The Annals of Applied Statistics
2023, Vol. 17, No. 1, 178–198
https://doi.org/10.1214/22-AOAS1621
© Institute of Mathematical Statistics, 2023

CONEX–CONNECT: LEARNING PATTERNS IN EXTREMAL BRAIN CONNECTIVITY FROM MULTICHANNEL EEG DATA

BY MATHEUS B. GUERRERO[a], RAPHAËL HUSER[b] AND HERNANDO OMBAO[c]

Statistics Program, CEMSE Division, King Abdullah University of Science and Technology (KAUST),
[a]*matheus.bartologuerrero@kaust.edu.sa,* [b]*raphael.huser@kaust.edu.sa,* [c]*hernando.ombao@kaust.edu.sa*

Epilepsy is a chronic neurological disorder; it affects more than 50 million people globally. An epileptic seizure acts like a temporary shock to the neuronal system, disrupting normal electrical activity in the brain. Epilepsy is frequently diagnosed with electroencephalograms (EEGs). Current methods study only the time-varying spectra and coherence but do not directly model changes in extreme behavior, neglecting the fact that neuronal oscillations exhibit non-Gaussian heavy-tailed probability distributions. To overcome this limitation, we propose a new approach to characterize brain connectivity based on the joint tail (i.e., extreme) behavior of the EEGs. Our proposed method, the conditional extremal dependence for brain connectivity (Conex–Connect), is a pioneering approach that links the association between extreme values of higher oscillations at a reference channel with the other brain network channels. Using the Conex–Connect method, we discover changes in the extremal dependence driven by the activity at the foci of the epileptic seizure. Our model-based approach reveals that, preseizure, the dependence is notably stable for all channels when conditioning on extreme values of the focal seizure area. By contrast, the dependence between channels is weaker during the seizure, and dependence patterns are more "chaotic." Using the Conex–Connect method, we identified the high-frequency oscillations as the most relevant features, explaining the conditional extremal dependence of brain connectivity.

1. Introduction. Electroencephalograms (EEGs) are multidimensional spatiotemporal signals that measure brain electrical activity from electrodes placed on the scalp. EEGs capture changes in brain signals, following a shock to the neuronal system, such as an external stimulus, stroke, or epilepsy. These shocks show a profound impact on the neuronal system, including changes in frequency content, wave amplitudes, and connectivity structure of the network of signals. Often, these shocks have an impact on the distributions of the observed signals (in particular, at the tails). Here, our goal is to develop a new statistical approach for investigating changes in the *extremal dependence* structure, that is, the dependence structure prevailing in the tail of the distribution, between EEG channels during an epileptic seizure. While classical methods mostly rely on the behavior in the bulk (or around the center) of the distribution (Acharya et al. (2013)), our proposed extreme-value method is natural and, theoretically, justified for modeling extremely large signal amplitudes that are observed during the onset of an epileptic seizure. Unlike classical methods, our proposed approach is able to properly quantify whether an extreme shock in a reference channel is likely to further induce extreme shocks in other channels. Therefore, the proposed method, which provides new insights into "extremal" brain connectivity, may have a potential impact on public health, given that epilepsy affects nearly 50 million people worldwide (WHO (2019)). The anticipated impact of this method will be a deeper understanding of the etiology of seizures and refined diagnosis brought about by the potential ability to differentiate between seizure subtypes and

Received June 2021; revised December 2021.
Key words and phrases. Epilepsy, extreme-value theory, conditional extremes, penalized likelihood, nonstationary time series.

