Supplementary derivations for

”Online Bayesian NARMAX identification: a free energy minimisation approach”

Wouter Kouw

March 19, 2021

NARMAX system

Let y_k be an output of and u_k be an input to a system at time k. Consider the following set of dynamics:

$$y_k = f_\theta \left(y_{k-1}, \dots, y_{k-M_1}, u_k, u_{k-1}, \dots, u_{k-M_2}, e_{k-1}, \dots, e_{k-M_3}\right) + e_k \qquad (1)$$

where $e_k \sim \mathcal{N}(0, \gamma^{-1})$ is a zero mean zero auto-correlation Gaussian noise component. The function f, parameterized by θ, is a nonlinear regression from input, output and noise components unto the current output. The constants M_1, M_2, and M_3 refer to the delays, for a total model order of $M = M_1 + 1 + M_2 + M_3$.

Generative model

We cast the NARMAX system equation to a likelihood:

$$\begin{aligned} p(y_k \mid \theta, y_{k-1}, \dots, y_{k-M_1}, u_k, \dots, u_{k-M_2}, \tau) = \\ \mathcal{N}\Big(y_k \mid f_\theta \left(y_{k-1}, \dots, y_{k-M_1}, u_k, \dots, u_{k-M_2}, e_{k-1}, \dots, e_{k-M_3}\right), \tau^{-1}\Big) . \end{aligned} \qquad (2)$$

Using the following priors

$$p(\theta) \triangleq \mathcal{N}\big(\theta \mid \mu_0, \Lambda_0^{-1}\big) , \qquad p(\tau) \triangleq \mathit{\Gamma}\big(\tau \mid \alpha_0, \beta_0\big) , \qquad (3)$$

we form the generative model for the entire time-series:

$$p(y_{1:T}, \theta, \tau \mid u_{1:T}) = \underbrace{p(\theta)p(\tau)}_{\text{priors}} \prod_{k=1}^{T} \underbrace{p(y_k \mid \theta, y_{k1}, .. y_{kM_1}, u_k, .. u_{kM_2}, \tau)}_{\text{likelihood}} . \qquad (4)$$

Recognition model

We employ a mean-field factorisation: $q(\theta, \tau) = q(\theta)q(\tau)$ where

$$q(\theta) \triangleq \mathcal{N}(\theta \mid \mu, \Lambda^{-1}) , \qquad q(\tau) \triangleq \Gamma(\tau \mid \alpha, \beta) . \tag{5}$$

Note that we employ mean-covariance and shape-rate parameterisations.

Message Computation

The NARMAX factor node sends out variational messages to its coefficients θ and its precision parameter τ. The general mathematical formula for a variational message in a factor graph is [21]:

$$\nu(x_i) \propto \exp\left(\mathbb{E}_{q(x_{j \neq i})}\left[\log p(x_i, \dots)\right]\right) , \tag{6}$$

where $p(x_i, \dots)$ represents the factor node function, which in our case is the likelihood (Equation 2).

Note that, in order to compute variational messages, we must be able to compute expected values with respect to the recognition distributions. In NARMAX models, the coefficients θ are part of a nonlinear function, which makes it challenging to compute expected values. In our paper, we employ a Taylor approximation of the nonlinear function f_θ to be able to compute expectations.

Taylor approximation

First-order Taylor approximation of f_θ at point μ:

$$f_\theta(x) \approx f_\mu(x) + J_\theta^\top(\theta - \mu) , \tag{7}$$

where J_θ represents the gradient of g with respect to θ evaluated at the approximating point:

$$J_\theta = \frac{\partial f_\theta(x)}{\partial \theta}|_{\theta=\mu} . \tag{8}$$

The expectation of the first moment of the function is:

$$\mathbb{E}_{q(\theta)}\left[f_\theta(x)\right] = f_\mu(x) + J_\theta^\top \underbrace{(\mu - \mu)}_{=0} = f_\mu(x) . \tag{9}$$

