
Research Article: Methods/New Tools | Novel Tools and Methods

Assessing the impacts of correlated variability with dissociated timescales

Toshiyuki Takahashi¹, Yoshiko Maruyama², Hiroyuki Ito³ and Keiji Miura¹

¹*Department of bioscience, School of Science and Technology, Kwansai Gakuin University, Hyogo, Japan*

²*National Institute of Technology, Hakodate College, Hokkaido, Japan*

³*Faculty of Information Science and Engineering, Kyoto Sangyo University, Kyoto, Japan*

<https://doi.org/10.1523/ENEURO.0395-18.2019>

Received: 14 October 2018

Revised: 30 January 2019

Accepted: 5 February 2019

Published: 11 February 2019

T.T. and K.M. performed research; T.T. and K.M. analyzed data; Y.M., H.I., and K.M. designed research; K.M. wrote the paper.

Funding: <http://doi.org/10.13039/501100001691> Japan Society for the Promotion of Science (JSPS)
18K11485
18K13251
16K01966 and 15H05878

Conflict of Interest: No. Authors report no conflict of interest.

JSPS KAKENHI Grant Numbers 18K11485, 18K13251, 16K01966, and 15H05878.

Correspondence should be addressed to Keiji Miura, miura@kwansai.ac.jp

Cite as: eNeuro 2019; 10.1523/ENEURO.0395-18.2019

Alerts: Sign up at www.eneuro.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Accepted manuscripts are peer-reviewed but have not been through the copyediting, formatting, or proofreading process.

Copyright © 2019 Takahashi et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

1 **Manuscript Title Page**

2 1. Assessing the impacts of correlated variability with dissociated timescales

3 2. Abbreviated Title: Assessing correlations with dissociated timescales

4 3. Toshiyuki Takahashi[§], Yoshiko Maruyama[†], Hiroyuki Ito[‡], Keiji Miura[§]

5 [§]Department of Bioscience, School of Science and Technology, Kwansai Gakuin
6 University, Hyogo, Japan

7 [†]National Institute of Technology, Hakodate College, Hokkaido, Japan

8 [‡]Faculty of Information Science and Engineering, Kyoto Sangyo University, Kyoto, Japan

9 4. YM, HI and KM Designed Research; TT and KM Performed Research and Analyzed
10 data; KM Wrote the paper.

11 5. Correspondence should be addressed to Keiji Miura (miura@kwansai.ac.jp)

12 6. Number of Figures: 6

13 7. Number of Tables: 0

14 8. Number of Multimedia: 0

15 9. Number of words for Abstract: 250

16 10. Number of words for Significance Statement: 111

17 11. Number of words for Introduction: 683

18 12. Number of words for Discussion: 1264

19 13. Acknowledgements: The authors declare no competing financial interests.

20 14. Conflict of Interest: No. Authors report no conflict of interest.

21 15. Funding sources: JSPS KAKENHI Grant Numbers 18K11485, 18K13251, 16K01966,
22 and 15H05878.

23

65 (Zohary et al. (1994), but see also Abbott & Dayan (1999), Sompolinsky et al. (2001), Miura (2012),
66 Moreno-Bote et al. (2014)). Therefore, it is extremely important to estimate noise correlations
67 accurately in the era of large-scale electrophysiology (Steinmetz et al. (2018)).

68 Although significant noise correlations have been observed in almost all recorded cortical areas, it
69 has been pointed out that nonstationarity such as drifts in signals can engender artificial correlations
70 even if no actual correlation exists (Bair et al. (2001), Ecker et al. (2010), Renart et al. (2010)).
71 Therefore, it is desired to dissociate the observed noise correlations into short- and long-term
72 components, where the latter is possibly caused by the background trends or fluctuations of the
73 baseline activity (Okun et al. (2015), Luczak et al. (2015), Fiser et al. (2004), Ikegaya et al. (2004),
74 Sasaki et al. (2007)). Although attempts to separate them and estimate purely short-term noise
75 correlations under changing environments have been made previously, they were ad hoc and applicable
76 only to specific cases (Bair et al. (2001), Ecker et al. (2010), Renart et al. (2010), Mitchell et al. (2009)).
77 Even the latest Bayesian method requires considerable numbers of simultaneously recorded neurons as
78 well as exponential computational costs to estimate instantaneous activities (Ecker et al. (2014),
79 Rosenbaum et al. (2017)). Thus, the estimation method, that requires only the recording of a pair of
80 neurons and works for arbitrary baseline drifts nonparametrically (Amari & Cardoso (1997)), is desired.

81 In addition to measuring the noise correlations, assessing their impacts is also very important. The
82 degree to which sensory information is represented reliably by neural responses has been
83 characterized by applying a decoding approach in a stochastic stimulus-response framework (Dayan &
84 Abbott (2001), Averbeck et al. (2006), Sharpee (2017)). That is, the decoding success rates can be used
85 as a measure of accuracy of neural representations. You can take different features of neural activities
86 as clues for decoding to see which feature carries information. Therefore, it is ideal, within this
87 framework, if the dissociation of short- and long-term correlations gives us a novel way to assess their
88 respective impacts on information representations.

89 In this paper, we propose an information-geometric method to unbiasedly estimate pure short-term
90 noise correlations irrespective of the background brain activities. One effective way to use the
91 information geometry, that generally finds orthogonal statistical parameters (Amari & Nagaoka (2001),
92 Miura (2011), Amari (2016)), is to estimate only finite parameters of interest irrespective of the other
93 infinite numbers of parameters (Miura et al. (2006*b,a*, 2007), Miura & Uchida (2008)). Here we utilized
94 this infinite-dimensional scheme (Amari & Kawanabe (1997), Miura (2013)) to dissociate the
95 parameter for short-term correlation from the infinitely many parameters for (all possible) long-term
96 baseline drifts nonparametrically. This allows us to estimate pure short-term correlations whatever the
97 baseline drift is without demanding considerable numbers of simultaneously recorded neurons and
98 high computational costs. Then, the accompanying statistical test as well as the existing nonstationarity
99 test enabled us to dissociate short- and long-term correlations. First, as benchmark simulations, we
100 demonstrated that the proposed estimator is more accurate and computationally efficient than the
101 conventional correlograms and the residual correlations with Kalman filters or moving averages of
102 length three or more, while the best moving average of length two coincided with the propose method
103 regarding correlation estimates. Next, when we excluded the spurious noise correlations of purely long-
104 term nature, only small fraction of V1 neuron pairs showed significant short-term noise correlations,
105 possibly reconciling the previous inconsistent observations on the existence of significant noise
106 correlations. Finally, with the additional help of the machine learning that classifies stimuli from neural
107 activities, we assessed the impacts on decoding of the presence of short- or long-term noise
108 correlations, separately. The presence of pure short-term correlations slightly improved the decoding
109 accuracy, while the pure long-term correlations deteriorated the generalization ability. However, the
110 decrease in decoding accuracy by the long-term correlations was recoverable by using the decoder with
111 offset, suggesting that the brain could overcome nonstationarity by detrending. Thus, our method
112 enables us to elucidate the impacts of short- and long-term correlations in a dissociated manner,
113 advancing a modern, component-wise information theoretic analysis (Schneidman et al. (2003), Latham
114 & Nirenberg (2005), Averbeck et al. (2006), Sharpee et al. (2006)).

115 2 Materials and Methods

116 All the simulations and data analyses in this paper were done by using R. Throughout the analyses in
 117 the paper, the firing rate for each trial was used as an activity feature. The firing rate was computed as
 118 the spike count divided by the trial duration with a visual stimulus, which varied by trials from 1.0 to
 119 1.7 seconds. Thus when we say correlation coefficients or (trial-shifted) correlograms, we solely consider
 120 spike count noise correlations.

121 2.1 Proposed estimator for short-term noise correlation

122 As a measure of short-term noise correlations, we proposed and used the following estimator,

$$123 \quad \hat{\rho}_{12} = \frac{\hat{\rho}_{12}}{\sqrt{\hat{\rho}_{11}\hat{\rho}_{22}}} \quad \hat{\rho} = (1)$$

124
125

126 where the covariances $\hat{\rho}_{ij}$ are estimated as

$$127 \quad \hat{\rho}_{11} = \frac{2}{N} \sum_{t=1}^{N/2} \{(x_{2t-1} - \bar{x}_{(2t)})^2 + (x_{2t} - \bar{x}_{(2t)})^2\}$$

$$128 \quad \hat{\rho}_{22} = \frac{2}{N} \sum_{t=1}^{N/2} \{(y_{2t-1} - \bar{y}_{(2t)})^2 + (y_{2t} - \bar{y}_{(2t)})^2\}$$

$$129 \quad \hat{\rho}_{12} = \frac{2}{N} \sum_{t=1}^{N/2} \{(x_{2t-1} - \bar{x}_{(2t)})(y_{2t-1} - \bar{y}_{(2t)}) + (x_{2t} - \bar{x}_{(2t)})(y_{2t} - \bar{y}_{(2t)})\} \quad (2)$$

130 Therein, x_t and y_t denote the neural responses in spike counts within a few seconds in the t -th trial,
 131 while the local mean activities were defined by

$$132 \quad \bar{x}_{(2t)} = \frac{x_{2t-1} + x_{2t}}{2} \quad \text{and}$$

$$133 \quad \bar{y}_{(2t)} = \frac{y_{2t-1} + y_{2t}}{2} \quad (3)$$

134 The proposed measure in Eq. 1 is comparable to the conventional correlation coefficient. When we
 135 plotted in the form of correlograms, we first shifted one of two time series by τ trials and then
 136 computed the proposed measure for them.