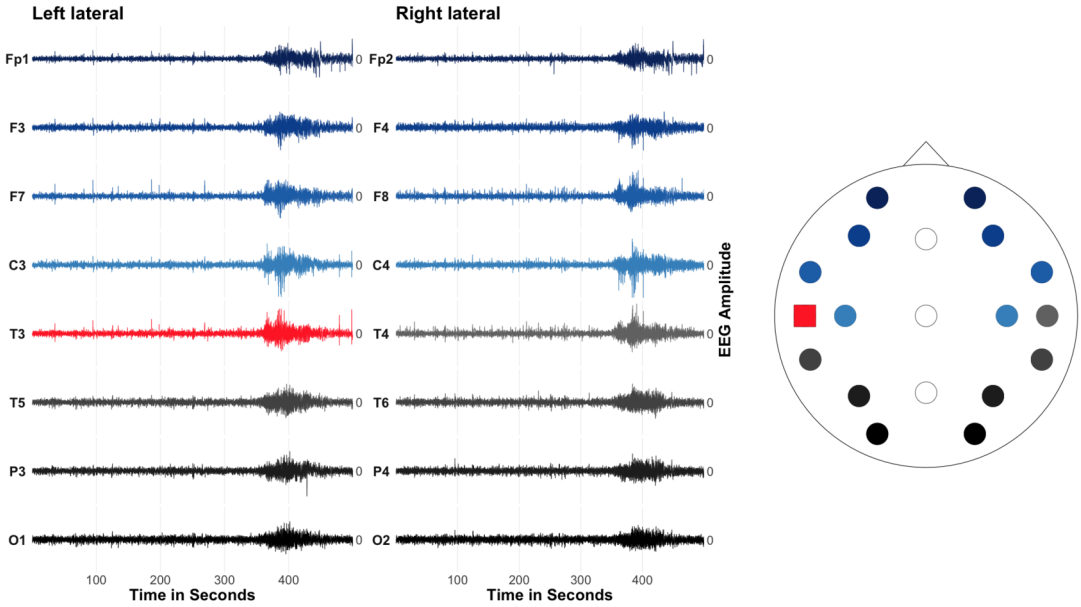

FIG. 1. *Recorded EEG signals during an epileptic seizure (on a 10–20 electrode system) from a female patient diagnosed with left temporal lobe epilepsy. The seizure onset is roughly at $t = 350$ seconds on the reference channel T3 (square), the most likely focal point of seizure discharges for this particular patient. The entire EEG recording consists of 50,000 time points at a sampling rate of 100 Hz. The central temporal (Fz), central (Cz) and central parietal (Pz) channels are not displayed in the figure, albeit they are included in the analysis.*

hence develop more targeted treatments. Figure 1 displays EEG traces of the left and right sides of the brain during an epileptic seizure from a patient previously diagnosed with left temporal lobe epilepsy. In this paper we set the left temporal channel T3 (denoted as the square) to be the reference channel.

One of the frequently-used measures of dependence is cross-correlation. Consider a bivariate random vector $(X, Y)^T$ with joint density $f(x, y)$ having support on the domain $\mathcal{D}$ with marginal means $\mathrm{E}(X) = \mathrm{E}(Y) = 0$ and unit variances $\mathrm{var}(X) = \mathrm{var}(Y) = 1$. Then, the cross-correlation between X and Y is $\rho_{XY} = \mathrm{E}(XY) = \int_{\mathcal{D}} xyf(x, y)\,\mathrm{d}x\,\mathrm{d}y$, where the integral is calculated across the entire domain $\mathcal{D}$. Indeed, classical time-domain and frequency-domain measures of dependence, in multivariate time series, are all derived by some averaging across the *entire* domain $\mathcal{D}$. The main limitation of these measures is that they cannot capture local dependence. It is blind to conflicting variations in the dependence structure across the domain $\mathcal{D}$. That is, the dependence structure (in a network of channels) at the "center" of the distribution might be different from that at the "tails." For a general overview of extreme-value theory, refer to Davison and Huser (2015).

The study of extremal dependence is especially important in analyzing EEGs recorded during an epileptic seizure when there is an abnormal disruption in electrical activity in the brain that results in increased amplitude and changes in spectral decomposition at the localized focal region. In this paper we develop a new statistical procedure for studying how an extreme signal amplitude, caused by an epileptic seizure at the focal region (left temporal lobe T3), may trigger a change in the signal amplitude of another channel of the brain network. In other words, the proposed method will be utilized to study the conditional extremal brain connectivity during an epileptic seizure.

Our proposed method, conditional extremal dependence for brain connectivity (*Conex–Connect*, in short), jointly utilizes extreme-value theory and spectral analysis to model the conditional extremal dependence of brain connectivity. Conex–Connect uses an improved version of the penalized piecewise constant approach (Ross et al. (2018)) of the conditional

extremes model (Heffernan and Tawn (2004)) for multivariate extremes, applying it to time-varying brain signals. We further extend this model to the spectral setting, where we study the impact of extreme amplitudes of oscillations in one channel on the behavior of other channels. To the best of our knowledge, this is the first paper that examines extremal behavior in the amplitudes of oscillatory components of a multivariate time series.

Consider a random vector $(X, Y)^T$ with marginal cumulative distribution functions (CDFs) F_X and F_Y, respectively. The Heffernan and Tawn (H&T) model is an asymptotically motivated model for the conditional distribution of X, given that Y is large, that is, $X \mid Y > F_Y^{-1}(p)$, as $p \uparrow 1$. In finite samples we condition on Y being larger than its p-quantile, for some (fixed) high-nonexceedance probability $p \in (0, 1)$. The H&T model can capture the two possible regimes of extremal dependence between X and Y: asymptotic independence, which arises when $\chi := \lim_{p\to 1} \Pr\{X > F_X^{-1}(p) \mid Y > F_Y^{-1}(p)\} = 0$, and asymptotic dependence which arises when $\chi > 0$. The distinction between these two regimes is key for extrapolating to higher levels of the conditioning variable Y. However, EEGs display sudden bursts or increases in amplitude at the onset of a seizure which may be expressed at some—but not all—frequency bands. One limitation of the H&T model is that it is unable to distinguish such features and hence needs to be adapted accordingly. Our proposed Conex–Connect method overcomes these limitations: it examines extremal dependence behavior in oscillatory components of brain signals for *specific frequency bands*.

One major novelty in this context is that we apply a conditional multivariate extreme-value model not only to the original time series (i.e., from a time-domain perspective) but also to the time series "dissected" into several frequency bands (i.e., from a frequency-domain perspective). This reveals hidden features of the extremal dependence structure between channels that only affect certain frequency bands but may not be visible by looking at the entire time series. Most of the current work on spectral analysis of electrophysiological data focus on the spectral estimation and its association with behavioral measures and brain states. None of these examine the downstream effect of extremal dependence in a reference channel on the entire network (e.g., Fiecas and Ombao (2016), Krafty and Collinge (2013), Krafty, Hall and Guo (2011), Krafty et al. (2017), Ombao et al. (2018), and Scheffler et al. (2020)). Similarly, none of the work on conditional extreme-value theory (e.g., Heffernan and Tawn (2004), Hilal, Poon and Tawn (2011), Keef, Papastathopoulos and Tawn (2013), and Ewans and Jonathan (2014)) examine extremal dependence from a spectral perspective. This paper fills this gap by developing a novel method for investigating the conditional extremal dependence in a brain network from a spectral-domain perspective. Our empirical study is a first step that paves the way toward a deeper understanding of how the extremal behavior of multivariate time series is related to their frequency content.

