

OPEN ACCESS

EDITED BY

Alberto Mazzoni,
Sant'Anna School of Advanced Studies,
Italy

REVIEWED BY

Ali Yousefi,
Worcester Polytechnic Institute,
United States
Jens-Oliver Muthmann,
The University of Texas at Austin,
United States

*CORRESPONDENCE

Lazaros Mitskopoulos
l.mitskopoulos@sms.ed.ac.uk

SPECIALTY SECTION

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

RECEIVED 31 March 2022

ACCEPTED 05 September 2022

PUBLISHED 23 September 2022

CITATION

Mitskopoulos L, Amvrosiadis T and
Onken A (2022) Mixed vine copula
flows for flexible modeling of neural
dependencies.
Front. Neurosci. 16:910122.
doi: 10.3389/fnins.2022.910122

COPYRIGHT

© 2022 Mitskopoulos, Amvrosiadis and
Onken. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Mixed vine copula flows for flexible modeling of neural dependencies

Lazaros Mitskopoulos^{1*}, Theoklitos Amvrosiadis² and Arno Onken¹

¹School of Informatics, Institute for Adaptive and Neural Computation, University of Edinburgh, Edinburgh, United Kingdom, ²Centre for Discovery Brain Sciences, Edinburgh Medical School: Biomedical Sciences, University of Edinburgh, Edinburgh, United Kingdom

Recordings of complex neural population responses provide a unique opportunity for advancing our understanding of neural information processing at multiple scales and improving performance of brain computer interfaces. However, most existing analytical techniques fall short of capturing the complexity of interactions within the concerted population activity. Vine copula-based approaches have shown to be successful at addressing complex high-order dependencies within the population, disentangled from the single-neuron statistics. However, most applications have focused on parametric copulas which bear the risk of misspecifying dependence structures. In order to avoid this risk, we adopted a fully non-parametric approach for the single-neuron margins and copulas by using Neural Spline Flows (NSF). We validated the NSF framework on simulated data of continuous and discrete types with various forms of dependency structures and with different dimensionality. Overall, NSFs performed similarly to existing non-parametric estimators, while allowing for considerably faster and more flexible sampling which also enables faster Monte Carlo estimation of copula entropy. Moreover, our framework was able to capture low and higher order heavy tail dependencies in neuronal responses recorded in the mouse primary visual cortex during a visual learning task while the animal was navigating a virtual reality environment. These findings highlight an often ignored aspect of complexity in coordinated neuronal activity which can be important for understanding and deciphering collective neural dynamics for neurotechnological applications.

KEYWORDS

neural dependencies, higher-order dependencies, heavy tail dependencies, vine copula flows, Neural Spline Flows, mixed variables

1. Introduction

Coordinated information processing by neuronal circuits in the brain is the basis of perception and action. Neuronal ensembles encode sensory and behavior-related features in sequences of spiking activity which can exhibit rich dynamics at various temporal scales (Gao and Ganguli, 2015). Acquiring an understanding of how multivariate interactions in neural populations shape and affect information transmission is not only important for neural coding theory but will also inform methodological frameworks for

clinically translatable technologies such as Brain Computer Interfaces (BCIs). Both of these research programs have enjoyed a surge of activity as a result of recent advances in imaging technologies (Chen et al., 2012) and high-yield electrophysiology both for human (McFarland and Wolpaw, 2017) and animal studies (Jun et al., 2017). Brain Computer Interfaces can mediate neural signal transduction for moving prosthetic limbs or external robotic devices in paralyzed patients or they can aid communication with patients suffering from locked-in syndrome (Chaudhary et al., 2016). Therefore, successful clinical use relies on accurate reading and relaying of information content transmitted via population spiking responses. Doing so can be quite challenging from a data analytic perspective as moderate to high-dimensional brain activity can be considerably complex, exhibiting non-trivial multivariate neural dependencies (Hurwitz et al., 2021). Moreover, the resulting behavioral output variables (e.g., limb movement) might display vastly different statistics to neural variables like spike trains or event-related potentials (ERPs). These challenges highlight the importance of developing novel analytical tools that can handle the complexity within neural population activity and its relation to behavior. Such tools should also have broad applicability over different types of data (e.g., spike counts, local field potentials, EPRs).

The present need for novel methods stems from the fact that the majority of past work on neural dependencies has focused on pairwise shared response variability between neurons, also known as noise correlations (Zohary et al., 1994; Brown et al., 2004; Moreno-Bote et al., 2014; Kohn et al., 2016). Neural responses are known to exhibit considerable variability even when repeatedly presented with the same stimulus, but this might be part of collective dynamical patterns of activity in a neural population. The typical assumption in this line of research is that the noise in neural responses is Gaussian, and thus, firing rates are modeled with multivariate normal distributions where a certain covariance matrix specifies all pairwise linear correlations (Averbeck et al., 2006; Ecker et al., 2010). While this approach may provide a reasonable approximation for correlations in coarse time-scales, its validity can be disputed for spike-counts in finer time-scales. First of all, real spike counts are characterized by discrete distributions and they exhibit a positive skew instead of a symmetric shape (Onken et al., 2009). Also, spike data do not usually display elliptical dependence as in the normal distribution but tend to be heavy tailed (Kudryashova et al., 2022), which geometrically translates to having probability mass concentrated on one of the corners. Finally, although the multivariate normal distribution is characterized by only second-order correlations, recent studies have indicated that higher order correlations are substantially present in local cortical populations and have a significant effect on the informational content of encoding (Pillow et al., 2008; Ohiorhenuan et al., 2010; Yu et al., 2011; Shimazaki et al., 2012; Montangie and Montani, 2017) as well as on the performance

of decoding models (Michel and Jacobs, 2006). Therefore, dissecting the structure of multivariate neural interactions is important to the study of neural coding and clinical applications such as BCIs that rely on accurate deciphering of how neural activity translates to action. This calls for an alternative approach that goes beyond pairwise correlations.

A statistical tool which is suited for the study of multivariate dependencies is that of copulas. Copula-based approaches have enjoyed wide usage in economics for modeling risk in investment portfolios (Jaworski et al., 2012), but have received limited attention in neuroscience. Intuitively, copulas describe the dependence structures between random variables, and in conjunction with models of the individual variables, they can form a joint statistical model for multivariate observations. When observations come from continuous variables their associated copulas are unique, independent from marginal distributions, and invariant to strictly monotone transformations (Faugeras, 2017). However, data in neuroscience research are often discrete (e.g., spike counts) or they may contain interactions between discrete and continuous variables with vastly different statistics (e.g., spikes with local field potentials or behavioral measurements such as running speed or pupil dilation). Despite the indeterminacy of copulas in these cases, they are still valid constructions for discrete objects and mixed interactions and one can apply additional probabilistic tools to obtain consistent discrete copulas (Genest and Nešlehová, 2007). Previous work has successfully applied copula-based methods to discrete or mixed settings (Song et al., 2009; de Leon and Wu, 2011; Panagiotelis et al., 2012; Smith and Khaled, 2012; Onken and Panzeri, 2016) using copulas from parametric families that assume a certain type of interaction. Although parametric models are a powerful tool for inference, their application can bear a risk of misspecification by imposing rather rigid and limiting assumptions on the types of dependencies to be encountered within a dataset with heterogeneous variables or multiscale dynamical processes. This risk is especially amplified for dependencies between more than two variables as available multivariate copulas are quite limited in number and assume a particular type of dependence structure for all variables which can ignore potentially rich interaction patterns. As the set of commonly used bivariate parametric copulas is much larger, a common alternative is to decompose multivariate dependencies into a cascade of bivariate copulas organized into hierarchical tree structures called vines or pair copula constructions (Aas et al., 2009). Nodes of the vines correspond to conditional distributions and edges correspond to copulas that describe their interaction. This formulation allows for a flexible incorporation of various dependence structures in a joint model. Previous studies that employed vine copulas in mixed settings used parametric models (Song et al., 2009; de Leon and Wu, 2011; Panagiotelis et al., 2012; Smith and Khaled, 2012; Onken and Panzeri, 2016). For the present study, given the aforementioned

intricacies of neuronal spiking statistics, we aim to explore the potential of non-parametric methods as a more flexible alternative for estimating discrete and continuous vine copulas. Existing non-parametric approaches have focused on kernel-based methods (Racine, 2015; Geenens et al., 2017) or jittering and continuous convolutions with specified noise models to obtain pseudo-continuous data (Nagler et al., 2017; Schallhorn et al., 2017).

Another model-free method that shows promise is that of normalizing flows, a class of generative models for density estimation that allow for flexible sampling (Rezende and Mohamed, 2015; Papamakarios et al., 2021). Recently, some authors have attempted to employ normalizing flows for non-parametric modeling of copulas using simulated and standard benchmark datasets (Wiese et al., 2019; Kamthe et al., 2021). An application of these models to recordings from neural populations is still missing and has the potential to shed light on the structure of coordinated neural activity, thereby potentially improving BCIs that take this structure into account.

In this study, we aimed to conduct a thorough investigation into flow-based estimation of vine copulas with continuous and discrete artificial data in various settings with different but known dependence structures and number of variables. Furthermore, we sought to demonstrate the potential of this framework to elucidate interaction patterns within neural recordings that contain heavy tails and extend beyond bivariate dependencies. For this reason, we chose to investigate neural responses in the mouse V1 while the animal is navigating in a virtual reality environment. Studying neural interfaces in rodents has been important for pre-clinical testing of BCIs to probe potential limitations that can inform applications in humans (Widge and Moritz, 2014; Bridges et al., 2018). The test case we chose serves as a proof-of-concept study but it can also provide meaningful insights on how spatial navigation related cues and/or behavioral variables modulate visual activity, which can inform future clinical research on BCIs.