137 2.1.1 Code Accessibility

138 The R code for computing the proposed correlation coefficient and its p-value as defined in Sec. 2.5 is
 139 freely available online at [URL redacted for double-blind review]. The code is available as Extended
 140 Data. The code is ready to run just by replacing the example data for Figure 4 with users' own data.

141 As there can be a minor style difference in coding the proposed measure, we unified the rule and
 142 adopted the one with the minimum errors throughout the paper and the downloadable code. That is,
 143 there are two possible ways for pairing two neighboring trials, (A) starting at the first trial as {3,4}, {5,6},
 144 ..., and (B) starting at the second trial as {2,3}, {4,5}, {6, 7}, In the adopted {1,2},

145 {style, we took the average of the two estimated covariances because we found it had smaller variances
 146 (estimation errors). This style difference only negligibly modifies the results and the overall conclusions
 147 never change.

148 2.1.2 Assumption and derivation of proposed estimator

149 The proposed estimator in Eq. 1 was derived for estimating parameters in a semiparametric statistical
 150 model. That is, the activities of two neurons were hypothesized to obey the following statistical model
 151 (Eq. 4) and the proposed estimator estimates the Gaussian covariance there in. Although the derivation
 152 and concise benchmark simulations were already shown elsewhere (Miura (2013)), the application to
 153 the real experimental data has not been done yet.

154 In this paper, we solely consider the spike count within a trial, where the spiking activity of a neuron
 155 is integrated over a couple of seconds and, thus, well approximated by a Gaussian distribution. This
 156 leads us to consider a bivariate normal distribution for activities of two neurons, $q(x, y; \mu_x, \mu_y, \hat{\Sigma})$, where μ_x
 157 and μ_y denote the means for two neurons' activities and $\hat{\Sigma}$ denotes the covariance matrix. The activities x
 158 and y denote the spike counts of two neurons for a trial. These analyses address the situation in which
 159 the covariance matrix $\hat{\Sigma}$ is constant whereas the signals μ can change over time. Especially, when the
 160 signals μ are distributed randomly, but two consecutive signals are the same from continuity condition,
 161 the distribution of activities at time $2t-1$ and $2t$ ($t = 1, 2, \dots$) can be described as a mixed model,

$$162 \quad p(\{x_{2t-1}, y_{2t-1}, x_{2t}, y_{2t}\}; \hat{\Sigma}, k(\mu_x, \mu_y)) = \int k(\mu_x, \mu_y) q(x_{2t-1}, y_{2t-1}; \mu_x, \mu_y, \hat{\Sigma}) q(x_{2t}, y_{2t}; \mu_x, \mu_y, \hat{\Sigma}) d\mu_x d\mu_y$$

163 where $k(\mu_x, \mu_y)$ denotes an unknown distribution of the signals. The only assumption made here is that
 164 the consecutive signals have equal value, at least approximately (see the practical discussion at the end
 165 of the following subsection). That assumption is minimal and realistic as it is satisfied, e.g., when the
 166 signal drift is continuous, and preferably, sufficiently low. From another viewpoint, this definition of
 167 noises as the activities which is not locally flat over time is quite convenient for estimation.

168 Furthermore, Eq. 4 is a semiparametric model (Bickel et al. (1993), van der Vaart (1998)) because it
 169 has both a vector $\hat{\Sigma}$ and a function $k(\mu_x, \mu_y)$ as parameters. It is generally not easy to estimate parameters
 170 in semiparametric models because a function space is fundamentally infinite dimensional (Neyman &
 171 Scott (1948)). However, it is known that, for some cases, only parameters of interest can be estimated e-
 172 ciently through differential geometric methods on the manifolds of a family of probability distributions
 173 (Amari & Nagaoka (2001), Amari & Kawanabe (1997), Miura et al. (2006b,a, 2007), Miura & Uchida
 174 (2008)). For this model, it is possible to estimate the three constant parameters
 175 $\hat{\Sigma} = \{\hat{\Sigma}_{11}, \hat{\Sigma}_{12} (= \hat{\Sigma}_{21}), \hat{\Sigma}_{22}\}$ whatever the signal drift $k(\mu_x, \mu_y)$ is.

176 After a lengthy calculation in Miura (2013), the estimator was obtained as in Eq. 2. As the proposed
 177 estimator looks so simple, you might think that you can easily construct an arbitrary local smoother
 178 similar to the proposed estimator. However, because any arbitrarily invented estimators have larger
 179 estimation errors (biases and variances) in general, it is actually very difficult to discover an optimal
 180 estimator from scratch. As far as we know, other than information geometry, there is no systematic way
 181 to analytically derive an optimal estimator that works under arbitrary trends nonparametrically.
 182 Fortunately, it is very easy to just prove the optimality of the derived estimator, once it was derived.
 183 Therefore we take advantage of this fact for the educational purpose in what follows. That is, we do not
 184 repeat the derivation but rather only check the answer and demonstrate the performance of the
 185 proposed estimator concisely in the following subsection.

187 **2.1.3 Optimality of proposed estimator from statistical viewpoint**

188 Although this section is not necessary to read the main text, here we summarize and prove the
 189 optimality property from the statistical viewpoint. Specifically, we show that the proposed estimator
 190 has no bias (i.e. correct on average) and minimum variances (i.e. smallest errors) among the estimators
 191 which work unbiasedly for arbitrary baseline drifts.

192 The unbiased nature of the proposed estimators is clear from the fact that the estimators in Eq. 2 are
 193 normalized by dividing not by $2 (= M)$ but by $1 (= M - 1)$. Normalization of this type is widely known to
 194 guarantee the unbiased estimation for the covariances of Gaussian distributions. In fact, with $X_t = (x_t, y_t)$
 195 and $\mu = (\mu_x, \mu_y)$, the expectation of $\hat{\Lambda}_{12}$ can be calculated as an integral over the probability distribution
 196 in Eq. 4 as

$$197 \quad E[\hat{\Lambda}_{12}] := \int \int (Z \hat{\Lambda}_{12}(X_{2t-1}, X_{2t}) q(X_{2t-1} | \hat{\Lambda}, \mu) q(X_{2t} | \hat{\Lambda}, \mu) dX_{2t-1} dX_{2t}) k(\mu) d\mu = \int \hat{\Lambda}_{12} k(\mu) d\mu = \hat{\Lambda}_{12}. \quad (5)$$

199 This means that the estimator is unbiased or the estimator works (at least) "on average". The variance
 200 of the estimate can be similarly computed as

$$201 \quad \text{Var}[\hat{\Lambda}_{12}] := \int \int (Z \hat{\Lambda}_{12}(X_{2t-1}, X_{2t}) \hat{\Lambda})^2 q(X_{2t-1} | \hat{\Lambda}, \mu) q(X_{2t} | \hat{\Lambda}, \mu) dX_{2t-1} dX_{2t}) k(\mu) d\mu = \hat{\Lambda}_{11} \hat{\Lambda}_{22} + \hat{\Lambda}_{12} \hat{\Lambda}_{21}. \quad (6)$$

203 Surprisingly, $\hat{\Lambda}$ has the minimum variance (= estimation error) among all the estimators. To prove
 204 this, assume that $\check{\Lambda}(X_{2t-1}, X_{2t})$ is an arbitrary estimator of $\hat{\Lambda}_{12}$, that is,

$$205 \quad E(\check{\Lambda}(X_{2t-1}, X_{2t})) := \int \check{\Lambda}(X_{2t-1}, X_{2t}) p(X_{2t-1}, X_{2t}; \hat{\Lambda}, k(\mu_x, \mu_y)) dX_{2t-1} dX_{2t} = \hat{\Lambda}_{12}.$$

206 Note that we assumed that the expectation is equal to the statistical parameter of interest because any
 207 estimator should work at least "on average". By using the Cauchy-Schwartz inequality in functional

$$208 \quad \text{space } (f|g) = \int f \cdot g \text{ with } f = \check{\Lambda}(X) \hat{\Lambda}_{12} \text{ and } g = \hat{\Lambda}_{12}, \text{ We get} \\ 209 \quad \frac{\int \check{\Lambda}(X) \hat{\Lambda}_{12} p(X) dX}{\text{Var}[\check{\Lambda}]} \frac{\int \hat{\Lambda}_{12} p(X) dX}{\text{Var}[\hat{\Lambda}_{12}]} = \frac{(\int \check{\Lambda}(X) \hat{\Lambda}_{12} p(X) dX)^2}{\text{Var}[\check{\Lambda}] \text{Var}[\hat{\Lambda}_{12}]} = \frac{(\int \hat{\Lambda}_{12} p(X) dX)^2}{\text{Var}[\hat{\Lambda}_{12}]} \quad (7)$$

$$210 \quad \int \check{\Lambda}(X) \hat{\Lambda}_{12} p(X) dX = \int \hat{\Lambda}_{12} p(X) dX = \hat{\Lambda}_{12} \\ 211 \quad \int \hat{\Lambda}_{12} p(X) dX = \hat{\Lambda}_{12} \quad (8)$$