The expectation of the second moment of the function is:

$$
\begin{aligned}
&\mathbb{E}_{q(\theta)}\left[f_\theta(x)^2\right] \\
&\quad = \mathbb{E}_{q(\theta)}\Big(f_\mu(x) + J_\theta^\top(\theta-\mu)\Big)\Big(f_\mu(x) + J_\theta^\top(\theta-\mu)\Big) && \text{(10a)} \\
&\quad = \mathbb{E}_{q(\theta)}\Big(f_\mu(x)^2 + 2f_\mu(x)J_\theta^\top(\theta-\mu) + J_\theta^\top(\theta-\mu)(\theta-\mu)^\top J_\theta\Big) && \text{(10b)} \\
&\quad = f_\mu(x)^2 + 2f_\mu(x)J_\theta^\top(\mu-\mu) && \text{(10c)} \\
&\qquad\qquad + \mathbb{E}_{q(\theta)}\Big[J_\theta^\top\big(\theta\theta^\top - \mu\theta^\top - \theta\mu^\top + \mu\mu^\top\big)J_\theta\Big] && \text{(10d)} \\
&\quad = f_\mu(x)^2 + J_\theta^\top\big(\Lambda^{-1} + \mu\mu^\top - \mu\mu^\top - \mu\mu^\top + \mu\mu^\top\big)J_\theta && \text{(10e)} \\
&\quad = f_\mu(x)^2 + J_\theta^\top\Lambda^{-1}J_\theta && \text{(10f)}
\end{aligned}
$$

Polynomial In the special case of a polynomial function with basis expansion ϕ,

$$f_\theta(x) = \theta^\top\phi(x)\,, \tag{11}$$

the Taylor approximation defaults to:

$$
\begin{aligned}
\theta^\top\phi(x) &\approx \mu^\top\phi(x) + \frac{\partial\theta^\top\phi(x)}{\partial\theta}|_{\theta=\mu}(\theta-\mu) && \text{(12a)} \\
&= \mu^\top\phi(x) + \phi(x)^\top(\theta-\mu) && \text{(12b)} \\
&= \theta^\top\phi(x)\,. && \text{(12c)}
\end{aligned}
$$

Message to θ

In the following, I use the shorthand $f_\theta = f_\theta(y_{k-1}, \dots, e_{k-M_3})$ and ignore terms that do not depend on θ (denoted with C).

$$
\begin{aligned}
\log \nu(\theta) &= \mathbb{E}_{q(\tau)} \log \mathcal{N}\Big(y_k \mid f_\theta(y_{k-1}, \dots, e_{k-M_3}), \tau^{-1}\Big) + \mathsf{C} && \text{(13a)}\\
&= -\frac{1}{2}\mathbb{E}_{q(\tau)}[\tau]\Big(y_k - f_\theta\Big)^2 + \mathsf{C} && \text{(13b)}\\
&= -\frac{1}{2}\frac{\alpha}{\beta}\Big(y_k^2 - 2y_k f_\theta + f_\theta^2\Big) + \mathsf{C} && \text{(13c)}\\
&\approx -\frac{1}{2}\frac{\alpha}{\beta}\Big(y_k^2 - 2y_k[f_\mu + J_\theta^\top(\theta-\mu)] + [f_\mu + J_\theta^\top(\theta-\mu)]^2\Big) + \mathsf{C} && \text{(13d)}\\
&= -\frac{1}{2}\frac{\alpha}{\beta}\Big(y_k^2 - 2y_k f_\mu - 2y_k J_\theta^\top \theta + 2y_k J_\theta^\top \mu + \\
&\qquad f_\mu^2 + 2f_\mu J_\theta^\top(\theta-\mu) + J_\theta^\top(\theta-\mu)(\theta-\mu)^\top J_\theta\Big) + \mathsf{C} && \text{(13e)}\\
&= -\frac{1}{2}\frac{\alpha}{\beta}\Big(y_k^2 - 2y_k f_\mu - 2y_k J_\theta^\top \theta + 2y_k J_\theta^\top \mu + f_\mu^2 + 2f_\mu J_\theta^\top \theta \\
&\qquad - 2f_\mu J_\theta^\top \mu + J_\theta^\top(\theta\theta^\top - \mu\theta^\top - \theta\mu^\top + \mu\mu^\top)J_\theta\Big) + \mathsf{C} && \text{(13f)}\\
&= -\frac{1}{2}\frac{\alpha}{\beta}\Big(- 2y_k J_\theta^\top \theta + 2f_\mu J_\theta^\top \theta \\
&\qquad + J_\theta^\top \theta\theta^\top J_\theta - J_\theta^\top \mu\theta^\top J_\theta - J_\theta^\top \theta\mu^\top J_\theta\Big) + \mathsf{C} && \text{(13g)}\\
&= -\frac{1}{2}\frac{\alpha}{\beta}\Big(- 2(y_k - f_\mu)J_\theta^\top \theta + \theta^\top J_\theta J_\theta^\top \theta - 2J_\theta^\top \mu J_\theta^\top \theta\Big) + \mathsf{C} && \text{(13h)}\\
&= -\frac{1}{2}\Big(- 2\underbrace{\frac{\alpha}{\beta}(y_k - f_\mu + J_\theta^\top \mu)J_\theta^\top}_{Wm}\theta + \theta^\top \underbrace{(\frac{\alpha}{\beta}J_\theta J_\theta^\top)}_{W}\theta\Big) + \mathsf{C}\,. && \text{(13i)}
\end{aligned}
$$