Moreover, the H&T model is designed for stationary data with respect to covariates. Thus, it needs to be adapted to seizure EEG signals which are highly nonstationary—both in their marginal and cross-dependence behavior. Thus, we use the piecewise version of the H&T model (Ross et al. (2018)) to flexibly capture the time dynamics of extremal brain connectivity, further adapting it to handle autocorrelated data. Note that it is quite a standard practice to either applying a moving window to a nonstationary time series or to segment into quasi-stationary blocks (see Ombao and Pinto (2021), Ombao, von Sachs and Guo (2005), Ombao et al. (2016)). Because of our novel use of extreme-value theory in the context of spectral analysis, we emphasize that it is natural to exploit this rigorous probabilistic framework to understand brain connectivity in epileptic patients since a seizure event creates a strong abnormal (i.e., extreme) disturbance in brain signals. Moreover, by looking at extreme amplitudes rather than the mean behavior, we can essentially probabilistically assess whether or not a shock in a reference channel is likely to create further big shocks in other channels. This fact cannot be quantified by classical methods that focus on the bulk (typically center of the

distribution) behavior of the EEG signals. Therefore, our Conex–Connect method, based on extreme-value theory, is theoretically and practically justified for analyzing epilepsy, and this paper is a call for further research in this direction.

The remainder of the paper is organized as follows. Section 2 presents the proposed Conex–Connect extreme-value approach to analyze EEG data. In Section 3 we provide an in-depth study of the extremal behavior of EEG signals using the Conex–Connect method. Section 4 concludes and discusses perspectives on future research.

2. Conditional extremal dependence for connectivity (Conex–Connect).

2.1. *General setting and method overview.* Let $\dot{Y}_{dt}$ be the EEG amplitude, in absolute value,[1] of channel $d \in \{1, \ldots, D\}$ at time $t \in \{1, \ldots, T\}$. Alternatively, it may also represent the amplitude of filtered signal component at a given frequency band; see Section 2.4 for more details on spectral decomposition. Figure 1 displays the behavior of these multichannel time series. In a nutshell, our proposed method proceeds as follows. The first step is to select a reference channel (i.e., the *conditioning variable*). In our case it is natural to select the left temporal channel T3 (square; Figure 1) as the reference channel because the neurologist has determined that to be the focal point of seizure discharges for this patient which is labeled as $\dot{Y}_1 = \{\dot{Y}_{1t}\}_{t=1}^{T}$. The remaining channels are defined to be the associated channels (*associated variables*) and are labeled as $\{\dot{Y}_d\}_{d=2}^{D}$, where $\dot{Y}_d = \{\dot{Y}_{dt}\}_{t=1}^{T}$. Next, since the dependence structure in the brain network is thought to evolve across time, we segment the multichannel EEG dataset into B (approximately) time-homogeneous blocks (separately for each pre- and postseizure phase). Following our enhanced version of the piecewise approach of Ross et al. (2018), we then apply a conditional multivariate extreme-value model to each phase separately (conditional on T3), each of which being characterized by autocorrelated nonstationary (but blockwise stationary) data. Observations within the same time block are assumed to have common extremal characteristics. Finally, we model the changes in both the marginal and dependence properties. In the Supplementary Material (Guerrero, Huser and Ombao (2023a)), a sensitivity analysis is presented in which we consider $B = 8$, 12, and 16 time blocks for the pre- and postseizure phases (modeled separately) and demonstrate that the results from Conex–Connect are robust concerning the choice of time block sizes. In the following we consider $B = 12$ blocks per phase as a good compromise providing a robust inference scheme that is still able to detect a lot of local detail.

More precisely, for each channel $d \in \{1, \ldots, D\}$ and each pre- and postseizure phase (modeled separately), we fit a piecewise marginal model by assuming that high-threshold exceedances follow a generalized Pareto (GP) distribution with a time-varying scale parameter $\nu_{db} > 0$, $b \in \{1, \ldots, B\}$. Using the empirical CDF below the threshold, the data are transformed from their original scale to a common marginal scale chosen to be the Laplace distribution, based on the probability integral transform. The original data are denoted by $\dot{Y}_d$, while the transformed Laplace-scale data are denoted by Y_d, $d \in \{1, \ldots, D\}$. The GP marginals are transformed to the same standard scale so that they become comparable. Moreover, the Laplace distribution is chosen because the H&T model expressed on this scale yields

[1]EEGs are often modeled as realizations of zero-mean stochastic processes. Thus, there is no interest in modeling the mean or first moment ($\mathrm{E}(X_t)$). Since EEGs are often characterized as a sum of oscillations at different frequencies, the interest lies in the (absolute) amplitudes of these various oscillations rather than the "direction" of the shocks. This is the justification for using absolute values. It is also worth noting that extremes are scarce by nature, and thus, it is valuable to merge information from both tails which contain the same type of information about extreme shocks to the brain network. This merger leads to a reduction in uncertainty in the estimators.

a wide range of extremal dependence structures, from asymptotic independence with negative association to asymptotic dependence with positive association (Keef, Papastathopoulos and Tawn (2013)).