212 where $X = (X_{2t-1}, X_{2t})$. This shows that any estimator $\check{\Lambda}$ of $\hat{\Lambda}_{12}$ at least have the minimum variance in the
 214 right hand side:

$$215 \quad \text{Var}[\check{\Lambda}] \geq \text{Var}[\hat{\Lambda}_{12}] = \hat{\Lambda}_{11} \hat{\Lambda}_{22} + \hat{\Lambda}_{12} \hat{\Lambda}_{21}. \quad (9)$$

216 Actually the proposed estimator $\hat{\mu}_{12}$ attains this minimum variance as shown in Eq. 6. Thus, for any
 217 estimator $\hat{\mu}$,

$$218 \quad \text{Var}[\hat{\mu}] \geq \text{Var}[\hat{\mu}_{12}]. \quad (10)$$

219 Similar relations hold for $\hat{\mu}_{11}$ and $\hat{\mu}_{22}$.

220 We have demonstrated that the proposed estimator is optimal as far as the assumption on the
 221 statistical model hold. Practically, due to the violation of the assumption that the consecutive two
 222 signals (means) are exactly the same, the biases can arise. However, it can be shown from a simple
 223 calculus that the biases are generally small. In fact, if the consecutive signals are

$$224 \quad \begin{aligned} E[x_{2t+1}] &= \mu + \epsilon \text{ and} \\ E[x_{2t}] &= \mu + \epsilon \end{aligned} \quad (11)$$

226 differing of order of ϵ then, the biases are of the second order of ϵ^2

$$227 \quad E[\hat{\mu}_{11}] = E\left[\frac{x_{2t+1} - x_{2t}}{2}\right] = \frac{\epsilon - \epsilon}{2} = 0 \quad (12)$$

230 Thus even if you assume that the biases accumulate over the time points whose size is of order $\frac{1}{\epsilon}$, the
 231 total bias is still negligible, being of order $\epsilon^2 * \frac{1}{\epsilon} = \epsilon$. This suggests that even if the signal drifts slowly
 232 $O(\epsilon)$ as in Eq. 11, keeping the difference between the first and the last activities finite $O(1)$ after a long
 233 time sequence $O(\frac{1}{\epsilon})$, the total bias is negligibly small $O(\epsilon)$. In fact, Figure 5 (d) demonstrates that the
 234 proposed statistical test detects no spurious short-term correlations even if signals drift in the real V1
 235 data.

236 2.2 Simulation of activities of two neurons with drifting baselines

237 The simulations of bivariate Gaussian noises added to the baselines generated by the ARIMA models for
 238 activities of two neurons in Figure 1(a) were performed with `mvrnorm()` and `arma.sim()` functions in
 239 R.

240 2.3 Conventional cross correlograms

241 The conventional cross correlograms were computed with `cor()` function in R for the manually
 242 trialshifted data. As the function returns NA (i.e. Not Available) when either of two neurons show no
 243 spike across all 40 trials, we excluded those pairs from the analyses in the paper. Note that the
 244 proposed method also returns NA for those pairs. We computed time-shifted noise correlations or cross
 245 correlation functions separately for different stimuli, because recent works indicated the stimulus
 246 dependency of noise correlations (Kohn & Smith (2005), Maruyama & Ito (2013), Ruoff & Cohen
 247 (2016)).

248 2.4 Kalman filter method

249 The smoothing by the Kalman filter to obtain the baseline trend of the simulated neural activities was
 250 computed with `dlmFilter()` function in `dlm` package for R (Petris et al. (2009)). The noise correlations
 251 in residuals was obtained by the maximum likelihood method for data fitting with `dlmMLE()` function
 252 in the same package. The statistical model for the baseline trend μ_t we assumed to decode with Kalman
 253 filter was,

$$254 \quad \mu_{t+1} = F \mu_t + \eta \quad (\text{Gaussian noise})$$

255
$$X_t^{(i)} = G^{(i)}\mu_t + \xi_i(\text{Gaussian noise}) \quad (13)$$

256 where $X^{(i)}$ denotes i -th neuron's activity and F and G are to be estimated by data fitting.

257 The computational time was measured by `proc.time()` function in R on iMac with 3.3 GHz Intel Core
258 i5 and 32 GB memory.

259 2.5 Statistical tests for short-term noise correlations

260 We detected neuron pairs that have significant short-term noise correlations by using the statistical test
261 accompanying our estimator. As is usual with statistical tests, we computed p-values under the null
262 hypothesis of no correlation.

263 One possible way, which we did not adopt, was to assume the asymptotic normality for the
264 distribution of the proposed estimator, whose mean and variance can be computed from Eqs. 5 and 6
265 (or from simulations). However, for the current case, each neuron has only finite 40 trials per stimulus
266 and, thus, the normality assumption holds only approximately. Therefore, for example, the control p-
267 value distribution for the 1-time-shifted data is not as flat as in Figure 5 (d), although it is approximately
268 flat. Although this method saves the computational time, it seems to lack the accuracy in p-values.

269 To pursue the full accuracy, we resorted to the computational method with the white Gaussian
270 Monte Carlo simulations for reference activities of neuron pairs. Here the test was based on the idea
271 that even if there is no short-term correlation, its estimate from finite 40 trials takes a nonzero value
272 (error), which varies according to some statistical distribution. First, we obtained the shape of the
273 distribution as accurate as possible by repeating the Monte Carlo simulations a million times. Next, the
274 p-value for a given estimate is defined as its percentile in this numerically obtained distribution. That is,
275 the p-value is defined so that the p-value distribution is completely flat for white Gaussian noises. To be
276 precise, the p-value varies by realization of the activities of two neurons, but, with many realizations,
277 you obtain the uniform distribution for the p-values. Note that the uniform p-value distribution is a
278 hallmark of a good statistical test. Finally, if an estimate is too high or too low within the numerically-
279 obtained distribution (typically top 2.5% for both sides, corresponding to positive and negative
280 correlations), it is detected as significant or violating the null hypothesis.

281 When we computed the control p-value distribution for "1-time-shifted" data in Figure 5 (d), we
282 actually shifted 2-trials. This is because our proposed estimator treat the time series by pairs of time
283 points as in Eqs. 2 and 4. This is also why we shifted 2, 4, 6, ..., trials in Figure 1 and 2.

284 The R codes for computing the proposed short-term correlation and the accompanying statistical
285 test was hand-written.

286 As the level of significance, $\frac{0.01}{16 \frac{N(N-1)}{2}}$, where N denotes the number of neurons in the session and
287 16 is for 16 stimuli, was entirely used in the paper (in Figure 6, specifically). That is, we employed
288 Bonferroni's multiple comparison technique, because we wanted to keep the number of neuron pairs
289 moderately. Note that if we remove a neuron, we lose many pairs in the same session.

290 2.6 Statistical tests for nonstationarity

291 We selected neurons with and without nonstationarity by using the serial correlation test for
292 randomness of fluctuations (CASE64 in Kanji (2006)). To remove the effect of stimulus presentation
293 from the time series of neural activities, we averaged local 16 trials within a single block where 16
294 different stimuli are presented pseudo-randomly. In this way, the length of the time series was reduced
295 from original 640 to 40 trials, to which we applied the test. The R code for the test was hand-written.
296 The validity of the test was confirmed by the observation that the test returns uniformly distributed p-
297 values for the Gaussian white noises or the completely random time series in which a random number
298 is generated according to the normal distribution at every time. Note that the resulting p-value varies
299 by (random) time series and, here, we confirmed that the distribution got flat with many realizations.

300 As the level of significance, 0.01 was entirely used in the paper (in Figure 5 and Figure 6, specifically).
301 We did not employ the multiple comparison techniques, as we wanted to categorize suspicious neurons
302 into the nonstationary neuron pool, conservatively.

303 **2.7 Classification analysis and principal component analysis**

304 For the classifications of 16 visual stimuli based on the firing rates of neurons, we solely used `lda()`
305 function in R in this paper, although the result did not change significantly when we used the support
306 vector machine. The classification was done session by session in order to utilize the simultaneity of the
307 recorded data. For the statistical significance, the means of classification success rates for all sessions
308 were compared between different conditions by the paired t-test. Only the sessions with more than five
309 neurons remaining after the selections by short- or long-term correlations were included in the
310 classification analyses for reliability.

311 For the principal component analysis, we used `prcomp()` command in R. As a preprocessing, we
312 first averaged the neural responses to each stimuli, in order not to include the trial-to-trial variability in
313 the visualization by principal components. That is, we essentially visualized the tuning curves. In
314 addition, here we did not standardize the activity of each neuron or tuning curve, because we did not
315 want to enlarge small noises within bad neurons who do not respond to any stimuli at all. That is, not to
316 listen to purely noisy neurons too much, we did not enlarge the tuning curves even if their amplitudes
317 are small. In Figure 6 (b), the same neuron pool as in Figure 6 (c, right), i.e. the neurons with pure long-
318 term correlations, was used (189 neurons from 23 sessions).