We recognise both a linear function, $(Wm)\theta$, and a quadratic function, $\theta^\top W\theta$, in the log-domain. Consider for a moment a multivariate Gaussian probability density function, $\mathcal{N}(x \mid m, W^{-1})$, in the log-domain and ignore the normalisation terms as well as all terms in the exponent that don't depend on x:

$$
\log \mathcal{N}(x \mid m, W^{-1}) \propto -\frac{1}{2}\Big(- 2x^\top Wm + x^\top Wx\Big)\,. \tag{14}
$$

With this, we recognise a Gaussian distribution in Equation 13i. If we left-multiply Wm with the inverse precision W^{-1}, we obtain the mean; $m = W^{-1}Wm$. This yields:

$$
\nu(\theta) \propto \mathcal{N}\Big(\theta \mid (\frac{\alpha}{\beta}J_\theta J_\theta^\top)^{-1}(\frac{\alpha}{\beta}(y_k - f_\mu + J_\theta^\top \mu)J_\theta^\top), (\frac{\alpha}{\beta}J_\theta J_\theta^\top)^{-1}\Big)\,. \tag{15}
$$

Note that we stick to the mean-covariance parametrisation in our $\mathcal{N}(\cdot)$ notation. To be precise: W is the precision and W^{-1} is the covariance matrix.

Polynomial In the case of a polynomial f_θ, the term f_μ corresponds to $\mu^\top \phi(y_{k-1}, \dots)$ and the term $J_\theta^\top \mu$ corresponds to $\phi(y_{k-1}, \dots)^\top \mu$, which cancel out. Therefore, the message defaults to:

$$\nu(\theta) \propto \mathcal{N}\Big(\theta \mid \Big(\frac{\alpha}{\beta}\phi\phi^\top\Big)^{-1}\big(\frac{\alpha}{\beta} y_k \phi^\top\big), \frac{\alpha}{\beta}\phi\phi^\top\Big)\,, \tag{16}$$

where ϕ is short for $\phi(y_{k-1}, \dots)$.

Message to τ

$$\begin{aligned}
\log \nu(\tau) &= \mathbb{E}_{q(\theta)} \log \mathcal{N}\Big(y_k \mid f_\theta(y_{k-1}, \dots, e_{k-M_3}), \tau^{-1}\Big) + \mathsf{C} && \text{(17a)}\\
&= \frac{1}{2}\log\tau - \frac{1}{2}\tau \mathbb{E}_{q(\theta)}\Big[y_k^2 - 2y_k f_\theta + f_\theta^2\Big] + \mathsf{C} && \text{(17b)}\\
&= \frac{1}{2}\log\tau - \tau\frac{1}{2}\Big(y_k^2 - 2y_k \mathbb{E}_{q(\theta)}[f_\theta] + \mathbb{E}_{q(\theta)}[f_\theta^2]\Big) + \mathsf{C} && \text{(17c)}\\
&= \underbrace{\frac{1}{2}}_{a-1}\log\tau - \tau\underbrace{\frac{1}{2}\Big(y_k^2 - 2y_k f_\mu + f_\mu^2 + J_\theta^\top \Lambda^{-1} J_\theta\Big)}_{b} + \mathsf{C}\,. && \text{(17d)}
\end{aligned}$$