To estimate the conditional extremal dependence of Y_d, given that Y_1 exceeds a high threshold, we finally fit the H&T model with a time-varying dependence parameter $\alpha_{db} \in [-1, 1]$ which captures the evolution (over time) of the strength of the linear dependence in the joint tails. The specific elements of the H&T model, including the parameter α_{db}, are discussed in Section 2.3. The penalized part comes from roughness-penalization parameters included in the likelihood functions to control the extent of variation of the time-varying parameter estimates for the marginal and dependence models. We assess the uncertainty of the estimates through a bootstrap technique that accounts for autocorrelation. The optimal penalty parameter is obtained using a walk-forward cross-validation procedure (Hyndman and Athanasopoulos (2019)) and designed to keep the temporal dependence in the data. The next subsections provide details on the two stages of the Conex–Connect method: the marginal modeling of EEG channels and their connectivity characterization.

2.2. *Marginal model.* For each channel $\dot{Y}_d = \{\dot{Y}_{dt}\}_{t=1}^{T}$, $d \in \{1, \dots, D\}$, each pre- and postseizure phase, and for some high nonexceedance probability $\tau_d \in \mathcal{T}_d \subset (0, 1)$, we define $\psi_{db}(\tau_d)$ as the empirical τ_d-quantile for the bth time block ($b = 1, \dots, B$). Then, motivated by extreme-value theory, we model the upper tail of $\dot{Y}_{dt}$ using the GP distribution, that is,

$$F_{GP}\big(\dot{y}_{dt}; \xi_d, \nu_{db}, \psi_{db}(\tau_d)\big) = \Pr\big\{\dot{Y}_{dt} \le \dot{y}_{dt} \mid \dot{Y}_{dt} > \psi_{db}(\tau_d)\big\}$$
$$= 1 - \left[1 + \frac{\xi_d}{\nu_{db}}\big\{\dot{y}_{dt} - \psi_{db}(\tau_d)\big\}\right]^{-\frac{1}{\xi_d}},$$

for $\dot{y}_{dt} \in (\psi_{db}(\tau_d), \dot{y}_{db}^{+}]$, where $\dot{y}_{db}^{+} = \psi_{db}(\tau_d) - \nu_{db}/\xi_d$ if $\xi_d < 0$, and $\dot{y}_{db}^{+} = \infty$, otherwise. Here, $\xi_d \in \mathbb{R}$ is the shape parameter, assumed constant across time blocks, while $\{\nu_{db}\}_{b=1}^{B} \in (0, \infty)^B$ are the blockwise (i.e., time-varying) scale parameters. In the above expression it is implicitly understood that time t is contained within the bth time block, but, in practice, we need to identify the correct time block for each time point and to assign parameter values accordingly. For each channel d, the shape parameter ξ_d controls the heaviness of the tail of the GP distribution (when compared to the tail of an exponential distribution), and we here keep it time-constant for reasons of parsimony and because this parameter is usually difficult to estimate. Note that there is a structural change at the seizure onset, but since we model the pre- and postseizure phases separately, the shape parameter can indeed be different across phases. Within each phase, however, the tail behavior is usually more regular, and it makes sense to keep the shape parameter constant, as commonly done in the extremes literature, to guarantee a robust and stable marginal model fit. Depending on the value of ξ_d, there are three types of upper tail: bounded ($\xi_d < 0$), light ($\xi_d = 0$), and heavy ($\xi_d > 0$). As special cases, the GP distribution contains the uniform distribution ($\xi_d = -1$), the exponential ($\xi_d = 0$), and the (shifted) Pareto ($\xi_d > 0$).

Note that each channel has its own nonexceedance probability τ_d, indicating the flexibility of the model in capturing the channel-specific extremal characteristics. In addition, despite τ_d being invariant over time blocks, it does not imply threshold invariance (over time) since the τ_d-quantile is specific to each time block.

Let $\boldsymbol{\theta}_d := (\{\nu_{db}\}_{b=1}^{B}, \xi_d)^T \in \boldsymbol{\Theta} := (0, \infty)^B \times \mathbb{R}$ be the parameter vector of the marginal GP model for channel d. Under the working assumption of independence, the likelihood function is

$$\mathcal{L}_{\tau_d}(\boldsymbol{\theta}_d) = \prod_{b=1}^{B} \prod_{\substack{t \in T_b \\ \dot{y}_{dt} > \psi_{db}(\tau_d)}} \frac{1}{\nu_{db}}\left[1 + \frac{\xi_d}{\nu_{db}}\big\{\dot{y}_{dt} - \psi_{db}(\tau_d)\big\}\right]^{-\frac{1}{\xi_d} - 1},$$

where T_b is the set of time points within block b and $\{\dot{y}_{dt}\}_{t=1}^{T}$ are the observed data. We here assume independence between and within time blocks. In the Supplementary Material (Guerrero, Huser and Ombao (2023a)), we provide tools to verify that the within-block time dependence is relatively short-range compared to block size.