2. Materials and methods

2.1. Copulas

Multivariate neuronal interactions can be described probabilistically by means of copulas, which embody how spiking statistics of individual neurons, i.e., the marginal distributions, are entangled in intricate ways to produce the observed joint population activity. The central theoretical foundation for copula-based approaches is Sklar's theorem (Sklar, 1959). It states that every multivariate cumulative distribution function (CDF) F_x can be decomposed into its margins, in this case the single-neuron distributions F_1, \dots, F_d , and a copula C (Figure 1A) such that:

$$F_x(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (1)$$

Copulas are also multivariate CDFs with support on the unit hypercube and uniform margins and their shape describes the dependence structure between random variables in a general sense which goes beyond linear or rank correlations (Faugeras, 2017). Following Sklar's theorem, it is possible to obtain copulas from joint distributions using:

$$C(u_1, \dots, u_d) = F_x(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)), \quad (2)$$

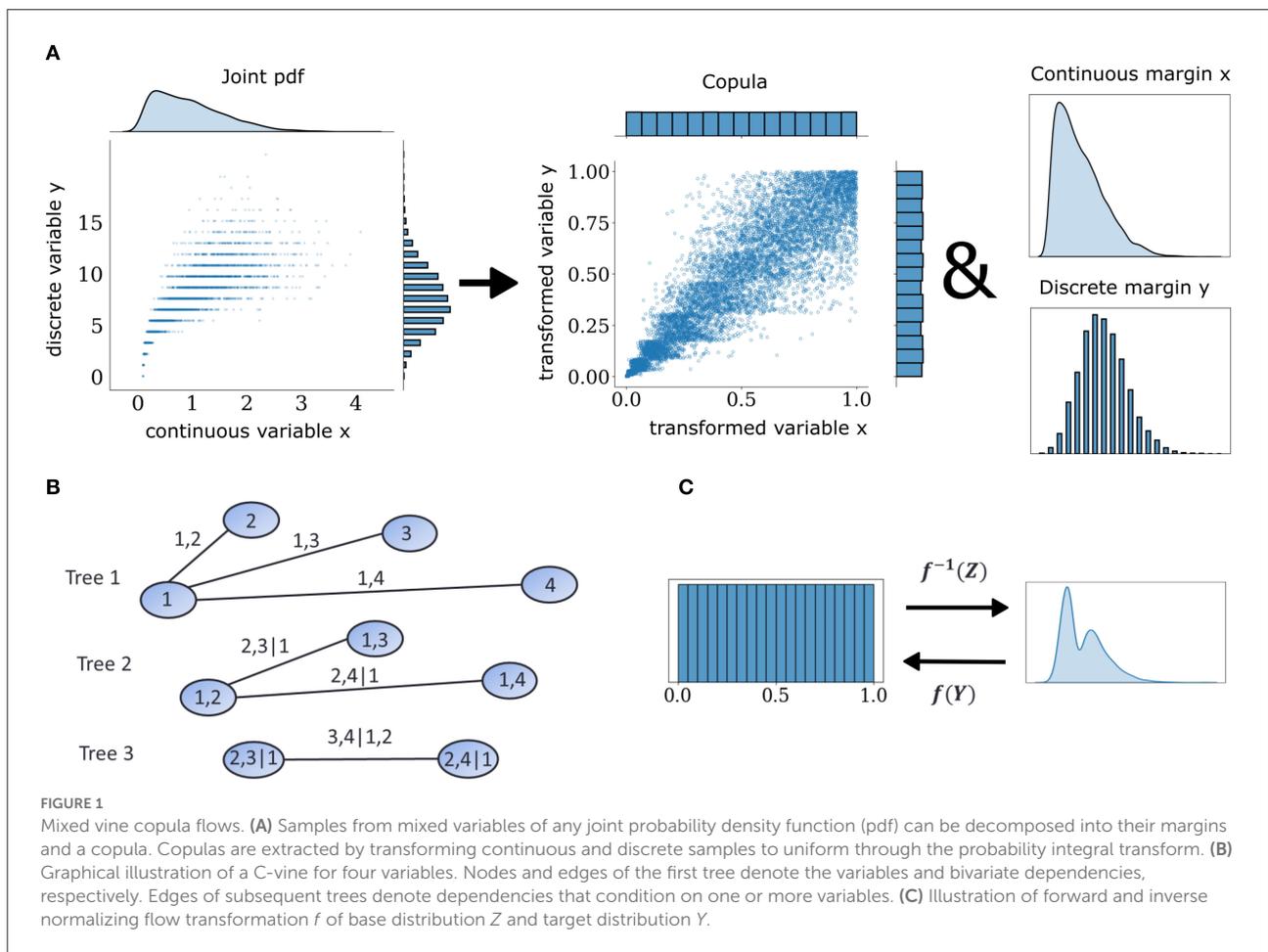
Conversely, it is also possible to construct proper joint distributions by entangling margins with copulas. These operations rely on the probability transform F and its generalized inverse, the quantile transform F^{-1} . The probability transform maps samples to the unit interval: $F(X) \rightarrow U \sim U_{[0,1]}$, where $U_{[0,1]}$ denotes the uniform distribution on the unit interval. Since copulas for discrete data depend on margins and are not unique (Genest and Nešlehová, 2007), additional tools are required to obtain consistent mapping to copula space. We employed the distributional transform:

$$G(X, V) = F_{x-}(x) + V(F_x(x) - F_{x-}(x)) \quad (3)$$

where $F_{x-}(x) = Pr(X < x)$ as opposed to the regular expression for the CDF, $F_x(x) = Pr(X \leq x)$, and V is a random variable uniformly distributed on $[0,1]$ independent of X . This extension to the probability transformation effectively converts a discrete copula to a pseudo-continuous variable by adding uniform jitter in between discontinuous intervals in the support of discrete variables and makes it possible to use non-parametric estimators designed for continuous data. When one is interested in the dependence structures within joint observations of multiple variables, copula densities are the object to work with, instead of their CDF. An example with continuous and discrete observations is illustrated in Figure 1A. This case of a mixed interaction is described by a Clayton copula which displays an asymmetric heavy tail. The empirical copula density can be discovered by subjecting the variables to the probability transform with an added distributional transform for the discrete one. This operation dissects the dependence information that is embedded in the joint probabilities of these two variables.

2.2. Pair copula constructions

The curse of dimensionality encountered in large datasets can pose considerable challenges for copula modeling. Pair copula constructions (Bedford and Cooke, 2002) offer a flexible way of scaling copula models by factorizing the multivariate distribution into a cascade of bivariate conditional distributions that can be described by bivariate copulas. The latter can be modeled parametrically, as previous studies have already done



(Panagiotelis et al., 2012; Onken and Panzeri, 2016) or with non-parametric tools, which is the present study's approach.

The space of possible factorizations is prohibitively large so vine copula structures, a special type of pair copula constructions can be employed to facilitate inference and sampling (Aas et al., 2009). They can be represented as hierarchical sets of trees which account for a specific graph of multivariate interactions among elements of the distributions and assume conditional independence for the rest. In the present study, we focused on the canonical vine or C-Vine (Figure 1B) in which each tree in the hierarchy has a node that serves as a connection hub to all other nodes. The C-Vine decomposes the joint distribution f into a product of its margins and conditional copulas c .

$$f_X(x_1, \dots, x_d) = \prod_{k=1}^d f(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|1, \dots, j-1}(F(x_j|x_1, \dots, x_{j-1}), F(x_{i+j}|x_1, \dots, x_{j-1})) \quad (4)$$

where $c_{i,j|A}$ denotes the pair copula between elements i and j given the elements in the set A , which is empty in the surface tree but it increases in number of elements with deeper trees.

2.3. Copula flows

We modeled the margin and copula densities non-parametrically using a specific class of normalizing flows, that is called Rational-Quadratic Neural Spline Flows (NSF) (Durkan et al., 2019). In general, normalizing flows are a class of generative models that construct arbitrary probability densities using smooth and invertible transformations to and from simple probability distributions (Rezende and Mohamed, 2015) (Figure 1C). In essence, this is an application of the change of variables formula:

$$p_x(x) = p_u(T^{-1}(x)) \det \left| \frac{\partial T^{-1}(x)}{\partial x} \right|, \quad (5)$$

where $p_x(x)$ is the density of the observations and p_u is the base density of random variable $U = T^{-1}(X)$, which is a known and convenient distribution such as the normal or the uniform distribution. The transformation T is usually a composition of invertible transformations that can be perceived and implemented as an artificial neural network with a certain number of layers and hidden units. Its parameters have to be learned through training in order to achieve an invertible mapping between the two distributions, while scaling

appropriately by the determinant of the Jacobian matrix which keeps track of volume changes induced by T .

