaključek

Globoki gausovski proces predstavlja fleksibilno strukturo modela za modeliranje z gaussovskimi procesi. Posledično lahko modelira bolj kompleksne procese na račun večje računalniške zahtevnosti. Rezultati vrednotenja globokih gaussovskih procesov za modeliranje dinamičnih sistemov so pozitivni. Spoznali smo, da globoki gausovski procesi v kombinaciji z avtoregresijskim modelom izboljšajo sposobnost modeliranja v primerjavi s stohastičnim variacijskim gaussovskim procesom, ki predstavlja ekvivalent globokega gaussovskega procesa z enim skritim nivojem. V prihodnosti želimo testirati avtoregresijske modele globokih gaussovskih procesov na bolj kompleksnih primerih. Zanimiva je tudi ideja o bolj prilagojenih strukturah globokih gaussovskih procesov za modeliranje dinamičnih sistemov.

## Literatura

- [1] C. E. Rasmussen: Evaluation of Gaussian Processes and other Methods for Non-Linear Regression, University of Toronto, 1999.
- [2] C. K. Williams, C. E. Rasmussen, Gaussian Processes for Machine Learning, no. 3, MIT press Cambridge, MA, 2006.
- [3] J. Kocijan, Modelling and Control of Dynamic Systems using Gaussian Process Models, Springer, 2016.
- [4] J. Quiñero-Candela, C. E. Rasmussen, A Unifying View of Sparse Approximate Gaussian Process Regression, Journal of Machine Learning Research 6 (Dec) (2005) 1939–1959.
- [5] J. Hensman, A. Matthews, Z. Ghahramani, Scalable Variational Gaussian Process Classification, JMLR, 2015.
- [6] H. Salimbeni, M. Deisenroth. Doubly stochastic variational inference for Deep Gaussian Processes, Advances in Neural Information Processing Systems, 2017.
- [7] A. Damianou, N. Lawrence. Deep gaussian processes, Artificial Intelligence and Statistics. 2013.
- [8] J. Noël, M. Schoukens, Hysteretic Benchmark with a Dynamic Nonlinearity, in: Workshop on nonlinear system identification benchmarks, 2016, pp. 7–14.
- [9] M. Bauer, M. van der Wilk, C. E. Rasmussen, Understanding Probabilistic Sparse Gaussian Process Approximations, in: Advances in neural information processing systems, 2016, pp. 1533–1541.
- [10] D. P. Kingma, J. Ba, Adam, A Method for Stochastic Optimization, arXiv preprint arXiv:1412.6980.
- [11] Nonlinear System Identification Benchmarks, <http://www.nonlinearbenchmark.org/>