Consider the Gamma probability density function, with a shape-rate parameterisation, in the log-domain:

$$\log \Gamma(x \mid a, b) = (a-1)\log x - xb + \mathsf{C}\,. \tag{18}$$

We recognise a log-term and a linear term in Equation 17d and can therefore say that the variational message towards the variable τ is proportional to a Gamma distribution:

$$\nu(\tau) \propto \Gamma\Big(\tau \mid \frac{3}{2}, \frac{1}{2}\Big(y_k^2 - 2y_k f_\mu + f_\mu^2 + J_\theta^\top \Lambda^{-1} J_\theta\Big)\Big)\,. \tag{19}$$

Polynomial The rate parameter of the message takes the following form for a polynomial f_θ:

$$\frac{1}{2}\Big(y_k^2 - 2y_k(\mu^\top\phi) + (\mu^\top\phi)^2 + \phi^\top\Lambda^{-1}\phi\Big)\,, \tag{20}$$

where ϕ is short for $\phi(y_{k-1}, \dots)$.

Free Energy Computation

The FE objective is defined as the KL-divergence between the recognition model and the generative model:

$$\mathcal{F}[q] = \iint q(\theta,\tau) \log \frac{q(\theta,\tau)}{p(y_{1:T}, u_{1:T}, \theta, \tau)} \, \mathrm{d}\theta \mathrm{d}\tau \tag{21a}$$

$$\begin{aligned} &= \underbrace{\iint q(\theta,\tau) \big[-\log p(y_{1:T}, u_{1:T} \mid \theta, \tau) \big] \, \mathrm{d}\theta \mathrm{d}\tau}_{\text{Energy of likelihood}} \\ &\quad + \underbrace{\iint q(\theta,\tau) \log p(\theta,\tau) \, \mathrm{d}\theta \mathrm{d}\tau}_{\text{Energy of priors}} \\ &\quad + \underbrace{\iint q(\theta,\tau) \log q(\theta,\tau) \, \mathrm{d}\theta \mathrm{d}\tau}_{\text{Entropy of variables}} \, . \end{aligned} \tag{21b}$$

Below, we derive each of these terms separately.

Energy of likelihood The energy of the likelihood is:

$$\iint q(\theta,\tau) \, \big[-\log p(y_{1:T}, u_{1:T} \mid \theta, \tau) \big] \, \mathrm{d}\theta \mathrm{d}\tau$$

$$= \mathbb{E}_{q(\theta)q(\tau)} \big[-\log \mathcal{N}(y_k \mid f_\theta(y_{k-1}, \dots), \tau^{-1}) \big] \tag{22a}$$

$$= \frac{1}{2} \log 2\pi - \frac{1}{2} \mathbb{E}_{q(\tau)} \big[\log \tau \big] + \frac{1}{2} \mathbb{E}_{q(\tau)} \big[\tau \big] \mathbb{E}_{q(\theta)} \big[y_k^2 - 2 y_k f_\theta + f_\theta^2 \big] \tag{22b}$$

$$\begin{aligned} &= \frac{1}{2} \log 2\pi - \frac{1}{2} \big(\psi(\alpha) - \log(\beta) \big) \\ &\quad + \frac{1}{2} \frac{\alpha}{\beta} \Big(y_k^2 - 2 y_k f_\mu + f_\mu^2 + J_\theta^\top \Lambda^{-1} J_\theta \Big) , \end{aligned} \tag{22c}$$

where $\psi(\cdot)$ refers to the digamma function.