Since our main goal was to apply normalizing flows on a copula-based framework for neural dependencies, it was natural to choose the uniform distribution on $[0,1]$ as a base distribution, so that backward and forward flow transformations for the margins approximate the probability transform and its inverse, the quantile transform, respectively, so as to map observations to copula space and back. Furthermore, a uniform base distribution for copula flows can be leveraged to generate new (simulated) observations via inverse sampling. Monte Carlo simulation with inverse sampling is much more flexible than with rejection sampling used in kernel-density based non-parametric estimators which are much harder to invert. We expected this flexibility to translate into faster sampling for NSF compared to the other estimators. Different types of normalizing flows exist in the literature which involve simple affine or more flexible non-affine transformations albeit with the cost of sacrificing invertibility in some cases (Papamakarios et al., 2021). Our choice of employing NSF (Durkan et al., 2019) in this study for modeling both margin and copula densities was in virtue of the fact that they combine the flexibility of non-affine flows while maintaining easy invertibility by approximating a quasi-inverse with piecewise spline-based transformations around knot points of the mapping.

2.4. Sequential estimation and model selection

To fit the C-vine model with NSF to data, we applied the Inference for Margins procedure, which first fits the margins and then fits the copulas deriving from pairs of margins and conditional margins. For the first step, we fit each margin with NSF and then proceeded with the copulas of a particular canonical vine formulation (Aas et al., 2009). Fitting NSF to bivariate copulas in the vine was conducted sequentially starting with the copulas of the surface layer of the tree. Subsequently, conditional marginals for the tree in the layer next in depth were estimated using the copulas of the previous layer via h-functions (Czado, 2019). Then, copulas for the conditional margins were constructed by transforming these margins in the same layer according to Equation (2). This procedure was followed until all copulas in the vine decomposition were estimated. We followed the simplifying assumption (Haff et al., 2010) according to which copulas of conditional margins are independent of the conditional variables but depend on them indirectly through the conditional distribution functions at the nodes of the vine. In other words, conditional margins do vary with respect to the conditioning variables but they are assumed to map to the same conditional copula. This assumption helps evade the curse of dimensionality when investigating multivariate

dependence structures. For each copula, we used a random search procedure (Bergstra and Bengio, 2012) to determine the best performing NSF hyperparameter configuration on a validation set containing 10% of the data. The hyperparameters that were tuned during training were the number of hidden layers and hidden units as well as the number of knots for the spline transforms. This sequential estimation and optimization scheme for NSF-based vine copulas was followed in both our analyses with artificial data as well as data from neuronal recordings.

2.5. Other non-parametric estimators

The other non-parametric estimators used for comparisons against NSF included four versions of Kernel Density Estimators (KDE), namely one with log-linear local likelihood estimation (tll1), one with log-quadratic local likelihood estimation (tll2), one with log-linear local likelihood estimation and nearest neighbor bandwidths (tll1nn), and one with log-quadratic local likelihood estimation and nearest neighbor bandwidths (tll2nn) (Nagler et al., 2017). Lastly, an estimator based on Bernstein polynomials (Sancetta and Satchell, 2004) was also used for the comparisons. The implementations for all these five non-parametric estimators are from kdecopula package (Nagler et al., 2017).

2.6. Artificial data

The NSF framework for vine copulas was validated on artificial data with known dependency structures and was compared against other non-parametric estimators. We constructed several test cases where data was continuous or discrete, consisting of four (low dimensional) or eight (higher dimensional) variables exhibiting weak or strong dependencies that were characterized by different copula families, namely either Clayton, Frank, or Gaussian copulas. These three types of parametric copulas display different behavior in the tail regions (Nelsen, 2007). Clayton copulas have an asymmetric tail dependence whereby probability mass is concentrated in one corner of the copula space indicating a single heavy tail region (see example copula in Figure 1A). On the contrary, Frank copulas do not have heavy tails and probability mass is allocated uniformly and symmetrically along the correlation path. Gaussian copulas are also symmetric and without particularly heavy tails, but probability mass concentration in the tail regions is larger compared to Frank copulas. The strength of dependencies was determined by varying the θ parameter for Clayton ($\theta = 2$ for weak and $\theta = 5$ for strong dependence) and Frank copulas ($\theta = 3$ for weak and $\theta = 7$ for strong dependence) and the off-diagonal entries of the 2×2 correlation matrix for Gaussian copulas (0.4

for weak dependence and 0.8 for strong dependence). We constructed and drew simulated samples from all vines with the aforementioned specifications using the mixed vines package developed by Onken and Panzeri (2016). Training for all the estimators was conducted with 5,000 simulated samples for each variable in the artificial data. The training procedure was repeated 10 times. Performance was measured with Kullback-Leibler (KL) divergences of 8,000 copula samples generated by each of the estimators to an equal number of ground truth copula samples from the mixed vines package (Onken and Panzeri, 2016). To estimate the KL divergences from sample data, a k-nearest neighbor algorithm was used (Wang et al., 2009), which was implemented in Hartland (2021). The resulting KL divergences from the 10 repetitions and the different copulas in each vine were aggregated to measure the overall performance of the framework in each test case. We statistically compared performances by all estimators via Kruskal-Wallis significance tests at level of significance equal to 0.05. Bonferonni correction was used for multiple comparisons. Moreover, we calculated copula entropies from the NSF copula densities via classical Monte Carlo estimation:

$$h(c(u_1, u_2)) = \mathbb{E}_c[-\log_2 c(u_1, u_2)] \approx -\frac{1}{K} \sum_{k=1}^K (\log_2 c(u_1^k, u_2^k)), \quad (6)$$

where h denotes the entropy and \mathbb{E}_c denotes the expectation with respect to the copula c . The expectation is approximated by summing over a K number of samples which needs to be sufficiently large ($K = 8,000$ in our study). Negative copula entropy provides an accurate estimate of mutual information between variables which does not depend on the marginal statistics of each variable (Sklar, 1959; Jenison and Reale, 2004) and is a measure of the degree to which knowing one variable can reduce the uncertainty of the other one. All analyses including sampling and entropy calculations were conducted on an ASUS laptop with Intel(R) 4 Cores, i5-8,300 CPU, 2.30 GHz.

2.7. Experimental data

In order to assess our framework's applicability to neuronal activity we used two-photon calcium imaging data of neuronal activity in the mouse primary visual cortex, that were collected at the Rochefort lab (see Henschke et al., 2020 for more details). Briefly, V1 layer 2/3 neurons labeled with the calcium indicator GCaMP6s were imaged while the animal was headfixed, freely running on a cylindrical treadmill and navigating through a virtual reality environment (160 cm). Mice were trained to lick at a specific location along the virtual corridor in order to receive a reward. In addition to neuronal activity, behavioral variables such as licking and running speed were monitored simultaneously. Over the course of 5 days, mice learned to lick within the reward zone to receive the water

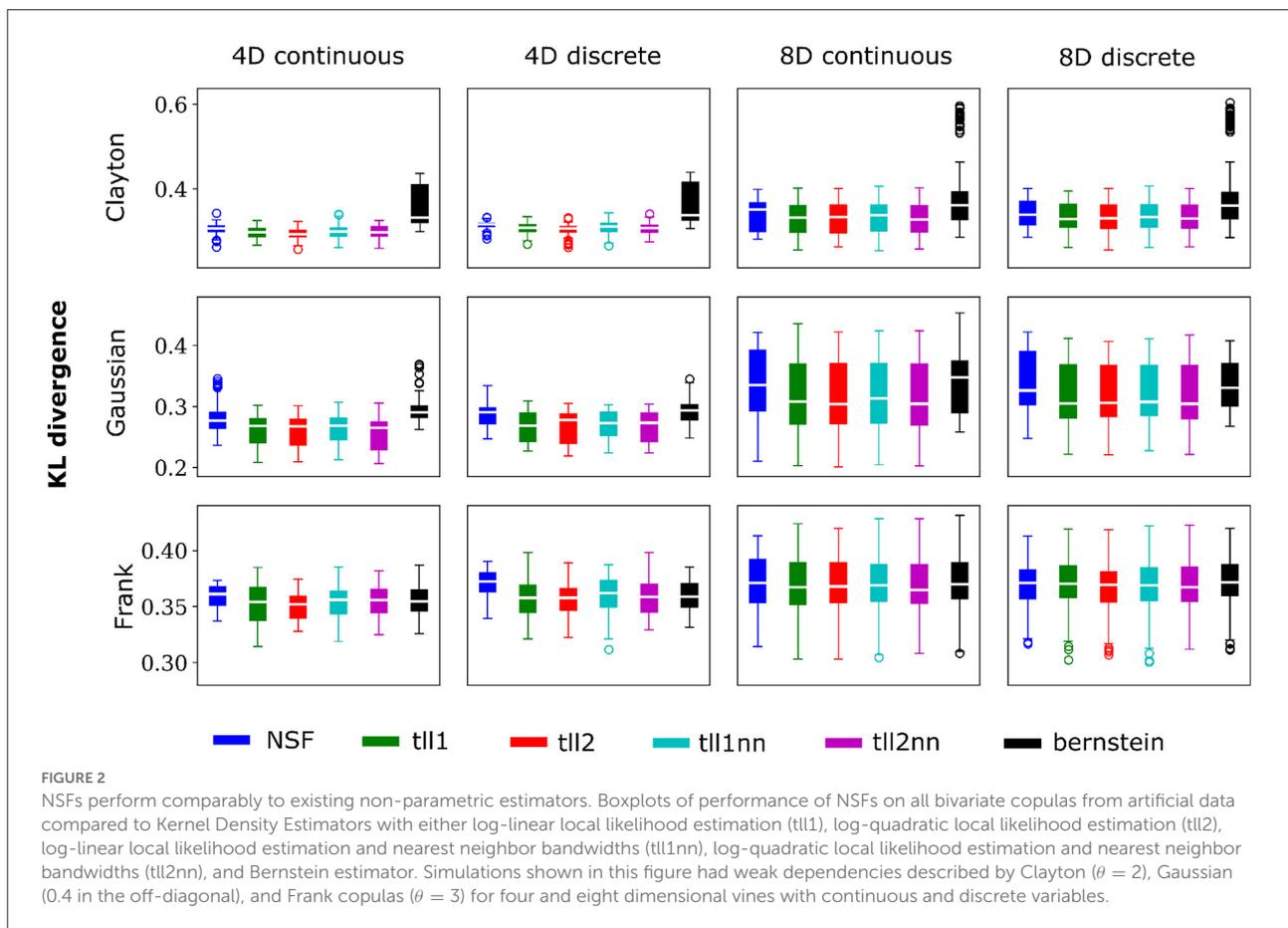
reward. This visual detection task was used to investigate V1 neuronal activity before, during and after learning (Henschke et al., 2020). The goal of the experiments was to elucidate how repeated exposure to a stimulus modulates neural population responses, particularly in the presence of a stimulus-associated reward. Our analysis was based on deconvolved spike trains instead of the calcium transients. Spiking activity had been reconstructed using the MLspike algorithm (Deneux et al., 2016) (see Henschke et al., 2020 for more details). The data we used in this study were limited to one mouse on day 4 of the experiment when the animal was an expert at the task. Moreover, in order to provide a proof-of-concept example and illustrate the complete vine copula decomposition as well as the performance of the NSF framework, we selected a subset of five neurons out of the 102 V1 neurons that were monitored in total for that particular mouse. Each of the five selected neurons showed non-negligible positive or negative rank correlation with every other neuron in that subset (Kendall's $\tau > 0.3$ or Kendall's $\tau < -0.3$), suggesting that they might be part of a module warranting a more detailed investigation of the dependence structures within.