In order to control the variance of temporal fluctuations in the estimated GP scale parameters, ν_{db}, we add a channel-specific roughness-penalization parameter $\lambda_d > 0$. The negative penalized log-likelihood is

$$\ell_{\tau_d,\lambda_d}(\boldsymbol{\theta}_d) = -\log \mathcal{L}_{\tau_d}(\boldsymbol{\theta}_d) + \lambda_d \left\{ \frac{1}{B}\sum_{b=1}^{B} \nu_{db}^2 - \left(\frac{1}{B}\sum_{b=1}^{B} \nu_{db} \right)^2 \right\}. \tag{1}$$

Marginal parameter estimates are obtained by minimizing (1), while λ_d is selected by cross-validation; see Section 2.5. Note that larger λ_d values imply an increased smoothness of estimates of the GP scale parameter ν_{db} across time blocks. Also, given λ_d, each channel has a different penalized log-likelihood; that is, each marginal fit has a different set of parameters, consisting of $(B + 1)$ parameters from the GP fit.

Now, using the probability integral transform, we transform data to the standard Laplace scale. For observations below the threshold $\psi_{db}(\tau_d)$, we use the empirical CDF, denoted here by $F_E(\cdot)$. First, the raw data, $\{\dot{y}_{dt}\}_{t=1}^{T}$, is transformed to the uniform scale, $\{u_{dt}\}_{t=1}^{T}$, as follows: $u_{dt} = \tau_d F_E(\dot{y}_{dt})$, if $\dot{y}_{dt} \le \psi_{db}(\tau_d)$, and $u_{dt} = \tau_d + (1 - \tau_d)F_{GP}(\dot{y}_{dt})$, if $\dot{y}_{dt} > \psi_{db}(\tau_d)$. Then, we use the inverse of the standard Laplace CDF, $F_L(\cdot)$, that is, $y_{dt} = F_L^{-1}(u_{dt}) = \text{sign}(0.5 - u_{dt}) \log(2 \min\{1 - u_{dt}, u_{dt}\})$, to obtain common standard Laplace margins, $\{y_{dt}\}_{t=1}^{T}$.

2.3. *Conditional extremal dependence model.* After fitting the marginal models for all D channels and each pre- and postseizure phase, we obtain a standard Laplace-scale sample $\{y_{1t}, y_{2t}, \ldots, y_{Dt}\}_{t=1}^{T}$, with y_{1t} representing the transformed time series of the reference channel (here, T3). Note that the lowercase $\{y_{1t}, \ldots, y_{Dt}\}_{t=1}^{T}$ denote (transformed) realized values, while the uppercase $\{Y_{1t}, \ldots, Y_{Dt}\}_{t=1}^{T}$ denote the corresponding random variables. The next goal is to study, separately for each phase, the conditional dependence of the associated variables, $\{Y_{2t}, \ldots, Y_{Dt}\}$, given that the conditioning variable, $\{Y_{1t}\}$, takes on a large value (in the upper tail of the distribution) since, in terms of EEG analysis, the interest lies in the dependence structure in the network when the signal amplitude at the reference channel is large. Define $\tilde{\tau} \in \tilde{\mathcal{T}} \subset (0, 1)$ to be a nonexceedance probability such that $\phi(\tilde{\tau})$ is the $\tilde{\tau}$-quantile of the standard Laplace distribution for the reference channel Y_1. Let $\tilde{\boldsymbol{\theta}} = (\{\alpha_{db}\}_{d=2,b=1}^{D,B}, \{\beta_d\}_{d=2}^{D}, \{\mu_d\}_{d=2}^{D}, \{\sigma_d\}_{d=2}^{D})^T \in \tilde{\boldsymbol{\Theta}} := [-1, 1]^{(D-1)B} \times (-\infty, 1]^{(D-1)} \times \mathbb{R}^{(D-1)} \times (0, \infty)^{(D-1)}$ be the parameter vector of the H&T model for all channels. Thus, according to the H&T model, for all $t \in T_b$ (within time block b) such that $y_{1t} > \phi(\tilde{\tau})$, conditional on $Y_{1t} = y_{1t}$, Y_{dt} may be expressed as

$$Y_{dt} = \alpha_{db} y_{1t} + y_{1t}^{\beta_d} W_{dt}, \quad d = 2, \ldots, D, b = 1, \ldots, B, \tag{2}$$

where, for model estimation purposes, the components of the random variable W_{dt} are assumed to be mutually independent and normally distributed with mean μ_d and standard deviation σ_d, both time-constant. In other words, the H&T model is similar to a particular nonlinear regression model of Y_{dt} onto $Y_{1t} = y_{1t}$ with normal errors.

Note that the dependence threshold, $\tilde{\tau}$, based on the reference channel Y_1 on the Laplace scale, is not necessarily the same as the marginal threshold τ_1 on the original scale. The latter is tuned to provide a good marginal fit, while the former is chosen so that there will be a sufficient number of exceedances to guarantee a good H&T dependence model fit.