319 **2.8 V1 neuronal spikes**

320 The experimental details for the cat V1 anesthetized recordings we reanalyzed have been previously
321 described (Maruyama & Ito (2013, 2017)). Briefly, 566 neurons were recorded in 48 sessions with 640
322 trials (40 repeats of 16 visual stimuli) from five adult male cats. Two types of electrode arrays were
323 adopted for the recordings: a four-tetrode array and an array of eight single microelectrodes, both of
324 which were fabricated in the laboratory.

325 The eyes were focused on the tangential screen at a distance of 57 cm using the tapetal reflection
326 technique and an appropriate set of gas-permeable contact lenses. The pupils were dilated using
327 phenylephrine hydrochloride (Neosynesis eye solution). All animal procedures were performed in
328 accordance with the [Author University] animal care committee's regulations.

329 Once stable recordings were obtained, the receptive field properties (location) of the multi-unit
330 activity recorded by each electrode were mapped, using a mouse-controlled moving light bar presented
331 on a 21-inch color monitor (1024 x 768 resolution, vertical refresh rate of 80 Hz) at a distance of 57 cm
332 from the eyes. Because the receptive fields of the units recorded by the high-density electrode arrays
333 had significant overlap, The units were stimulated by moving the light bars on a dark background
334 crossing over the region covering all of the receptive fields. The stimuli consist of the light bars of 16
335 orientations equally spaced (i.e. with an angular separation of 22.5°) that move along the direction of
336 the normal. We ran 40 trial blocks in which each of the 16 stimuli were presented in a pseudo-random
337 order with an intertrial interval of 3s. The bars traveled an angular distance of 3–5° over a period of 1.0–
338 1.7 s (speed 3°/s).

339 Multi-unit activities recorded by each electrode were sorted to recover the activities of individual
340 single units using custom spike sorting software (Gray et al. (1995)).

341 **3 Results**

342 For the purpose of measuring the spike-count noise correlations in different timescales and assessing
343 their respective impacts on neural representations, we utilized the novel information geometric

344 estimators of pure short-term correlations, which can be dissociated from long-term correlations in a
345 nonparametric manner, that is, whatever the baseline drifts are. Before we applied this proposed
346 method to the neural responses in V1, we checked if and how it worked for the simulated time series as
347 a benchmark.

348 **3.1 Proposed estimator works irrespective of baseline drifts**

349 First we randomly generated the artificial time series which mimic the activities of two neurons,
350 whose baselines drift across many trials. Note that nonstationarity, often observed experimentally in an
351 unreproducible manner, was indispensable for the simulation, as we wanted to see if the proposed
352 method can overcome it. In the numerical simulation in Figure 1 (a), the activities of two neurons were
353 created by adding the bivariate Gaussian noises to the smoothly drifting trends, which, in turn, were
354 independently generated for the two neurons by ARIMA(0,2,1) model whose moving average coefficient
355 was 0.6 (Harvey (1993)). Here a significant short-term noise correlation ($\rho = 0.3$) was induced only
356 between simultaneous noises for two neurons, mimicking typical neuroscience experiments where
357 significant trial intervals of seconds order wash out inter-trial temporal correlations in spike counts. An
358 example realization of the simulation in Figure 1 (a), that mimics one recording session, shows
359 hallmark drifting baselines, which is definitely unreproducible and hard to estimate with limited
360 sample number or from this "single snapshot" data. Note that here we exclusively consider trials as a
361 unit for time axes, instead of fine scale windows such as 1ms bins.

362 Figure 1 (b) shows the cross correlation functions for the realization of simulated activities for two
363 neurons in Figure 1 (a) computed by both the conventional correlogram and the proposed method (Eqs.
364 1 and 2). Here, the correlation coefficient ρ was estimated for each time-shifted data, where the activities
365 of one neuron was time-shifted while those for the other neuron was kept. Because of the wrong
366 assumption of the constant baselines, the conventional correlation coefficients caused a broad cross
367 correlation function attributable to the temporal correlations in the baselines. That is, the correlation
368 coefficients is positive because when the activity of neuron 1 is higher (lower) than its average at a late
369 (early) trial, that of neuron 2 is also higher (lower). Adding a time shift does not affect this situation as
370 there is a global trend in Figure 1 (a). Note that broad cross correlation functions have been observed
371 for the experimental data (Bair et al. (2001)). On the other hands, the proposed method gave a
372 satisfiable result, correctly causing 0 for the time shifted data and the short-term correlation $\rho (= 0.3)$
373 for the simultaneous data as demonstrated by a clear peak in Figure 1 (b). Note that the estimated
374 correlation coefficient $\hat{\rho} (= 0.3)$ is not only useful for statistical tests but also interpretable as a
375 simultaneous covariation of Gaussian noises because our method is statistical model-based.

376 The reason for the flexible estimation by the proposed method is that it estimates the covariance for
377 two neurons within each local window, where the background activity is assumed to be almost
378 constant, and, then, averages the local estimates across sliding windows as in Figure 1 (c). Note that our
379 method is based on the assumption that the short-term correlation (or the covariance parameter of
380 Gaussian noises) is constant over time. Consequently, the proposed method enables estimation of the
381 short-term correlations existing in the simultaneous activities independently of the drifting baselines.
382 Figure 1 (c) shows how this method works for the cases with pure long- (top) or short-term
383 correlations (bottom). In the case of pure long-term correlations in Figure 1 (c, top), the estimate of the
384 correlation in the short window is zero (on average), as there is no real short-term correlation and the
385 baseline drift is negligible in this short timescale. Note that an implicit assumption in the proposed
386 method is that within a short window, the baseline drift is absent or negligible, although the violation of
387 this assumption, if small enough, actually does not matter (Materials and Methods). In the case of pure
388 short-term correlations in Figure 1 (c, bottom), the estimate of the correlation in the short window is
389 nonzero (on average), as there is a real short-term correlation although the baseline drift is absent. In
390 this way, the proposed "local" estimates, that can be unaffected by the slow, long-term trends, work
391 fairly well even if the baseline activities drift arbitrarily over time.

392 **3.2 Proposed estimator requires less neurons and computational powers than**
393 **conventional Kalman filters**

394 The key idea for the proposed estimator of noise correlations resides in the local detrending. However
395 there are other types of detrending methods such as Kalman filters. The latest studies also computed
396 the correlations in residuals after the neural activities were smoothed and detrended by the Kalman
397 filter-like methods (Ecker et al. (2014), Rosenbaum et al. (2017)). Therefore we performed another
398 benchmark simulation to compare the conventional Kalman filter method with the proposed method.
399 Specifically, we checked if the two methods work in the presence of sinusoidal baseline drifts in
400 simulations.

401 Figure 2 (a) shows the activities of two neurons, simulated as the time series of length 100 with the
402 common sinusoidal baseline trend. The activities of two neurons at each time are generated as the sum
403 of the baseline trend and the bivariate Gaussian noises with unit variances and 0.3 correlation coefficient.
404 When we simulated more than two neurons simultaneously, the additional neurons shared the trend
405 but did not have noise correlations. Thus, among N simulated neurons, only neuron 1 and 2 have a
406 nonzero correlation coefficient, which is to be estimated.

407 Figure 2 (b) shows the residual activities after the removal of the estimated trend by the Kalman
408 filter from the activities in (a). The dark horizontal line indicates the estimated trend, which has been
409 already removed from the activities.

410 Figure 2 (c) shows the noise correlations in the residuals averaged across 100 realizations of the
411 simulated data. The horizontal dotted gray line for the true correlation coefficient ($=0.3$) indicates that
412 the conventional Kalman filter method does not work when the number of simultaneously simulated
413 neurons are small. Naturally, recording from more neurons helps to estimate the current baseline trend,
414 which is essentially the average activities of neurons in this easiest situation. If you do not know
415 baseline trends accurately, the estimation of noise correlations fails as well. In more realistic situations,
416 in which neurons do not necessarily share baseline trends, more neurons would be required to estimate
417 the noise correlation by the Kalman filter-like methods.

418 Figure 2 (d) shows the noise correlations estimated by the proposed method from the same data.
419 The horizontal dotted gray line for the true correlation coefficient ($=0.3$) indicates that the proposed
420 method always works. Note the proposed method only requires the activities of the two relevant
421 neurons as evident in Eqs. 1 and 2.

422 Furthermore, the Figure 2 (e) shows that the Kalman filter can be fairly expensive in computational
423 time with as small as 15 neurons. Given the number of simultaneously recorded neurons is increasing
424 rapidly, the computational costs can easily constitute a limiting factor. Thus, the proposed method is
425 advantageous not only in the estimation accuracy, but also in the computational cost as demonstrated in
426 Figure 2 (f).

427 The results obtained here are fairly general. Although the sinusoidal trend with 7 cycles was entirely
428 used in this paper, qualitatively the same results were obtained for a wide range of numbers of cycles (4
429 \leftarrow 10, not shown). Imagine that the sinusoidal waves with different periods can exhaust the different
430 possible timescales. In fact, it has been numerically demonstrated that the proposed method worked
431 also for linear as well as stepwise trends in the previous work (Miura (2013)), although all these
432 numerical simulations just confirmed the mathematical statement that the proposed method is robust
433 against arbitrary drifts. Although the proposed method might look too easy at first glance, any other ad
434 hoc estimators of covariances cannot achieve the unbiasedness (i.e. correctness) under arbitrary drifts.
435 Moreover, although the latest best Bayesian methods can be regarded as variants of Kalman filter
436 methods and some of them might improve the estimation accuracy slightly, we believe that the problem
437 in computational costs is unavoidable in any case.