3. Results

3.1. Validation on artificial data

In order to demonstrate the potential of the NSF vine copula framework to capture multivariate dependencies of arbitrary shape and data type (continuous vs. discrete) we conducted a simulation study. The set of generated samples that were used for training the NSF ($n = 5,000$) included cases with three different types of copulas (Clayton, Frank, and Gaussian) with each dictating a different set of dependence structures for weak and stronger dependencies, continuous and discrete data as well as four and eight dimensions. This ensemble provided a wide range of settings from simpler symmetric dependencies to skewed and heavy tailed interactions that can be encountered in neural population spiking responses.

Overall, performance of NSF as assessed by the KL divergence of the NSF-generated copula samples with those from the ground truth copulas was broadly within comparable levels to that of the KDE estimators while Bernstein estimators often performed the worst (Figures 2, 3). However, it is worth noting that relative performances varied slightly on a case-by-case level. For example, with weaker dependencies in four dimensional data with all copulas, NSF performed slightly but significantly worse than the other estimators (Kruskal-Wallis tests, $p < 0.05$) except Bernstein estimators (Figure 2). The latter even outperformed NSF in one exceptional case with weakly dependent 4D discrete data with Frank copulas. This trend of slightly worse NSF performance relative to all else except Bernstein estimators was also observed in 4D continuous and discrete data for all copulas with stronger dependencies

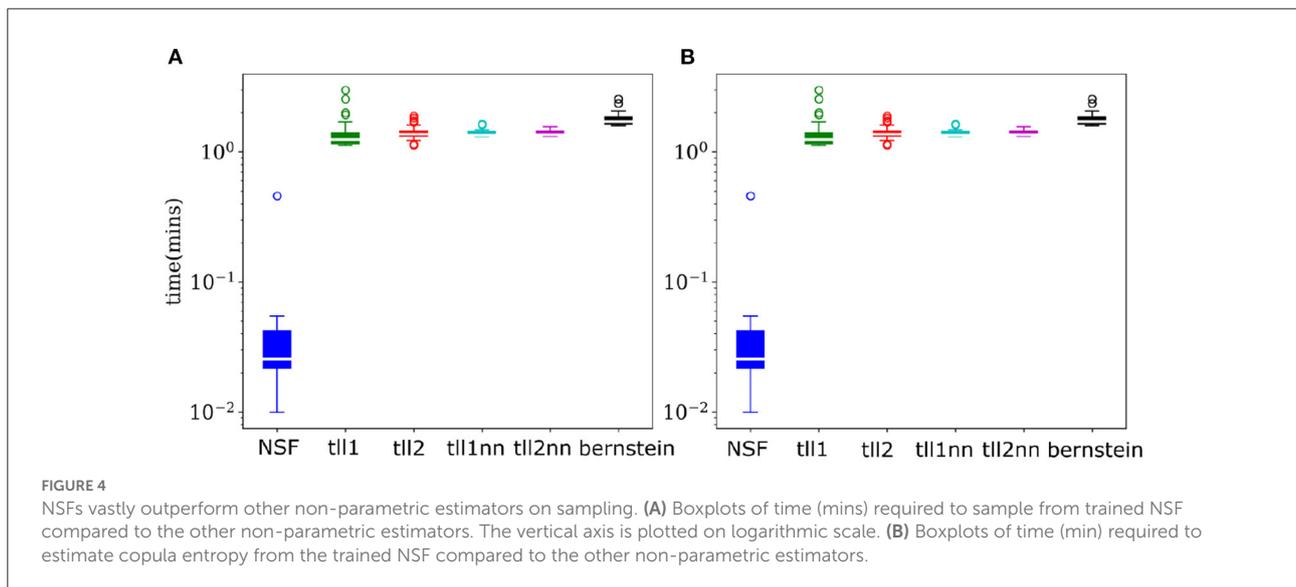
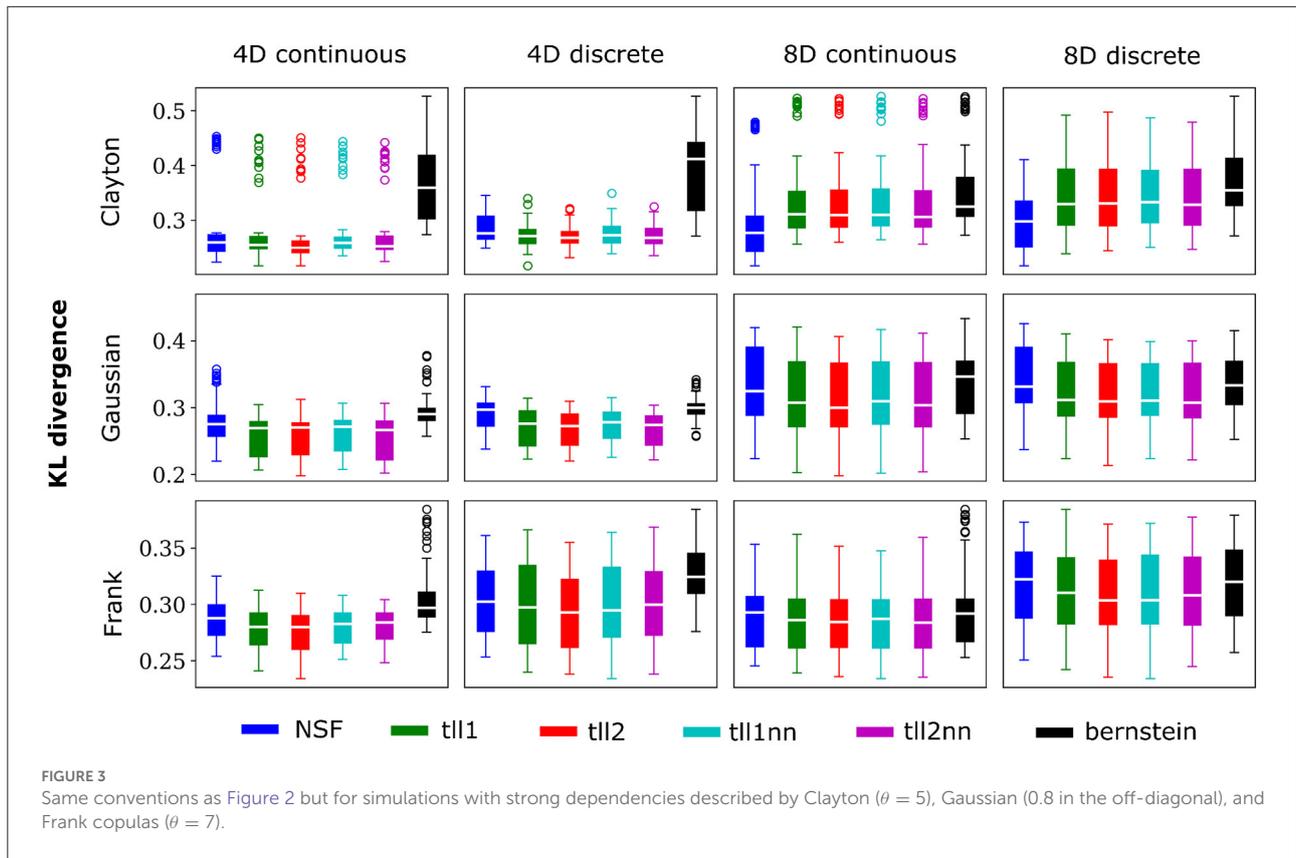


(Figure 3). However, NSF closed that gap and performed similarly with the group of KDE estimators in cases where data was 8D, either continuous or discrete and was entangled with Clayton copulas with weak dependencies or Frank copulas with both weak and stronger dependencies (Kruskal-Wallis tests, $p > 0.05$). Notably, in the case of Clayton copulas with stronger dependencies for 8D data, NSF outperformed all other estimators (Kruskal-Wallis tests, $p < 0.05$). These findings might suggest that the flexibility of the NSF framework can be more beneficial with higher data dimensionality and dependence structures that are characterized by heavier tails.