Energies of priors The priors are independent of each other and split into two distributions. Therefore, the energy of the priors splits into two as well:

$$\iint q(\theta,\tau) \log p(\theta,\tau) \, \mathrm{d}\theta \mathrm{d}\tau = \iint q(\theta) q(\tau) \log p(\theta) p(\tau) \, \mathrm{d}\theta \mathrm{d}\tau \tag{23a}$$

$$= \int q(\theta) \log p(\theta) \, \mathrm{d}\theta + \int q(\tau) \log q(\tau) \, \mathrm{d}\tau \, . \tag{23b}$$

If we plug in the parameterisations of the prior defined in Equation 3, then we get:

$$\int q(\theta) \log \mathcal{N}\big(\theta \mid \mu_0, \Lambda_0^{-1}\big)\, \mathrm{d}\theta \tag{24a}$$

$$= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\det(\Lambda_0^{-1}) - \frac{1}{2}\mathbb{E}_{q(\theta)}\big[(\theta-\mu_0)^\top \Lambda_0^{-1}(\theta-\mu_0)\big] \tag{24b}$$

$$\begin{aligned} = &-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\det(\Lambda_0^{-1}) \\ &- \frac{1}{2}\mathbb{E}_{q(\theta)}\big(\theta^\top \Lambda_0^{-1}\theta - \mu_0^\top \Lambda_0^{-1}\theta - \theta^\top \Lambda_0^{-1}\mu_0 + \mu_0^\top \Lambda_0^{-1}\mu_0\big) \end{aligned} \tag{24c}$$

$$\begin{aligned} = &-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\det(\Lambda_0^{-1}) \\ &- \frac{1}{2}\Big(\mathrm{tr}\big(\Lambda_0^{-1}(\Lambda^{-1}+\mu\mu^\top)\big) - \mu_0^\top \Lambda_0^{-1}\mu - \mu^\top \Lambda_0^{-1}\mu_0 + \mu_0^\top \Lambda_0^{-1}\mu_0\Big)\,. \end{aligned} \tag{24d}$$

and

$$\int q(\tau) \log \mathit{\Gamma}\big(\tau \mid \alpha_0, \beta_0\big)\, \mathrm{d}\tau \tag{25a}$$

$$= -\log \mathbf{\Gamma}(\alpha_0) + \alpha_0 \log\beta_0 + (\alpha_0 - 1)\mathbb{E}_{q(\tau)}[\log\tau] - \beta_0\mathbb{E}_{q(\tau)}[\tau] \tag{25b}$$

$$= -\log \mathbf{\Gamma}(\alpha_0) + \alpha_0 \log\beta_0 + (\alpha_0 - 1)(\psi(\alpha) - \log\beta) - \beta_0\frac{\alpha}{\beta}\,. \tag{25c}$$

where $\mathbf{\Gamma}(\cdot)$ refers to the gamma function, not the Gamma distribution.

Entropies of variables The entropies of the recognition factors, as defined in Equation 5, can be looked up. They are:

$$\int q(\theta)\log q(\theta)\, \mathrm{d}\theta = -H_q[\theta] = -\big[\frac{1}{2}\log\det(2\pi\mathbf{e}\Lambda^{-1})\big] \tag{26a}$$

$$\int q(\tau)\log q(\tau)\, \mathrm{d}\tau = -H_q[\tau] = -\big[\alpha - \log\beta + \log\mathbf{\Gamma}(\alpha) + (1-\alpha)\psi(\alpha)\big]\,. \tag{26b}$$

Note that $\mathbf{e}$ refers to Euclid's number, or $\exp(1)$.

References

[1] S. A. Billings, *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.

[2] H.-S. Tang, S.-T. Xue, R. Chen, and T. Sato, "Online weighted LS-SVM for hysteretic structural system identification," *Engineering Structures*, vol. 28, no. 12, pp. 1728–1735, 2006.

[3] P. Kshirsagar, D. Jiang, and Z. Zhang, "Implementation and evaluation of online system identification of electromechanical systems using adaptive filters," *IEEE Transactions on Industry Applications*, vol. 52, no. 3, pp. 2306–2314, 2016.

[4] *Preparing for the unknown: Learning a universal policy with online system identification*, 2017.

[5] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013, vol. 3.

[6] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[7] V. Peterka, "Bayesian approach to system identification," in *Trends and Progress in System identification*. Elsevier, 1981, pp. 239–304.