438 3.3 Proposed estimator has less errors than conventional moving averages

439 As some of the previous works (ex. Cohen & Newsome (2008), Mitchell et al. (2009)) simply used the
 440 moving average for detrending, we next compared the conventional moving average method with the
 441 proposed method (Figure 3).

442 In the comparison, as in Figure 2, the simulated neural activities had the sinusoidal trend with five
 443 waves (left) or four waves in 100 trials (right). For the moving average method, the neural activities
 444 were first smoothed by the moving average with various window sizes and then the correlation coe
 445 cients were computed for the residuals. The horizontal dotted gray lines for the true correlation coe
 446 cient (=0.3) indicate that the biases are prominent for longer window sizes and for rapidly changing
 447 trends.

448 Although the moving average method is uniquely defined for odd window sizes, some variants can
 449 be considered when the window size is two (and even lengths in general). When the window size is two,
 450 however, you can carefully define the moving average method so that it coincides with the proposed
 451 method regarding the correlation coe cients. To be precise, the moving average method actually fails
 452 and underestimates both the variances ($\hat{\sigma}_{11}$, $\hat{\sigma}_{22}$) and the covariance ($\hat{\sigma}_{12}$) by half, although

453
 454 the correlation coe cients as their ratio is intact as $\hat{\rho} = \hat{\sigma}_{12} / \sqrt{\hat{\sigma}_{11} \hat{\sigma}_{22}}$. For example, when the true variances
 455 for the activities of two neurons are both 1 and the true covariance is 0.2, the moving average method
 456 on average estimates them as 0.5, 0.5 and 0.1 while the correlation coe cients estimated as their ratio
 457 coincides with that of the proposed estimator, which is always near 0.2.

458 Some previous works used longer window lengths for detrending (previous and future 20 trials for
 459 Cohen & Newsome (2008), and Gaussian kernels with = 5 trials for Mitchell et al. (2009)). Although it is
 460 not clear if the actual drift is as drastic as in Figure 3, our message in this paper is that, in fact, you can
 461 safely shorten the window length to the minimum size, i.e., two.

462 3.4 Examples of noise correlations in V1 neuron pairs

463 Here we applied the proposed method for estimating pure short-term noise correlations to the pairs of
 464 the neural activities in the primary visual cortex. Figure 4 shows the inter-neuronal noise correlations
 465 of two example pairs of neurons estimated by the proposed method as well as the conventional
 466 crosscorrelogram. We computed time-shifted noise correlations or cross correlation functions. Note
 467 that we solely computed noise correlations for a fixed stimulus in this paper, because recent works
 468 indicated the stimulus dependency of noise correlations (Kohn & Smith (2005), Maruyama & Ito (2013),
 469 Ru^e & Cohen (2016)). For the putatively nonstationary neuron pairs in Figure 4 (a), the time series for
 470 the activities of both neurons showed significant drifts. The conventional correlogram showed the
 471 spurious correlations across wide shifts of trials, while the proposed method indicated no short-term
 472 correlation successfully. Note that similar broad cross correlation functions have been observed
 473 previously (Bair et al. (2001)). For the putatively stationary neuron pairs in Figure 4 (b), the time series
 474 for the both neurons did not show significant drifts but the simultaneous activities tended to
 475 synchronize. Both the conventional correlogram and the proposed method correctly detected the short-
 476 term noise correlation at the origin. Thus the proposed method succeeded to clarify the fine structure of
 477 noises in real V1 data by detecting purely short-term correlations.

478 3.5 Both long- and short-term correlations are widely observed in V1

479 Next we investigate the noise correlations for the entire population of pairs of simultaneously recorded
 480 neurons. Figure 5 (a) and (b) plot the short-term noise correlations estimated by the proposed method
 481 against the conventional correlation coe cients for all the pairs within the stationary or nonstationary
 482 neurons. The stationary or nonstationary neurons were selected by the statistical serial correlation test
 483 for nonstationarity. In Figure 5 (a) for the stationary neuron pool, the correlations are highly

484 reproducible, located along the diagonal line. Meanwhile, in Figure 5 (b) for the nonstationary neuron
485 pool, they are not reproducible, scattered apart from the diagonal line, with smaller absolute values for
486 the proposed method. The result suggests that the proposed method successfully removes longterm
487 components of noise correlations essentially by detrending. Note that some of the smallest noise
488 correlations reported in the previous works were obtained for the detrended time series (Bair et al.
489 (2001), Ecker et al. (2010), Renart et al. (2010)), consistent with our observation. Thus the
490 nonstationarity or a baseline drift may engender spurious correlations even if no actual short-term
491 correlation exists.

492 Figure 5 (c) shows the p-value histogram for the statistical significance of the proposed short-term
493 noise correlations for V1 data. The non-uniformity of the distribution indicates that the significant
494 short-term correlations for some pairs are not obtained by chance. Furthermore, all types of pairs,
495 irrespective of stationary and nonstationary neurons, show significant short-term correlations. As a
496 control to check the validity of our statistical test, Figure 5 (d) shows the p-value histogram for the
497 same test obtained for the 1-time-shifted V1 data that cannot have short-term correlations. The
498 resulting uniform distribution demonstrates that, desirably, the statistical test detects no spurious
499 short-term correlation even if the signals drift in the V1 data. Remember, in contrast, the conventional
500 correlogram in Figure 1(b) resulted in the nonzero correlations even for time-shifted data.

501 In total, significant fractions of noise correlations seem to be explainable by the long-term
502 components while there are some pairs with significant short-term correlations as well. We next pursue
503 if each component is either helpful or harmful for the sensory information representation in the brain.

504 **3.6 Impacts of short- and long-term noise correlations are dissociable**

505 Finally, we assessed the impacts on decoding of the presence of short- or long-term correlations,
506 separately. Our estimator enables us to elucidate the impacts of short- and long-term correlations in a
507 dissociated manner, as we will see. Here we performed the linear discriminant analysis of stimuli based
508 on the neural responses and used the classification success rates as a measure of the accuracy of neural
509 coding. That is, the higher the classification success rate is, the more accurate the neural coding should
510 be.

511 To elucidate the impact of short-term correlations, we compared the classification success rates in
512 the absence and presence of pure short-term correlations in Figure 6 (a). For that purpose, we first
513 selected the neurons who have no long-term correlation. That is, we selected the neurons whose
514 baselines did not drift significantly by using the serial correlation statistical test for randomness of
515 fluctuations (Materials and Methods). For those selected neurons, that cannot have long-term
516 correlations, we compared the classification success rates before and after trial shuffling, which was
517 supposed to remove short-term correlations. We computed the classification success rate session by
518 session, as we wanted to include only simultaneously recorded pairs. We found that the impact of pure
519 short-term correlations was small but significantly positive in Figure 6 (a).

520 Next, as the origin of long-term correlations, we visualized the baseline drifts in Figure 6 (b). For the
521 neurons with significant baseline drifts (to be precise, the same neuron pool as used in Figure 6 (c,
522 right)), we performed the principal component analysis for the average responses to 16 visual stimuli
523 (i.e. tuning curves). The activities of the 189 neurons with baseline drifts were concatenated and
524 transformed ("rotated") to the same numbers of 189 principal components, from which we chose the
525 first two as the (most informative) axes for visualization. Figure 6 (b) plots the average responses
526 (tuning curves) for 1-20th trials (turquoise blue) and 21-40th trials (green) separately and
527 demonstrates the baseline drifts over trials shifted the entire activities of neurons. However it is still
528 unclear from the simple visualization, if this drift is, taking form of long-term correlations, significant in
529 decoding.

530 To elucidate the impact of long-term correlations on decoding, we compared the classification
531 success rates in the absence and presence of pure long-term correlations in Figure 6 (c). Specifically, we

532 compared the cross-validated classification success rates for four types of learning: (1) when trained by
533 former trials and tested by former trials, (2) when trained by former trials and tested by latter trials, (3)
534 when trained by former trials and tested by latter trials after the respective global means were
535 subtracted for detrending (i.e. centering and equating the means of former and latter trials in (b)), (4)
536 when trained by even-numbered trials and tested by odd-numbered trials. Note that the conventional
537 sampling of odd-numbered 20 trials (1st, 3rd, 5th, ..., 39th) included both former and latter trials as a
538 part and, thus, can be inhomogeneous under baseline drifts. For that purpose, we first selected the
539 neurons who have no short-term correlation. That is, we selected the neurons whose short-term
540 correlation is not significant by using the statistical test accompanying our estimator (Materials and
541 Methods). Note that although the test applies to a pair, we eventually selected the neurons who have no
542 short-term correlation in any pair in the session. For those selected neurons, that cannot have short-
543 term correlations, we compared the classification success rates for the neurons with and without long-
544 term correlations (i.e. statistically significant baseline drifts) as in Figure 6 (c). As a control, no
545 significant difference was observed among four types of learning in the absence of pure long-term
546 correlations, that is, when both short- and long-term correlations were absent (left, not significant for
547 all pairs, paired t-test, 11 sessions with 77 neurons). On the other hand, the significant decrease at the
548 green bar in the presence of pure long-term correlations demonstrates that the long-term correlations
549 do harm for generalization (right, \ast : $p < 0.05$, paired t-test, 23 sessions with 189 neurons). The recovery
550 of the classification success by the detrending or the conventional inhomogeneous sampling (trained by
551 even- and tested by odd-numbered trials) suggests that the decrease in decoding accuracy is due to the
552 baseline drift ($\ast\ast\ast$: $p < 0.001$, paired t-test). Note that the last two types of learning may mimic brains'
553 possible decoding strategies under changing environments, suggesting that the brain could overcome
554 nonstationarity by detrending.