As copulas offer a detailed and invariant to marginal statistics view into multivariate dependencies, their negative entropy provides an accurate estimate of mutual information between variables. Thus, calculating copula entropies can be useful not only in understanding coordinated information processing but also in BCI settings where dependencies might be an important feature needing to be accounted for. Despite the largely similar performance to KDEs, NSFs showed a remarkable advantage regarding drawing samples from the trained models and estimating copula entropies. Inverting the kernel transformation in KDE estimators and rejection sampling in Bernstein estimators are considerably more computationally

expensive compared to a flexible and substantially faster sampling scheme in NSFs which directly approximate the CDF and inverse CDF of margins and copulas (Figure 4). Flexible and fast sampling is an attractive property useful for estimation of information theoretic quantities and compensates for the cases where NSF performs slightly worse compared to the rest.

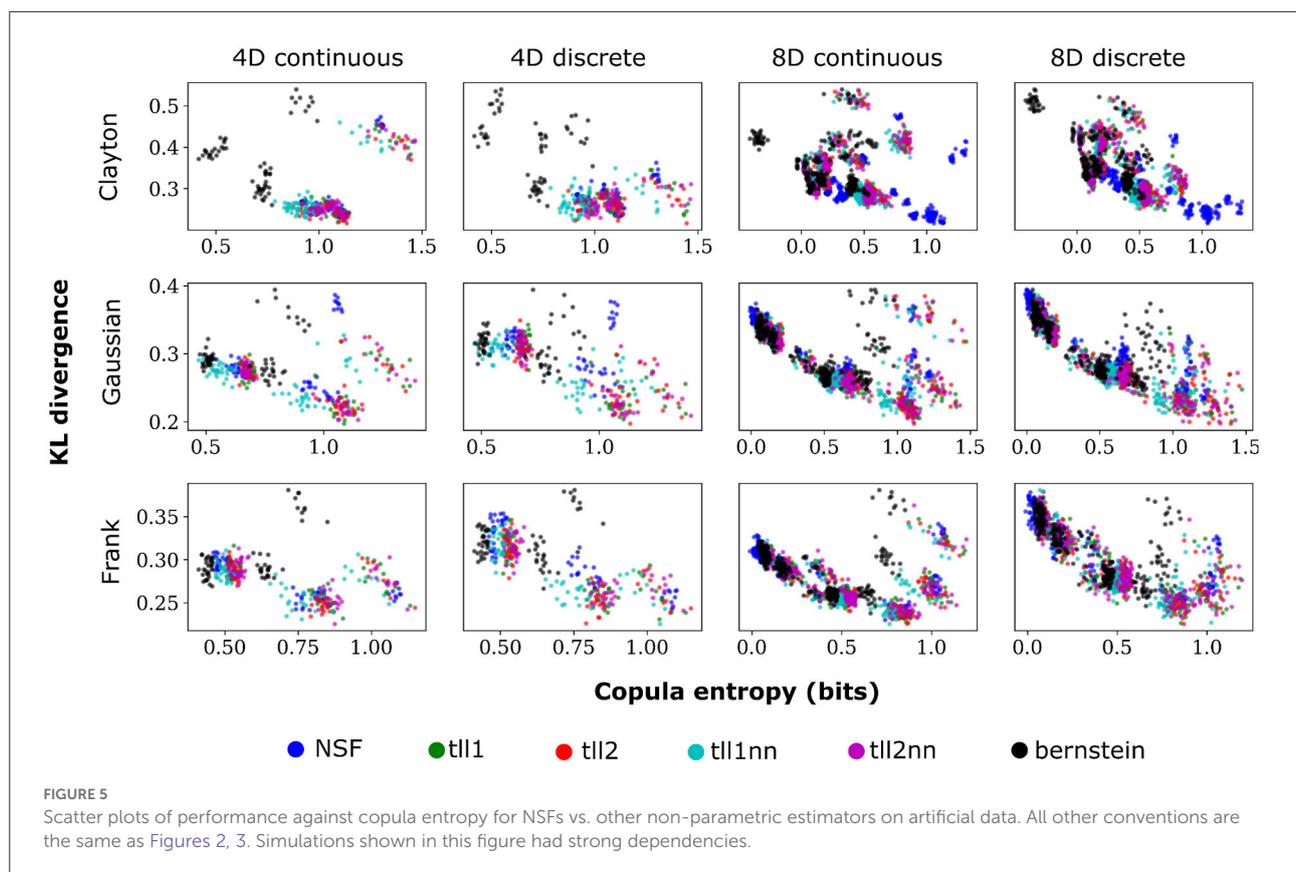
Plotting the copula entropies from all the estimators against the KL divergence for every particular iteration of fitting in every bivariate copula from the vine revealed an inverse relationship between the two quantities. Namely, better performance appeared to relate to higher copula entropy and thus mutual information for all distinct copulas, vine dimensions, and data types (Figure 5). This could mean that bivariate copulas with higher KL divergence were overfit to the point of diminishing the informational content of the interaction captured by the copula. It is noteworthy that Clayton copulas from 8D vines that were well fit by NSFs exhibited significantly higher copula entropy compared to the other estimators (Kruskal-Wallis, $p < 0.01$). This could indicate an potential advantage of NSFs over the other estimators with heavy-tailed data, which might warrant further future investigation.



3.2. NSF elucidate dependence structures in rodent V1 spiking activity and behavior

Having validated the vine flow copula framework with artificial data, we subsequently focused on spiking activity from a subset of 5 V1 layer 2/3 neurons while the animal was navigating

a virtual reality corridor (Figure 6A) (Henschke et al., 2020). A parametric vine copula framework with Gaussian processes (Kudryashova et al., 2022) was recently applied to calcium transients from the same V1 layer 2/3 neurons (Henschke et al., 2020). In the present study, we focused instead on modeling the deconvolved spiking activity with flow-based vine copulas.

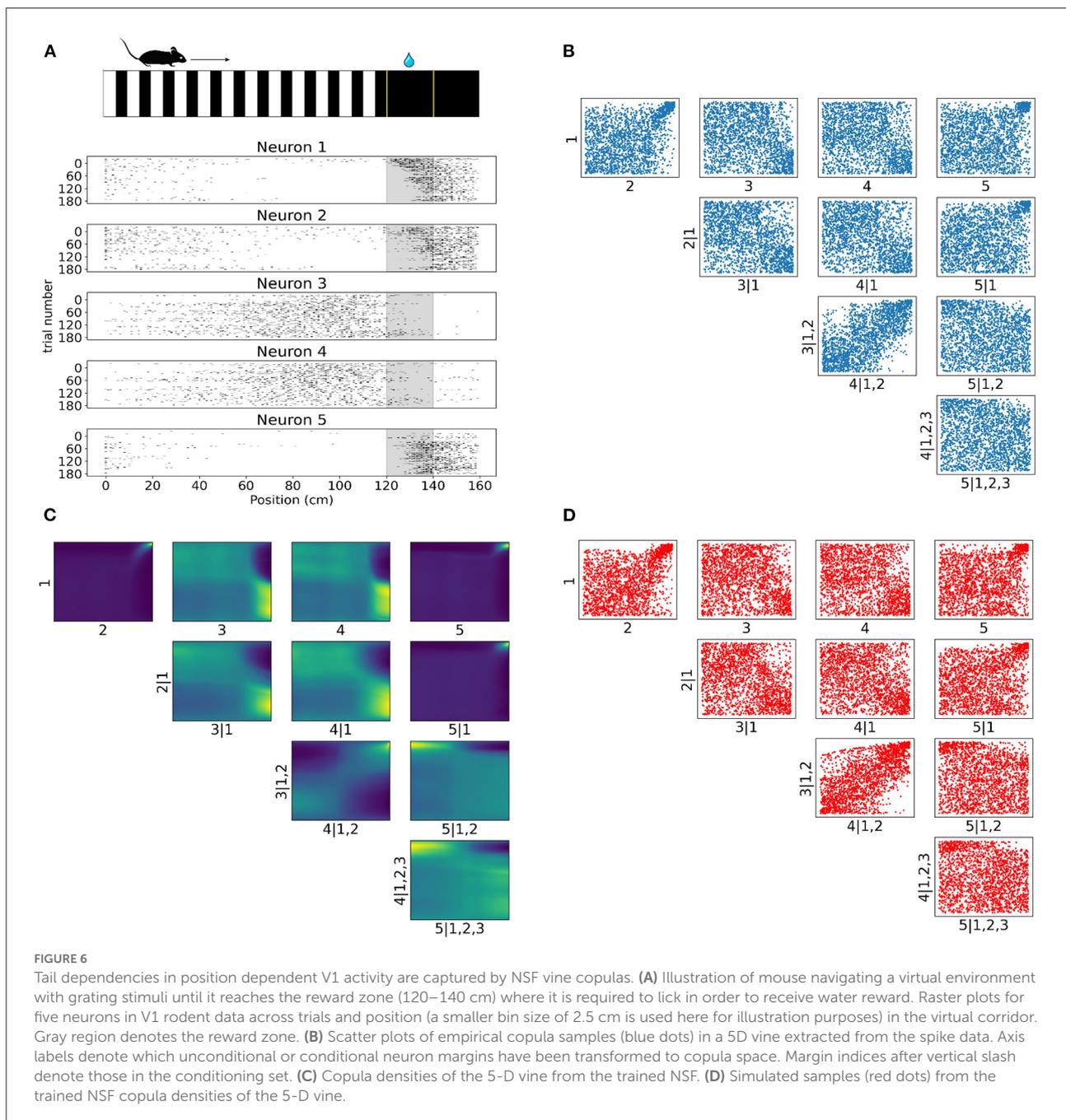


Our use of a non-parametric tool aimed at escaping potentially restricting assumptions that parametric copula families might place on the dependence structures among neurons. Our analysis aimed to detect such dependencies both as a function of time and position since previous findings indicate that V1 neuronal activity can be modulated by behaviorally relevant variables such as a reward being given at a particular location.