[8] T. Hill, P. Green, A. Cammarano, and S. Neild, "Fast Bayesian identification of a class of elastic weakly nonlinear systems using backbone curves," *Journal of Sound and Vibration*, vol. 360, pp. 156–170, 2016.

[9] T. B. Schön, F. Lindsten, J. Dahlin, J. Wågberg, C. A. Naesseth, A. Svensson, and L. Dai, "Sequential Monte Carlo methods for system identification," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 775–786, 2015.

[10] J. Schoukens and L. Ljung, "Nonlinear system identification: A user-oriented road map," *IEEE Control Systems Magazine*, vol. 39, no. 6, pp. 28–99, 2019.

[11] J. N. Hendriks, F. K. Gustafsson, A. H. Ribeiro, A. G. Wills, and T. B. Schön, "Deep energy-based NARX models," *arXiv:2012.04136*, 2020.

[12] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[13] Y. Lu, S. Khatibisepehr, and B. Huang, "A variational Bayesian approach to identification of switched ARX models," in *IEEE Conference on Decision and Control*, 2014, pp. 2542–2547.

[14] W. R. Jacobs, T. Baldacchino, T. Dodd, and S. R. Anderson, "Sparse Bayesian nonlinear system identification using variational inference," *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4172–4187, 2018.

[15] K. J. Friston, J. Daunizeau, J. Kilner, and S. J. Kiebel, "Action and behavior: a free-energy formulation," *Biological Cybernetics*, vol. 102, no. 3, pp. 227–260, 2010.

[16] A. Imohiosen, J. Watson, and J. Peters, "Active inference or control as inference? A unifying view," in *International Workshop on Active Inference*. Springer, 2020, pp. 12–19.

[17] A. A. Meera and M. Wisse, "Free energy principle based state and input observer design for linear systems with colored noise," in *American Control Conference*. IEEE, 2020, pp. 5052–5058.

[18] W. M. Kouw, "Online system identification in a Duffing oscillator by free energy minimisation," in *International Workshop on Active Inference*. Springer, 2020, pp. 42–51.

[19] M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, University College London, 2003.

[20] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2282–2312, 2005.

[21] J. Dauwels, "On variational message passing on factor graphs," in *IEEE International Symposium on Information Theory*, 2007, pp. 2546–2550.

[22] S. Korl, "A factor graph approach to signal modelling, system identification and filtering," Ph.D. dissertation, ETH Zurich, 2005.

[23] M. Cox, T. van de Laar, and B. de Vries, "Forneylab.jl: Fast and flexible automated inference through message passing in julia," in *International Conference on Probabilistic Programming*, 2018.

[24] A. Podusenko, W. M. Kouw, and B. de Vries, "Online variational message passing in hierarchical autoregressive models," in *IEEE International Symposium on Information Theory*, 2020, pp. 1343–1348.

[25] I. Senoz, A. Podusenko, W. M. Kouw, and B. de Vries, "Bayesian joint state and parameter tracking in autoregressive models," in *Conference on Learning for Dynamics and Control*, 2020, pp. 1–10.

[26] D. Khandelwal, M. Schoukens, and R. Tóth, "On the simulation of polynomial narmax models," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 1445–1450.

[27] M. H. Hayes, *Statistical digital signal processing and modeling*. John Wiley & Sons, 2009.

[28] R. Pintelon and J. Schoukens, *System identification: a frequency domain approach*. John Wiley & Sons, 2012.

[29] T. Wigren and J. Schoukens, "Three free data sets for development and benchmarking in nonlinear system identification," in *European Control Conference (ECC)*, 2013, pp. 2933–2938.

[30] J. Paduart, L. Lauwers, J. Swevers, K. Smolders, J. Schoukens, and R. Pintelon, "Identification of nonlinear systems using polynomial nonlinear state space models," *Automatica*, vol. 46, no. 4, pp. 647–656, 2010.

[31] D. Khandelwal, "Automating data-driven modelling of dynamical systems: an evolutionary computation approach," Ph.D. dissertation, TU Eindhoven, 2020.

[32] J. Daunizeau, K. J. Friston, and S. J. Kiebel, "Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models," *Physica D: Nonlinear Phenomena*, vol. 238, no. 21, pp. 2089–2118, 2009.