555 Here we solely compared the classification success rates obtained for the same neuron pool with
556 different types of learning. This is because we believe that it is dangerous to compare different pools
557 even if the numbers of neurons are equated, as the sensitivity to stimuli varies by neurons, leading to
558 considerable sampling biases. For example, if we compare the two green bars in Figure 6 (c), the
559 classification success rate per neuron trained by former and tested by latter is higher in the presence of
560 long-term correlations (not shown), suggesting that the overall high classification success in the
561 presence of the long-term correlations can be explained by the sampling biases, i.e. simply because the
562 neuron pool with long-term correlations have more smart neurons.

563 Taken together, the proposed method enables us to elucidate the impacts of short- and long-term
564 noise correlations in a dissociated manner. The well-designed decoding analysis with dissociated
565 correlated activities may help to gain insight into the brains' decoding strategies under changing
566 environments.

567 **4 Discussion**

568 In this paper, we proposed an information-geometric method to unbiasedly estimate pure short-term
569 noise correlations irrespective of arbitrarily drifting baselines. The simulation demonstrated the
570 robustness of the proposed estimator against the slow, long-term drift. The accompanying statistical
571 test as well as the existing nonstationarity test enabled us to dissociate short- and long-term
572 correlations. When we exclude the spurious noise correlations of purely long-term nature, only small
573 fraction of V1 neuron pairs showed significant short-term correlations, possibly reconciling the
574 previous inconsistent observations on existence of significant noise correlations. Finally, with the
575 additional help of the machine learning that classifies stimuli from neural activities, we assessed the
576 impacts on decoding of the presence of short- or long-term correlations, separately. The presence of
577 pure short-term correlations slightly improved the decoding accuracy, while the pure long-term
578 correlations deteriorated the generalization ability. However, the decrease in decoding accuracy by the
579 long-term correlations was recoverable by using the decoder with offset, suggesting that the brain could
580 overcome nonstationarity by detrending. Thus, our method enables us to elucidate the functions of
581 short- and long-term correlations in a dissociated manner and the well-designed decoding analysis with

582 dissociated correlated activities may help to gain insight into the brain's decoding strategies under
583 changing environments.

584 Our observation that only small fraction of neuron pairs have short-term noise correlations after
585 detrending may, at first glance, inconsistent with previous works, which reported significant noise
586 correlations. However, the previous works which detrended the time series before calculating noise
587 correlations reported small short-term noise correlations (Bair et al. (2001), Ecker et al. (2010), Renart
588 et al. (2010)). In this sense, our result is consistent with the previous results. The previous modeling
589 studies implied that even if short-term noise correlations are small, it can have a big impact in a large
590 network (Zohary et al. (1994), Sompolinsky et al. (2001), Miura (2012)). As far as our V1 data set, the
591 impact of short-term noise correlations was small but significantly positive.

592 The classification analysis in this paper demonstrated that the presence of baseline drifts decreased
593 the generalizability of the classifier. However, the further analysis showed that the classification success
594 rate can be recovered by detrending data or including more inhomogeneous training data. Note that the
595 generalizability should depend on the training data set: the more different conditions are learned, the
596 higher the classification success rate becomes. In other words, if the future (test) conditions are
597 completely different from the past (learned) conditions, the baseline drifts do harm. Thus, we
598 essentially showed two possible decoding strategies that can overcome nonstationarity. It is interesting
599 to know how the brain decodes visual stimuli from small responses under big spontaneous
600 fluctuations? Can the downstream neurons separate stimuli in the high dimensional space or,
601 alternatively, cancel out the baseline drifts suitably? These questions remained and leave future works,
602 possibly, with well-designed decoding analyses.

603 In this paper, we solely treated spike count correlations as a measure of synchrony. We did not use
604 spike timing cross-correlograms with milli-second bins (Toyama et al. (1981*a,b*), Ito et al. (2010)) as we
605 took advantage of our proposed method, which is limited to spike count correlations. The limitation
606 comes from the assumption of no temporal auto-correlation in the time series. The assumption is
607 necessary to dissociate short- and long-term cross-correlations successfully. If we consider millisecond
608 bins, temporal auto-correlations exist, which violates the assumption. Here we rather focused on spike
609 count correlations to utilize our proposed method in depth to the extent to elucidate the
610 componentwise functions of short- and long-term correlations. Thus we did not say anything on
611 temporal coding in this paper, although previous papers suggested the relationship that the spike count
612 correlations increase with coupling strengths (Cossell et al. (2015), Bharmauria et al. (2016)).

613 There are considerable merits for our proposed estimator of short-term correlations. It guarantees
614 the smallest estimation error among all the estimators which "works" for arbitrary baseline drifts. It
615 utilizes differential geometry essentially and otherwise it is generally impossible to cope with infinitely
616 many cases programmatically even with the fastest computers. As a practical advantage, it enables us to
617 perform a statistical test from a single trial or a snapshot of time series with baseline drifts, which is
618 usually unreproducible. The estimating equation given in an analytically closed form as well as the
619 accompanying statistical test, are quite simple and implementable within a few lines of programming
620 codes, easier than shuffling-based methods which have longer lines and computational time. The
621 underlying statistical model allows us not only to test statistical significance but also to interpret the
622 correlation coefficients quantitatively, which is unrealizable for other ad hoc or shuffling-based methods.

623 Meanwhile, to fully exploit the temporal order and continuity of trials without assuming specific
624 statistical models for trends nonparametrically, we had to consider a simplified additive Gaussian noise
625 model. However, some previous works used more realistic models such as mixed Poisson distributions
626 and, for example, estimated the contributions of additive as well as multiplicative noises to explore the
627 underlying biological processes (Goris et al. (2014), Arandia-Romero et al. (2016)). Thus, it is desired to
628 pursue temporal structures with more realistic statistical models in the future work.

629 It is important to check if the spike count data to be analyzed satisfy the model assumptions of the
630 proposed method. For example, the normality assumption is satisfied by high firing neurons in general
631 due to the central limit theorem. However, strictly speaking, when we checked if the spike count data

632 used in the paper obey the normal distribution by using the Shapiro-Wilk test, only 70.0% of the
 633 neuron-stimulus pairs that are stationary (i.e. without drifts) and modest-firing (i.e. more than 5Hz)
 634 satisfied the normality assumption. One possible solution might be to apply the proposed method only
 635 to the high firing neurons as low firing neurons tend to violate the normality assumption. If the data do
 636 not satisfy the assumption of the normal noises, the proposed noise correlation is no more an optimal
 637 parameter estimate of the statistical model. It is also important to check if the assumption of the
 638 constant covariances is satisfied, at least for some time range. For example, strongly nonstationary
 639 neurons, whose firing rates grow twofold over an hour, might violate the assumption. We leave the
 640 detailed examination of the model assumptions with statistical model selection procedures for the
 641 future works. However, if the violation is weak, the proposed measure could still be used as a rough
 642 measure. For example, even if the data were actually non-Gaussian spike counts with multiplicative
 643 drifts (Goris et al. (2014)), the sign of the proposed measure, excitatory or inhibitory, could still be
 644 meaningful.

645 Another assumption for the proposed estimator was that the baseline activities for the consecutive
 646 two trials are (almost) the same. This assumption in our analysis was the clue to separate short- and
 647 long-timescales. Strictly speaking, however, as we computed noise correlations separately for different
 648 stimuli, the intervals between the trials for the same stimulus are variable. Note that stimuli were
 649 presented in a pseudo-random order. In fact, for the worst case, the effective trial interval can be as
 650 large as 90 sec (3 sec \rightarrow 15 stimulus \rightarrow 2). Although it is generally hard to characterize the effects of
 651 drifts on these medium timescales, no difference was observed between randomized and repeated
 652 orders of stimulus presentations (Kohn & Smith (2005)). Thus, we assumed that the drifts on these
 653 medium timescales were ignorable. Practically, if the assumption of the constant baseline is doubtful for
 654 a trial pair due to the long interval between them, you could remove the pair from the calculation of the
 655 proposed estimator. That is, you could exclude unreliable trial pairs from the summation in Eq. 2. This
 656 type of exception handling could also work for avoiding the change point where the baselines jump
 657 suddenly. Developing a more flexible algorithm for the proposed method can be a future work.