Raster plots across all trials on day 4 for the subset of the selected five neurons in Figure 6A illustrate how some of the neurons were more active for a short span of the virtual corridor within and after the reward zone (120–140 cm) while the activity of others was more spread out across the corridor and fell off at the onset of the reward zone. The strength of rank correlation between pairs of these neurons can be assessed by measuring the Kendall's τ from their spiking activities. However, this approach reduces the study of dependencies into single numbers that only provide a limited description of the interactions. In contrast, a copula-based approach can provide a detailed account of the actual shapes of neuronal dependencies which are usually characterized by a concentration of probability mass in the tail regions.

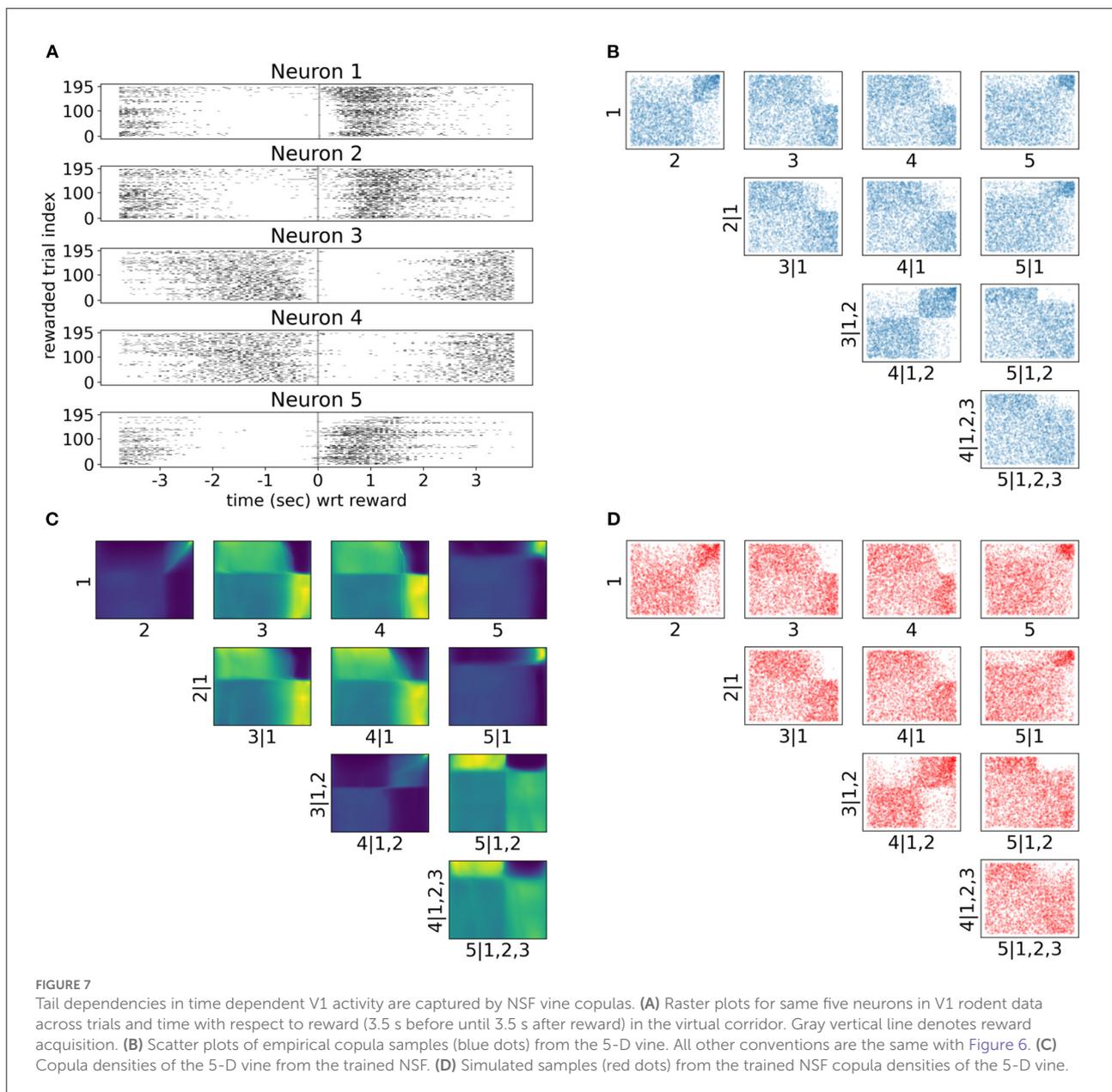
In similar fashion to the analysis on artificial data, we fit a C-vine whereby NSFs were fit to the spiking distributions of the neurons binned with respect to position (bin size was 20 cm), i.e., the margins as a first step and then NSFs were sequentially

fit to the empirical copulas (blue scatter plots in Figure 6B) from the surface tree level to the deeper one. These empirical copulas were obtained by transforming the data with the distributional transform. The five neurons were ordered according to the degree to which they exhibited statistical dependence with the other neurons as measured by the sum of Kendall's τ for each neuron. The copula densities (Figure 6C) and samples (red scatter plots in Figure 6D) from the trained NSFs were able to accurately capture the dependence structures between the neurons. All interactions were characterized by the presence of heavy tails in different corners of the uniform cube. Heavy tail dependencies at the top right part of the cube signified neurons that were more co-dependent when both neurons were active compared to other activity regimes (e.g., Figure 6C top row 1,2 and 1,5). For example, neurons like neuron 1 and neuron 2 displayed weak or no significant interaction until their activity was modulated by experimental or behavioral variables that are associated with the reward zone. Conversely, heavy tails at the bottom right (Figure 6C 2,3|1) or top left (Figure 6C 4,5|1,2,3) signified inverse relations of spiking activity between these neurons, i.e., being active at different locations in the corridor. Furthermore, it is worth noting that the probability mass concentration in the tail regions was different among neurons, with some pairs displaying lighter tails than others (e.g., Figure 6C 3,4|1,2).



As neuronal activity in V1 is modulated by introducing a reward in a certain location of the virtual corridor, it is also interesting to investigate neural responses and dependencies in time around that reward. Therefore, we also analyzed neural interactions as revealed by the spike counts of the same five neurons 3.5 s before and after the reward (bin size was 300 ms). Only successful trials where the mouse licked within the reward zone were included in this analysis. Raster plots in [Figure 7A](#) show a variety of spiking patterns relative to the timing of the reward across trials. The copula densities ([Figure 7C](#)) and