The parameters $\{\alpha_{db}\}_{d=2,b=1}^{D,B}$ are the "first-order" dependence parameters, and they capture the extent of blockwise linear extremal dependence, between channels Y_d and large Y_1, which is allowed to change over time (blocks). When α_{db} is in the interval $(0, 1]$, there is a positive linear association between Y_d and large Y_1 within time block b. This association becomes stronger as α_{db} increases, with (positive) asymptotic dependence corresponding to $\alpha_{db} = 1$ and $\beta_d = 0$. For $\alpha_{db} \in [-1, 0)$, the association is negative linear and becomes stronger as $\alpha_{db} \to -1$. As $\alpha_{db} \to 0$, the linear dependence in (2) weakens. In particular, when Y_d and large Y_1 are independent, then $\alpha_{db} = \beta_d = 0$. On the other hand, the time-constant parameters $\{\beta_d\}_{d=2}^{D}$ may be considered as capturing the "second-order" dependence characteristics since they specify the spread of the data around the linear relationship given by $\{\alpha_{db}\}$ for increasing values of Y_1. If $\beta_d < 0$, the distribution of the data (around the linear relationship dictated by the α's) becomes tighter for higher values of the conditioning variable. If $\beta_d > 0$, we have the opposite behavior; the distribution of the data has a wider spread around the linear relationship (given by the α's) for higher values of the conditioning variable. Notice that, among all dependence parameters, α_{db} is the only one that is allowed to vary across time-blocks $b = 1, \ldots, B$. The other parameters $(\beta_d, \mu_d, \sigma_d)$ are intentionally kept constant over time in order to reduce the overall model complexity (and thus, estimation uncertainty) and to avoid interferences with the estimation of α_{db} which drives the main dependence feature.

REMARK. The parameters α_{db} and β_d are more informative than classical extremal dependence measures, such as the tail correlation coefficient χ (recall its definition in Section 1), and its complementary version, $\bar{\chi}$ (Coles, Heffernan and Tawn (1999)), which is linked to the coefficient of tail dependence, often denoted by the symbol η (Ledford and Tawn (1996)). While χ is merely a (limiting) conditional exceedance probability that provides information in the asymptotic dependence setting and $\bar{\chi}$ summarizes the joint tail decay rate in the asymptotic independence setting, the parameters α_{db} and β_d describe how a variable grows and fluctuates with the extremes of a conditioning variable, both in the asymptotic dependence and independence settings. Specifically, these parameters provide detailed information about the strength, the "direction" (i.e., positive or negative association), and dispersion (in terms of the conditioning variable) of tail (in)dependence. The asymptotic dependence setting actually corresponds to $\alpha_{db} = 1$ and $\beta_d = 0$, whereas the other valid parameter combinations lead to various asymptotic independence scenarios, which makes the H&T model very general and much more informative than tail dependence measures like χ, $\bar{\chi}$ (or η), or other summary statistics, such as the tail quotient correlation coefficient proposed by Zhang (2008, 2021) and Zhang, Zhang and Cui (2017).

From (2) we may rewrite the model as

$$Y_{dt} \mid Y_{1t} = y_{1t} \sim \mathrm{N}\big(\alpha_{db} y_{1t} + \mu_d y_{1t}^{\beta_d}, (\sigma_d y_{1t}^{\beta_d})^2\big), \tag{3}$$

for all $t \in T_b$ (within time block b) such that $y_{1t} > \phi(\tilde{\tau})$, $d = 2, \ldots, D$, and $b = 1, \ldots, B$. The parameters for each channel, and each pre- and postseizure phase, are estimated by minimizing the negative log-likelihood based on (3), that is,

$$\tilde{\ell}_{\tilde{\tau},d}(\tilde{\boldsymbol{\theta}}) = \sum_{b=1}^{B} \sum_{\substack{t \in T_b \\ y_{1t} > \phi(\tilde{\tau})}} \left\{ \frac{1}{2} \log(2\pi) + \log\big(\sigma_d y_{1t}^{\beta_d}\big) + \frac{1}{2} \left(\frac{y_{dt} - (\alpha_{db} y_{1t} + \mu_d y_{1t}^{\beta_d})}{\sigma_d y_{1t}^{\beta_d}} \right)^2 \right\}. \tag{4}$$

Here, there are $(B+3)$ parameters per channel d and phase. Keef, Papastathopoulos and Tawn (2013) proposed additional constraints on $\tilde{\boldsymbol{\Theta}}$, leading to a smaller set of feasible parameters which reduce the variance of the estimators and overcome complications on the modeling

of negatively associated variables and parameter identifiability. Also, these additional constraints avoid drawing conditional inferences inconsistent with the marginal distributions. Here, we use a pragmatic approach and restrict $\{\beta_d\}_{d=2}^{D}$ to the interval [0, 1], to stabilize estimation and avoid unrealistic dependence behaviors. This restriction provides a reasonably constrained yet flexible model, still capturing the most important different types of extremal dependence (Tawn et al. (2018)). Moreover, note that the choice of a normal distribution may appear arbitrary, but it was suggested by Heffernan and Tawn (2004) as a convenient working assumption that ensures valid inference. Indeed, under mild assumptions, even if the normal distribution is misspecified, the parameters are still guaranteed to be consistent (Das and Resnick (2011), Heffernan and Resnick (2007)); see also Davison (2003), page 147, who studies the asymptotic distribution of the maximum likelihood under wrong model assumptions. Furthermore, for subasymptotic features of the conditional extremes model, see Lugrin, Davison and Tawn (2019).