658 Acknowledgments

659 This work was supported by ... The authors declare no competing financial interests.

660 References

- 661 Abbott, L. F. & Dayan, P. (1999), 'The effect of correlated variability on the accuracy of a population
 662 code', *Neural Comput.* **11**, 91–101.
- 663 Amari, S. (2016), *Information Geometry and Its Applications*, Springer Japan, Tokyo.
- 664 Amari, S. & Cardoso, J. (1997), 'Blind source separation-semiparametric statistical approach', *IEEE*
 665 *Trans. on Signal Processing* **45**, 2692–700.
- 666 Amari, S. & Kawanabe, M. (1997), 'Information geometry of estimating functions in semi-parametric
 667 statistical models', *Bernoulli* **3**, 29–54.
- 668 Amari, S. & Nagaoka, H. (2001), *Methods of Information Geometry*, American Mathematical Society,
 669 Providence, RI.
- 670 Arandia-Romero, I., Tanabe, S., Drugowitsch, J., Kohn, A. & Moreno-Bote, R. (2016), 'Multiplicative and
 671 additive modulation of neuronal tuning with population activity affects encoded information', *Neuron*
 672 **89**, 1305–1316.
- 673 Averbeck, B. B., Latham, P. E. & Pouget, A. (2006), 'Neural correlations, population coding and
 674 computation', *Nat. Rev. Neurosci.* **7**(5), 358–366.

- 675 Bair, W., Zohary, E. & Newsome, W. T. (2001), 'Correlated firing in macaque visual area mt: time scales
676 and relationship to behavior', *J Neurosci.* **21**, 1676–97.
- 677 Bharmauria, V., Bachatene, L., Cattan, S., Chanauria, N., Rouat, J. & Molotchniko, S. (2016), 'High noise
678 correlation between the functionally connected neurons in emergent v1 microcircuits', *Exp Brain Res*
679 **234**, 523–32.
- 680 Bickel, P. J., Klaassen, C. A. J., Ritov, Y. & Wellner, J. A. (1993), *Efficient and adaptive estimation for*
681 *semiparametric models*, Johns Hopkins University Press, Baltimore, MD.
- 682 Cohen, M. R. & Kohn, A. (2011), 'Measuring and interpreting neuronal correlations', *Nat Neurosci.* **14**,
683 811–9.
- 684 Cohen, M. R. & Newsome, W. T. (2008), 'Context-dependent changes in functional circuitry in visual area
685 mt', *Neuron* **60**, 162–73.
- 686 Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., Hofer, S. B. & MrsicFlogel, T. D.
687 (2015), 'Functional organization of excitatory synaptic strength in primary visual cortex', *Nature* **518**,
688 399–403.
- 689 Dayan, P. & Abbott, L. F. (2001), *Theoretical Neuroscience*, MIT Press, Cambridge.
- 690 Doiron, B., Litwin-Kumar, A., Rosenbaum, R., Ocker, G. K. & Josic, K. (2016), 'The mechanics of state-
691 dependent neural correlations', *Nat Neurosci* **19**, 383–93.
- 692 Ecker, A. S., Berens, P., Cotton, R. J., Subramanian, M., Denfield, G. H., Cadwell, C. R., Smirnakis, S. M.,
693 Bethge, M. & Tolias, A. S. (2014), 'State dependence of noise correlations in macaque primary visual
694 cortex', *Neuron* **82**, 235–48.
- 695 Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., Logothetis, N. K. & Tolias, A. S. (2010), 'Decorrelated
696 neuronal firing in cortical microcircuits', *Science* **327**, 584–7.
- 697 Fiser, J., Chiu, C. & Weliky, M. (2004), 'Small modulation of ongoing cortical dynamics by sensory input
698 during natural vision', *Nature* **431**, 573–8.
- 699 Goris, R. L., Movshon, J. A. & Simoncelli, E. P. (2014), 'Partitioning neuronal variability', *Nat Neurosci* **17**,
700 858–65.
- 701 Gray, C. M., Maldonado, P. E., Wilson, M. & McNaughton, B. (1995), 'Tetrodes markedly improve the
702 reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex', *J*
703 *Neurosci* **63**, 43–54.
- 704 Harvey, A. C. (1993), *Time Series Models*, MIT Press, Cambridge.
- 705 Ikegaya, Y., Aaron, G., Cossart, R., Aronov, D., Lampl, I., Ferster, D. & Yuste, R. (2004), 'Synfire chains and
706 cortical songs: temporal modules of cortical activity', *Science* **304**, 559–64.
- 707 Ito, H., Maldonado, P. E. & Gray, C. M. (2010), 'Dynamics of stimulus-evoked spike timing correlations in
708 the cat lateral geniculate nucleus', *J Neurophysiol* **104**, 3276–92.
- 709 Kanji, G. K. (2006), *100 Statistical Tests*, Sage, London.
- 710 Kohn, A. & Smith, M. A. (2005), 'Stimulus dependence of neuronal correlation in primary visual cortex of
711 the macaque', *J Neurosci* **25**, 3661–73.
- 712 Latham, P. E. & Nirenberg, S. (2005), 'Synergy, redundancy, and independence in population codes,
713 revisited', *Journal of Neuroscience* **25**, 5195–5206.

- 714 Latham, P. E. & Roudi, Y. (2013), 'Role of correlations in population coding', in R. Q. Quiroga and S. Panzeri
715 (eds.) *Principles of Neural Coding*. Boca Raton, Florida: CRC Press pp. 121–38.
- 716 Luczak, A., McNaughton, B. L. & Harris, K. D. (2015), 'Packet-based communication in the cortex', *Nat*
717 *Rev Neurosci* **16**, 745–55.
- 718 Maruyama, Y. & Ito, H. (2013), 'Diversity, heterogeneity and orientation-dependent variation of spike
719 count correlation in the cat visual cortex', *Eur J Neurosci* **38**, 3611–27.
- 720 Maruyama, Y. & Ito, H. (2017), 'Design of multielectrode arrays for uniform sampling of different
721 orientations of tuned unit populations in the cat visual cortex', *Neurosci Res* **122**, 51–63.
- 722 Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. (2009), 'Spatial attention decorrelates intrinsic activity
723 fluctuations in macaque area v4', *Neuron* **63**, 879–88.
- 724 Miura, K. (2011), 'An introduction to maximum likelihood estimation and information geometry',
725 *Interdisciplinary Information Sciences* **17**, 155–174.
- 726 Miura, K. (2012), 'Effects of noise correlations on population coding', *Proc. of SCIS-ISIS 2012* pp. 1072–
727 1075.
- 728 Miura, K. (2013), 'A semiparametric covariance estimator immune to arbitrary signal drift',
729 *Interdisciplinary Information Sciences* **19**, 35–41.
- 730 Miura, K., Mainen, Z. F. & Uchida, N. (2012), 'Odor representations in olfactory cortex: distributed rate
731 coding and decorrelated population activity', *Neuron* **74**.
- 732 Miura, K., Okada, M. & Amari, S. (2006a), 'Estimating spiking irregularities under changing
733 environments', *Neural Comput.* **18**, 2359–86.
- 734 Miura, K., Okada, M. & Amari, S. (2006b), 'Unbiased estimator of shape parameter for spiking
735 irregularities under changing environments', *Advances in Neural Information Processing Systems* **18**,
736 891–
737 8.
- 738 Miura, K., Tsubo, Y., Okada, M. & Fukai, T. (2007), 'Balanced excitatory and inhibitory inputs to cortical
739 neurons decouple firing irregularity from rate modulations', *J Neurosci.* **27**, 13802–12.
- 740 Miura, K. & Uchida, N. (2008), 'A rate-independent measure of irregularity for event series and its
741 application to neural spiking activity', *47th IEEE Conference on Decision and Control* pp. 2006–11.
- 742 Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P. & Pouget, A. (2014), 'Information-
743 limiting correlations', *Nat Neurosci* **17**, 1410–7.
- 744 Neyman, J. & Scott, E. L. (1948), 'Consistent estimates based on partially consistent observations',
745 *Econometrica* **16**, 1–32.
- 746 Okun, M., Steinmetz, N. A., Cossell, L., Iacaruso, M. F., Ko, H., Bartho, P., Moore, T., Hofer, S. B., Mrsic-
747 Flogel, T. D., Carandini, M. & Harris, K. D. (2015), 'Diverse coupling of neurons to populations in
748 sensory cortex', *Nature* **521**, 511–5.
- 749 Petris, G., Petrone, S. & Campagnoli, P. (2009), *Dynamic Linear Models with R*, Springer New York, New
750 York.
- 751 Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A. & Harris, K. D. (2010), 'The
752 asynchronous state in cortical circuits', *Science* **327**, 587–90.