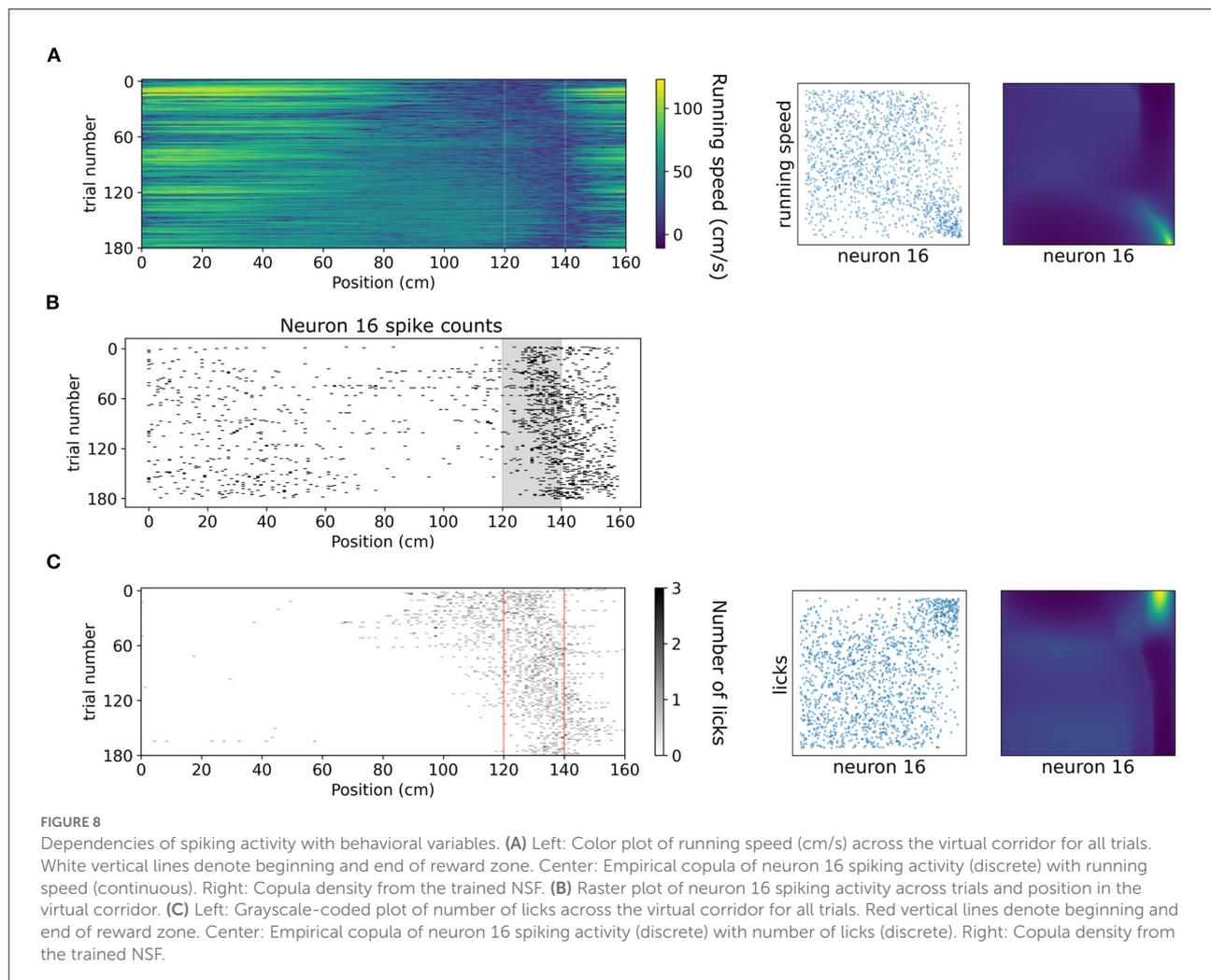
samples (red scatter plots in [Figure 7D](#)) from the trained NSFs closely captured the dependence structures as before ([Figure 7B](#)). Moreover, as before, these dependence structures were characterized by heavy tailed regions either in the top right (e.g., 1, 2 in [Figure 7C](#)) or top left corners (e.g., 2, 5|1 in [Figure 7C](#)) indicating stronger neuronal interactions for some regimes of spiking activity and not others. The more apparent block structure in this case compared to the previous one was a result of the fewer states of spike counts with the bin size we chose. For example, probability mass in the copula space



for neurons that would fire from 0 to 5 spikes in a given time bin, would have to be assigned to blocks that correspond to the relative proportions of jointly observed spike counts.

A copula-based approach can also offer a tool for studying dependencies of neuronal activity with behavioral variables which might exhibit vastly different statistics, such as running speed and number of licks. In Figure 8 we showcase how copulas modeled with NSFs can elucidate the shape of interaction of running speed and licks with the spiking activity of an example neuron (neuron 16 from the dataset, which is not included in the five neurons in the previous analysis) binned with respect to position (bin size was 20 cm) in the virtual corridor. This neuron increased its activity considerably within the reward

zone (Figure 8B) which coincided with greatly reduced running speed (Figure 8A) as the mouse stopped to lick (Figure 8C) and receive the reward. Licking was thus positively co-dependent with spiking activity, $\tau = 0.4, p < 0.001$ for number of licks and running speed was negatively co-dependent with spiking activity $\tau = -0.27, p < 0.001$. Transforming the variables to copula space revealed regions of mostly uniform probability mass concentration with a relatively heavier tail in the bottom right for running speed (copulas in Figure 8A) and in the top right for licking (copulas in Figure 8C). This finding indicated that these behavioral variables had a rather weak or no interaction for most of the span of the virtual corridor except for the part when the neuron was mostly active, which was the reward zone.



Considering all the aforementioned, the flow-based vine copula framework shows remarkable promise for flexibly capturing complicated interaction patterns within neural populations. The rich picture of these interactions and potential insights into neural circuit function would have remained undetected by only measuring pairwise rank correlations or pairwise Pearson correlations which would not have indicated any heavy tailed interactions.

4. Discussion

In this study, we proposed a fully non-parametric approach for estimating vine copula densities for both continuous and discrete data using NSF, a subtype of normalizing flows. Overall, the framework performed comparably to existing non-parametric approaches on artificial data while benefiting from more flexible and faster sampling. There were some cases where NSF performed slightly worse than the other estimators. We mostly observed this in vine models with dependencies

characterized by lighter tails (Frank, Gaussian) compared to those with heavy tail dependencies (Clayton). These findings, in combination with NSF being superior in the eight dimensional Clayton vine, might potentially point to performance differences between the estimators in different regimes. The added flexibility provide by NSF might come at the cost of fitting noise to some extent and simpler cases like the vines described by Frank and Gaussian copulas might be more easily captured by the other estimators. Conversely, complex dependencies induced by more interacting components and heavier tails might benefit from that flexibility. Furthermore, we demonstrated how flow-based vine copulas can shed light on the structure of dependencies in neural spiking responses.

The intricate shapes in bivariate copulas of spikes involving skewness and heavy tails would have been assumed as non-existent in a conventional approach based on pairwise linear correlations. Additionally, the arrangement of the discovered copulas into block structures (Figure 6B) would have been harder to capture by commonly applied parametric copula models (e.g., Clayton, Frank, or Gaussian copulas) and thus

lead to misleading conclusions. Therefore, we showed that non-parametric modeling with normalizing flows can be a valuable tool especially in the case of copula-based methods for discrete data and mixed data.

The development of such tools is crucial for understanding how coordinated neural activity transmits information about external stimuli or internal signals that guide planning and action. Copula-based methods have the capacity to provide a description of neural activity that includes the intricacies of dependencies and allows for higher-order interactions to occur within a certain set of conditions specified by the vine structure. Such descriptions can potentially offer significant insights that will aid and inform the development of neural coding theory. At the same time, decoding models that take into account the aforementioned features elucidated by copula methods, could potentially be valuable for the development of reliable Brain-Computer Interfaces. Some lines of evidence suggest that higher-order interaction patterns might have a significant presence and role in shaping collective neural dynamics and conveying information (Ohiorhenuan et al., 2010; Yu et al., 2011; Shimazaki et al., 2012) but further research needs to be conducted to gain a deeper understanding of the processes involved.

A limitation of our study is that the type of normalizing flows we used was designed to model variables with continuous data. We therefore included an additional rounding operation for discrete margins that were generated by trained NSF. This issue with discrete margins could be addressed in future work that can improve upon the framework by incorporating normalizing flow models that are more naturally suited to handle discrete or mixed data and do not need the *ad hoc* modifications that were employed in the present project. SurVAE flows that were developed recently (Nielsen et al., 2020) as an attempt to unify variational autoencoders and normalizing flows might potentially be better suited for discrete data as the kind of surjective transformations may better account for discontinuities in discrete CDF. Another limitation derives from the fact that our analysis was based on deconvolved spikes. Due to the slow time constant of the underlying calcium traces, spike deconvolution is unlikely to reconstruct any potential short time scale dependencies that might have existed within a neural population. Thus, the resulting copulas are very likely to look uniform if bin size is not sufficiently big. We were therefore confined to consider only larger bin sizes in our analyses.

Future directions include the analysis of entire neural populations as opposed to small subsets of a population. A possible extension to modeling population dependencies with vine copula methods can also involve leveraging the power of dimensionality reduction methods. A well-established trait of neural population activity is that latent collective dynamics that drive responses occupy a subspace that is much lower in dimensionality than the number of neurons recorded (Cunningham and Byron, 2014). Thus, dimensionality reduction methods could be applied prior to copula modeling

so that the latter can be employed on a set of low dimensional latent factors that capture the most prominent latent processes. Combining such methods can potentially provide even less computationally demanding frameworks that can be useful for clinical translation.

Data availability statement

Data is available upon reasonable request from the authors of the original experimental study. Requests to access these datasets should be directed to t.amvrosiadis@ed.ac.uk.

Ethics statement

The animal study was reviewed and approved by Animal Welfare and Ethical Review Board (AWERB) of the University of Edinburgh.

Author contributions

LM and AO conceptualized the study and designed the methodology. LM conducted data analysis, visualization of findings under the supervision of AO, and wrote the first draft version. TA collected and curated the neuronal recordings data. LM, TA, and AO reviewed and edited the article. All authors approved the submitted version.