Although the parameters in (2) have their own individual interpretations, sometimes it can be challenging to draw conclusions on the real behavior of the conditional extremal dependence. Thus, a preferable approach might be to analyze $\tilde{\boldsymbol{\theta}}_{db} = (\alpha_{db}, \beta_d, \mu_d, \sigma_d)^T$, the set of blockwise parameters, jointly through a relevant functional, $f_{db}(\tilde{\boldsymbol{\theta}}_{db})$, involving them all together for $d = 2, \ldots, D$, and $b = 1, \ldots, B$. To do so, we exploit the stochastic representation in (3), but we relax the normal assumption by estimating an upper conditional quantile semiparametrically. To be more specific, we here estimate the 0.975-quantile of Y_{dt}, given that Y_{1t} equals its own 0.975-quantile, by suitably combining the estimated H&T model parameters with the empirical 0.975-quantile of model residuals (obtained after standardizing the data by subtracting the fitted mean and dividing by the fitted standard deviation in (3)). The distribution of this estimated conditional quantile is assessed using a block bootstrap procedure, detailed below in Section 2.5.

While the parameters may be estimated separately for each channel from (4), it is also possible to estimate them jointly (yet, still separately for each pre- and postseizure phase), using a penalized likelihood enforcing a similar degree of smoothness for estimated dependence parameters α_{db} across time blocks, which stabilizes their fluctuations and reduces uncertainty. The joint negative penalized log-likelihood used here is

$$\tilde{\ell}_{\tilde{\tau},\tilde{\lambda}}(\tilde{\boldsymbol{\theta}}) = \sum_{d=2}^{D} \tilde{\ell}_{\tilde{\tau},d}(\tilde{\boldsymbol{\theta}}) + \tilde{\lambda} \sum_{d=2}^{D} \left\{ \frac{1}{B} \sum_{b=1}^{B} \alpha_{db}^2 - \left(\frac{1}{B} \sum_{b=1}^{B} \alpha_{db} \right)^2 \right\},$$

where $\tilde{\lambda} > 0$ is the overall roughness penalty parameter.

Alternatively, suppose a single (joint) penalization parameter imposes too severe shrinkage in the extent to which the extremal dependence varies. In this case, one may allow each channel to have its own roughness penalty parameter, $\tilde{\lambda}_d$, performing $D - 1$ model fits in parallel.

2.4. *Extremal spectral analysis*. EEGs are zero-mean signals that can be expressed as a mixture of frequencies oscillating at the following standard frequency bands: Delta-band (Ω_1: 1–4 Hz), Theta-band (Ω_2: 4–8 Hz), Alpha-band (Ω_3: 8–12 Hz), Beta-band (Ω_4: 12–30 Hz), and Gamma-band (Ω_5: 30–50 Hz), assuming a sampling rate of 100 Hz (Nunez and Srinivasan (2007), Ombao et al. (2016)). The spectral decomposition of channel T3 is displayed in Figure 2 for both the pre- and postseizure onset phases.

Following Gao et al. (2020), Ombao and Van Bellegem (2008), and Ombao and Pinto (2021), two EEG signals, $Y_1(t)$ and $Y_2(t)$, may be decomposed into the five rhythms as $Y_1(t) \approx \sum_{k=1}^{5} Y_{1,\Omega_k}(t)$ and $Y_2(t) \approx \sum_{k=1}^{5} Y_{2,\Omega_k}(t)$, where $Y_{d,\Omega_k}(t)$ denotes the Ω_k-waveform in channel $Y_d(t)$. In practice, these rhythms are obtained by applying a linear filter, such as

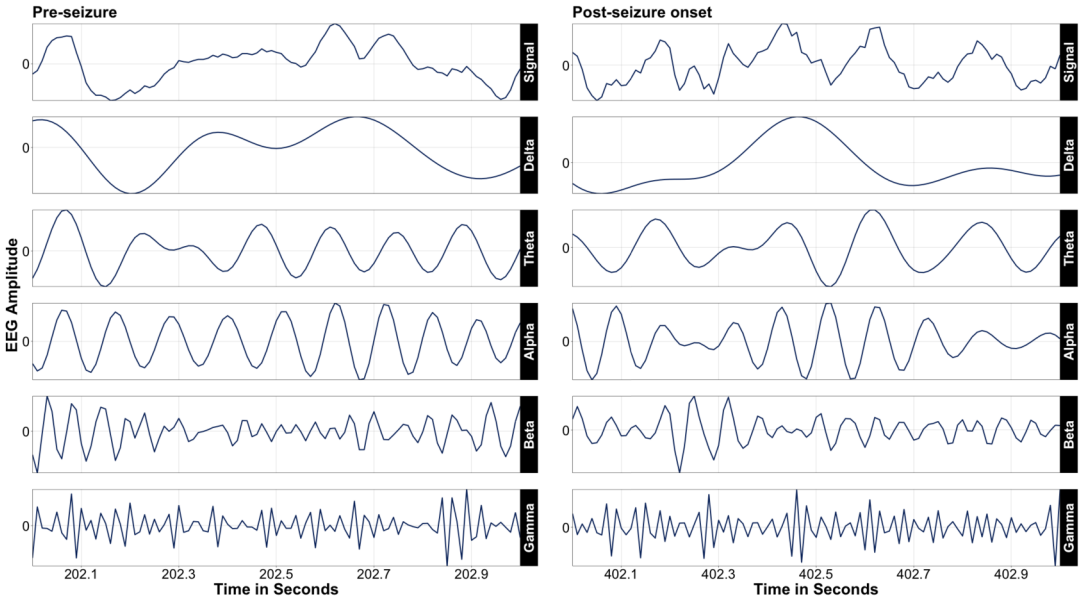

FIG. 2. *Spectral decomposition of channel T3 signal during one second of both pre- and postseizure onset phases from the EEG recording of a patient diagnosed with left temporal lobe epilepsy. The frequency bands are*: *Delta-band* (Ω_1: 1–4 *Hz*), *Theta-band* (Ω_2: 4–8 *Hz*), *Alpha-band* (Ω_3: 8–12 *Hz*), *Beta-band* (Ω_4: 12–30 *Hz*), *and Gamma-band* (Ω_5: 30–50 *Hz*).