- 753 Rosenbaum, R., Smith, M. A., Kohn, A., Rubin, J. E. & Doiron, B. (2017), 'The spatial structure of
754 correlated neuronal variability', *Nat Neurosci* **20**, 107–114.
- 755 Ru^e, D. A. & Cohen, M. R. (2016), 'Stimulus dependence of correlated variability across cortical areas', *J*
756 *Neurosci* **36**, 7546–56.
- 757 Sasaki, T., Matsuki, N. & Ikegaya, Y. (2007), 'Metastability of active ca3 networks', *J Neurosci* **27**, 517–28.
- 758 Schneidman, E., Bialek, W. & Berry, M. J. (2003), 'Synergy, redundancy, and independence in population
759 codes', *Journal of Neuroscience* **23**, 11539–11553.
- 760 Sharpee, T. O. (2017), 'Optimizing neural information capacity through discretization', *Neuron* **94**, 954–
761 960.
- 762 Sharpee, T. O., Sugihara, H., Kurgansky, A. V., Rebrik, S. P., Stryker, M. P. & Miller, K. D. (2006), 'Adaptive
763 filtering enhances information transmission in visual cortex', *Nature* **439**, 936–42.
- 764 Sompolinsky, H., Yoon, H., Kang, K. & Shamir, M. (2001), 'Population coding in neuronal systems with
765 correlated noise', *Phys. Rev. E* **64**, 051904.
- 766 Steinmetz, N. A., Koch, C., Harris, K. D. & Carandini, M. (2018), 'Challenges and opportunities for large-
767 scale electrophysiology with neuropixels probes', *Curr Opin Neurobiol* **50**, 92–100.
- 768 Toyama, K., Kimura, M. & Tanaka, K. (1981a), 'Cross-correlation analysis of interneuronal connectivity
769 in cat visual cortex', *J Neurophysiol* **46**, 191–201.
- 770 Toyama, K., Kimura, M. & Tanaka, K. (1981b), 'Organization of cat visual cortex as investigated by cross-
771 correlation technique', *J Neurophysiol* **46**, 202–14.
- 772 van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press, Cambridge.
- 773 Zohary, E., Shadlen, M. N. & Newsome, W. T. (1994), 'Correlated neuronal discharge rate and its
774 implications for psychophysical performance', *Nature* **370**, 140–35.
775
776
777
778
779
780
781
782
783
784
785

786

787

788

789

790 Figure 1: Comparison of conventional cross-correlogram and proposed method. (a) Artificial activities
791 of two neurons, simulated as sums of baselines and trial-to-trial noises. The thick gray smooth curves
792 denote time-dependent baselines μ generated by the ARIMA(0,2,1) model, on which the bivariate
793 Gaussian noises were added to generate the neural activities. The added noises have significant spacial
794 or inter-neuronal correlations but no temporal correlation because inter-trial intervals are assumed to
795 be fairly long (≈ 3 s). (b) The estimated cross correlations for the simulated activities in (a) by the
796 proposed method (red) and the conventional correlogram (black). Only the proposed method works
797 and shows a proper peak at the origin. (c) Schematic illustrations of how the proposed method works
798 for the cases with pure long- or short-term correlations. The cross correlation computed within each
799 local window, where the baselines are instantaneously constant, are averaged across sliding windows
800 to capture only short-term correlations whatever the baseline drift is.

801

802

803

804 Figure 2: Comparison of conventional Kalman filter method and proposed method. (a) The simulated
805 activities of two neurons (red and blue) for 100 trials with the common sinusoidal baseline trend. The
806 thick gray line denotes the model trend used for the data generation. The activities of two neurons at each
807 time are generated as the sum of the baseline trend and the bivariate Gaussian noises with unit variances
808 and 0.3 correlation coefficient. When we simulated more than two neurons simultaneously, the additional
809 neurons shared the trend but did not have noise correlation (not shown). Thus, among N simulated
810 neurons, only neuron 1 and 2 have nonzero correlation coefficient, which is to be estimated. (b) The
811 residual activities after the removal of the estimated trend by the Kalman filter from the activities in (a).
812 (c) The noise correlations in the residuals averaged across 100 realizations of the simulated data. The
813 horizontal dotted gray line for the true correlation coefficient ($=0.3$) indicates that the conventional Kalman
814 filter method does not work when the number of simultaneously simulated neurons are small. The
815 errorbars representing the standard deviation demonstrate the large trial-to-trial variability in the results.
816 (d) The noise correlations estimated by the proposed method from the same data. The horizontal dotted
817 gray line for the true correlation coefficient ($=0.3$) indicates that the proposed method always works. The
818 errorbars representing the standard deviation demonstrate the small variability in the results. (e) The
819 computational time for the conventional Kalman filter method. (f) The computational time for the
820 proposed method.

821

822 Figure 3: Comparison of conventional moving average method and proposed method. As in Figure 2, the
823 simulated neural activities had the sinusoidal trend with five waves (left) or four waves in 100 trials
824 (right). For the moving average method, the neural activities were first smoothed by the moving
825 average with various window sizes and then the correlation coefficients were computed for the residuals.

826 The mean \pm SD of the estimated noise correlations across 100 realizations of the simulated data
827 were correct ($=0.3$) indicate that the

828 plotted. The horizontal dotted gray lines for the true correlation coefficients are
 829 prominent for longer window sizes and for rapidly changing trends.
 830

831

832

833

834

835 Figure 4: Examples of noise correlations for two V1 neuron pairs. (a) For the nonstationary case where
 836 the time series for the both neurons show significant drifts (left), the broad cross correlation was
 837 estimated by the conventional cross correlogram ($p = 0.00012$ at the origin) but no short-term
 838 correlation by the proposed method ($p = 0.92$ at the origin). (b) For the stationary case where the time
 839 series for the both neurons do not show significant drifts but the simultaneous activities tend to
 840 synchronize, the narrow cross correlation at the origin was estimated by both the conventional
 841 correlogram ($p < 10^{-5}$) and the proposed method ($p < 10^{-5}$).
 842
 843

844 Figure 5: Population summary of noise correlations for all recorded V1 neurons. (a) The proposed
 845 short-term noise correlations plotted against the conventional correlation coefficients for all the
 846 simultaneously recorded pairs of stationary V1 neurons ($n = 12931$ pairs). The stationary or
 847 nonstationary neurons were selected by the statistical serial correlation test for nonstationarity. The
 848 numbers along the axes denote the mean \pm SEM. (b) Same plot for all the simultaneously recorded pairs
 849 of nonstationary V1 neurons ($n = 18891$ pairs). Note that the correlations are highly reproducible located along the
 850 diagonal for the stationary neuron pool but not reproducible for the nonstationary neuron
 851 pool, suggesting that the proposed method successfully removes long-term noise correlations by
 852 detrending. (c) The distribution of the p-values for the statistical significance of the proposed short-
 853 term noise correlations for V1 data. ss denotes the pair of two stationary neurons. sn denotes the pair
 854 of stationary and nonstationary neurons. nn denotes the pair of two nonstationary neurons. The non-
 855 uniformity of the distribution indicates that the significant short-term correlations for some pairs are
 856 not obtained by chance. (d) The control distribution of the p-values for the same test obtained for the 1-
 857 time-shifted V1 data that cannot have short-term correlations. The uniform distribution demonstrates
 858 that, desirably, the statistical test detects no spurious short-term correlation even if the signals drift in
 859 the V1 data. Note that, in contrast, the conventional correlogram in Figure 1(b) resulted in the nonzero
 860 correlations even for time-shifted data.
 861

862 Figure 6: Impacts of short- and long-term components of correlated activities of V1 neurons. (a) The
 863 classification success rates in the absence and presence of pure short-term correlations (mean \pm SEM).
 864 In the presence of pure short-term correlations, the decoding accuracy was slightly improved ($p =$
 865 0.0044 , paired t-test, 15 sessions with 134 neurons). Note that the chance level is $\frac{1}{16}$ (= 6.25%), as 16
 866 stimuli were decoded. (b) The baseline drifts, which cause long-term correlations, were visualized by
 867 the principal component analysis for the average responses to 16 visual stimuli (tuning curves) of the
 868 neurons with significant baseline drifts (to be precise, the same neuron pool as in (c, right)). The
 869 average responses for 1-20th trials (turquoise blue) and 21-40th trials (green) demonstrate that the
 870 entire activities of neurons shift over trials. (c) Decoding accuracy in the absence and presence of pure
 871 long-term correlations. The cross-validated classification success rates for four types of learning were

872 compared: (1) when trained by former trials and tested by former trials, (2) when trained by former
873 trials and tested by latter trials, (3) when trained by former trials and tested by latter trials after the
874 respective global means were subtracted for detrending (i.e. centering and equating the means of
875 former and latter trials in (b)), (4) when trained by even-numbered trials and tested by odd-numbered
876 trials. Note that the conventional sampling of odd-numbered 20 trials (1st, 3rd, 5th, ..., 39th) included
877 both former and latter trials as a part and, thus, can be inhomogeneous under baseline drifts. No
878 significant difference was observed among four types of learning in the absence of pure long-term
879 correlations, that is, when both short- and long-term correlations were absent (left, not significant for
880 all pairs, paired t-test, 11 sessions with 77 neurons). The significant decrease at the green bar in the
881 presence of pure long-term correlations demonstrates that the long-term correlations do harm for
882 generalization (right, $p < 0.05$, paired t-test, 23 sessions with 189 neurons). The recovery of the
883 classification success by the detrending or the conventional inhomogeneous sampling (trained by even-
884 and tested by odd-numbered trials) suggests that the brain can decode stimulus information under
885 changing environments by using a sophisticated decoder ($p < 0.001$, paired t-test).