Funding

This work was supported by the Engineering and Physical Sciences Research Council (grant [EP/S005692/1], to AO) and the Precision Medicine Doctoral Training Programme (Medical Research Council grant number [MR/N013166/1], to TA). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We would like to thank Nathalie Rochefort and Tom Flossman for constructive feedback on the analysis and manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insur. Math. Econ.* 44, 182–198. doi: 10.1016/j.insmatheco.2007.02.001
- Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366. doi: 10.1038/nrn1888
- Bedford, T., and Cooke, R. M. (2002). Vines—a new graphical model for dependent random variables. *Ann. Stat.* 30, 1031–1068. doi: 10.1214/aos/1031689016
- Bergstra, J., and Bengio, Y. (2012). Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* 13, 281–305.
- Bridges, N. R., Meyers, M., Garcia, J., Shewokis, P. A., and Moxon, K. A. (2018). A rodent brain-machine interface paradigm to study the impact of paraplegia on BMI performance. *J. Neurosci. Methods* 306, 103–114. doi: 10.1016/j.jneumeth.2018.05.015
- Brown, E. N., Kass, R. E., and Mitra, P. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.* 7, 456–461. doi: 10.1038/nn1228
- Chaudhary, U., Birbaumer, N., and Ramos-Murguialday, A. (2016). Brain-computer interfaces for communication and rehabilitation. *Nat. Rev. Neurol.* 12, 513–525. doi: 10.1038/nrneuro.2016.113
- Chen, X., Leischner, U., Varga, Z., Jia, H., Deca, D., Rochefort, N. L., et al. (2012). Lotos-based two-photon calcium imaging of dendritic spines *in vivo*. *Nat. Protoc.* 7, 1818–1829. doi: 10.1038/nprot.2012.106
- Cunningham, J. P., and Byron, M. Y. (2014). Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* 17, 1500–1509. doi: 10.1038/nn.3776
- Czado, C. (2019). *Analyzing dependent data with vine copulas. Lecture Notes in Statistics* (Cham: Springer). doi: 10.1007/978-3-030-13785-4
- de Leon, A. R., and Wu, B. (2011). Copula-based regression models for a bivariate mixed discrete and continuous outcome. *Stat. Med.* 30, 175–185. doi: 10.1002/sim.4087
- Deneux, T., Kaszas, A., Szalay, G., Katona, G., Lakner, T., Grinvald, A., et al. (2016). Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations *in vivo*. *Nat. Commun.* 7, 1–17. doi: 10.1038/ncomms12190
- Durkan, C., Bekasov, A., Murray, I., and Papamakarios, G. (2019). “Neural spline flows,” in *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)* (Vancouver, BC).
- Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., Logothetis, N. K., and Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science* 327, 584–587. doi: 10.1126/science.1179867
- Faugeras, O. P. (2017). Inference for copula modeling of discrete data: a cautionary tale and some facts. *Depend. Model.* 5, 121–132. doi: 10.1515/demo-2017-0008
- Gao, P., and Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* 32, 148–155. doi: 10.1016/j.conb.2015.04.003
- Genens, G., Charpentier, A., and Paindaveine, D. (2017). Probit transformation for nonparametric kernel estimation of the copula density. *Bernoulli* 23, 1848–1873. doi: 10.3150/15-BEJ798
- Genest, C., and Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bull. J. IAA* 37, 475–515. doi: 10.2143/AST.37.2.2024077
- Haff, I. H., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction—simply useful or too simplistic? *J. Multivar. Anal.* 101, 1296–1310. doi: 10.1016/j.jmva.2009.12.001
- Hartland, L. (2021). *KL Divergence Estimators*. Github. Available online at: <https://github.com/nhartland/KL-divergence-estimators.git>
- Henschke, J. U., Dylida, E., Katsanevaki, D., Dupuy, N., Currie, S. P., Amvrosiadis, T., et al. (2020). Reward association enhances stimulus-specific representations in primary visual cortex. *Curr. Biol.* 30, 1866–1880. doi: 10.1016/j.cub.2020.03.018
- Hurwitz, C., Kudryashova, N., Onken, A., and Hennig, M. H. (2021). Building population models for large-scale neural recordings: Opportunities and pitfalls. *Curr. Opin. Neurobiol.* 70, 64–73. doi: 10.1016/j.conb.2021.07.003
- Jaworski, P., Durante, F., and Härdle, W. K. (2012). “Copulae in mathematical and quantitative finance,” in *Proceedings of the Workshop Held in Cracow*, Vol. 10. doi: 10.1007/978-3-642-35407-6
- Jenison, R. L., and Reale, R. A. (2004). The shape of neural dependence. *Neural Comput.* 16, 665–672. doi: 10.1162/089976604322860659
- Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232–236. doi: 10.1038/nature24636
- Kamthe, S., Assefa, S., and Deisenroth, M. (2021). Copula flows for synthetic data generation. *arXiv [Preprint] arXiv:2101.00598*.
- Kohn, A., Coen-Cagli, R., Kanitscheider, I., and Pouget, A. (2016). Correlations and neuronal population information. *Annu. Rev. Neurosci.* 39, 237–256. doi: 10.1146/annurev-neuro-070815-013851
- Kudryashova, N., Amvrosiadis, T., Dupuy, N., Rochefort, N., and Onken, A. (2022). Parametric copula-gp model for analyzing multidimensional neuronal and behavioral relationships. *PLoS Computat. Biol.* 18:e1009799. doi: 10.1371/journal.pcbi.1009799
- McFarland, D., and Wolpaw, J. (2017). EEG-based brain-computer interfaces. *Curr. Opin. Biomed. Eng.* 4, 194–200. doi: 10.1016/j.cobme.2017.11.004
- Michel, M. M., and Jacobs, R. A. (2006). The costs of ignoring high-order correlations in populations of model neurons. *Neural Comput.* 18, 660–682. doi: 10.1162/neco.2006.18.3.660
- Montangie, L., and Montani, F. (2017). Higher-order correlations in common input shapes the output spiking activity of a neural population. *Phys. A Stat. Mech. Appl.* 471, 845–861. doi: 10.1016/j.physa.2016.12.002
- Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. *Nat. Neurosci.* 17, 1410–1417. doi: 10.1038/nn.3807
- Nagler, T., Schellhase, C., and Czado, C. (2017). Nonparametric estimation of simplified vine copula models: comparison of methods. *Depend. Model.* 5, 99–120. doi: 10.1515/demo-2017-0007
- Nelsen, R. B. (2007). *An Introduction to Copulas*. New York, NY: Springer Science & Business Media.
- Nielsen, D., Jaini, P., Hoogeboom, E., Winther, O., and Welling, M. (2020). Survae flows: surjections to bridge the gap between vaes and flows. *Adv. Neural Inform. Process. Syst.* 33, 12685–12696.
- Ohiorhenuan, I. E., Mechler, F., Purpura, K. P., Schmid, A. M., Hu, Q., and Victor, J. D. (2010). Sparse coding and high-order correlations in fine-scale cortical networks. *Nature* 466, 617–621. doi: 10.1038/nature09178
- Onken, A., Grünewälder, S., Munk, M. H., and Obermayer, K. (2009). Analyzing short-term noise dependencies of spike-counts in macaque prefrontal cortex using copulas and the flashlight transformation. *PLoS Computat. Biol.* 5:e1000577. doi: 10.1371/journal.pcbi.1000577
- Onken, A., and Panzeri, S. (2016). Mixed vine copulas as joint models of spike counts and local field potentials. *30th Conference on Neural Information Processing Systems (NIPS 2016)* (Barcelona).
- Panagiotelis, A., Czado, C., and Joe, H. (2012). Pair copula constructions for multivariate discrete data. *J. Amer. Stat. Assoc.* 107, 1063–1072. doi: 10.1080/01621459.2012.682850

- Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., and Lakshminarayanan, B. (2021). Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.* 22, 1–64.
- Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., and et al. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454, 995–999. doi: 10.1038/nature07140
- Racine, J. S. (2015). Mixed data kernel copulas. *Empiric. Econ.* 48, 37–59. doi: 10.1007/s00181-015-0913-3
- Rezende, D., and Mohamed, S. (2015). “Variational inference with normalizing flows,” in *Proceedings of the 32nd International Conference on Machine Learning* Vol. 37 (Lille), 1530–1538.
- Sancetta, A., and Satchell, S. (2004). The bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econ. Theory* 20, 535–562. doi: 10.1017/S026646660420305X
- Schallhorn, N., Kraus, D., Nagler, T., and Czado, C. (2017). D-vine quantile regression with discrete variables. *arXiv [Preprint] arXiv:1705.08310*.
- Shimazaki, H., Amari, S.-i., Brown, E. N., and Grün, S. (2012). State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data. *PLoS Comput. Biol.* 8:e1002385. doi: 10.1371/journal.pcbi.1002385
- Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 8, 229–231.
- Smith, M. S., and Khaled, M. A. (2012). Estimation of copula models with discrete margins via bayesian data augmentation. *J. Amer. Stat. Assoc.* 107, 290–303. doi: 10.1080/01621459.2011.644501
- Song, P. X.-K., Li, M., and Yuan, Y. (2009). Joint regression analysis of correlated data using gaussian copulas. *Biometrics* 65, 60–68. doi: 10.1111/j.1541-0420.2008.01058.x
- Wang, Q., Kulkarni, S. R., and Verdú, S. (2009). Divergence estimation for multidimensional densities via k -nearest-neighbor distances. *IEEE Trans. Inform. Theory* 55, 2392–2405. doi: 10.1109/TIT.2009.2016060
- Widge, A. S., and Moritz, C. T. (2014). Pre-frontal control of closed-loop limbic neurostimulation by rodents using a brain–computer interface. *J. Neural Eng.* 11:024001. doi: 10.1088/1741-2560/11/2/024001
- Wiese, M., Knobloch, R., and Korn, R. (2019). Copula & marginal flows: disentangling the marginal from its joint. *arXiv [Preprint] arXiv:1907.03361*.
- Yu, S., Yang, H., Nakahara, H., Santos, G. S., Nikolić, D., and Plenz, D. (2011). Higher-order interactions characterized in cortical activity. *J. Neurosci.* 31, 17514–17526. doi: 10.1523/JNEUROSCI.3127-11.2011
- Zohary, E., Shadlen, M. N., and Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370, 140–143. doi: 10.1038/370140a0