the Butterworth filter (see Cohen (2013)). One measure of dependence between two channels is coherence, which is essentially the frequency band-specific squared correlation between a pair of rhythms, that is,

$$\text{Coherence}_{\Omega_k}(t) = \max_{\ell} \left|\text{cor}\big(Y_{1,\Omega_k}(t), Y_{2,\Omega_k}(t+\ell)\big)\right|^2.$$

Another major contribution of this paper is a novel measure of nonlinear cross-dependence to characterize the joint behavior of the oscillations in brain signals, that is, the joint interpretation of the H&T parameter estimates, obtained from extremal models defined from both time and frequency domain perspectives. These parameter estimates reveal the impact of unusually large frequency band amplitudes in the T3 channel (the reference channel projecting from the seizure foci on the temporal lobe) on the other channels. In order to study the effect of the extremal behavior in the Ω_k-waveform, $k = 1, \ldots, 5$, of the associated channel d, we study the conditional distribution of $|Y_{d,\Omega_k}(t)|$, given $|Y_{1,\Omega_i}(t)| > \tau$, separately for each $i = 1, \ldots, 5$; remember that Y_d represents the associated channel d while Y_1 is the reference channel T3. Our pioneering way of analyzing the spectral decomposition of EEG signals produces new interesting results that give us deeper insights into the highly nonlinear interactions and dependence between channels during an epileptic seizure event.

2.5. *Cross-validation, bootstrapping, uncertainty and diagnostics.* This section gives details on both the block bootstrap and walk-forward cross-validation procedures, designed to assess the uncertainty of estimated parameters and to select the optimal roughness penalty parameters respectively, while handling the autocorrelation of the time series appropriately.

In both the marginal and the H&T models, we need to select the optimal value of the roughness penalty parameter in an objective manner. Note that we have one λ_d for each margin, $d = 1, \ldots, D$, and a single $\tilde{\lambda} > 0$ for the extremal dependence model. In both cases, for each channel, and within each time block, we rely on a 10-fold walk-forward cross-validation procedure (Hyndman and Athanasopoulos (2019)) since the data are autocorrelated across time.

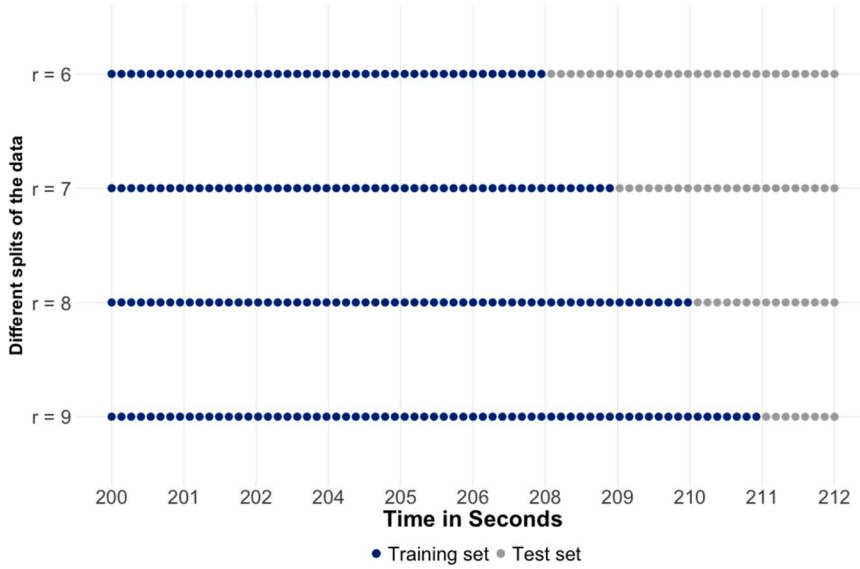

FIG. 3. *Schematic for a* 10*-fold walk-forward cross-validation procedure for the autocorrelated data within time block* 1 *of the preseizure phase of the EEG application. The darker dots represent the training sets, while the lighter ones are the test sets.*

For this procedure we divide each time block b into $k = 10$ adjacent nonoverlapping folds. Then, we consider the first r folds as the training set and the last $k - r$ folds as the test set. For a grid of λ values and for each split configuration, we then fit the model to the training set. We use the parameter estimates to obtain the value of the log-likelihood function on the test set, and we finally add up these log-likelihood values of each split configuration to get an overall score. The value of λ that maximizes the overall score is taken as the optimal value for the penalization parameter. Figure 3 shows a schematic illustration of this cross-validation procedure.

To assess the uncertainty of parameter estimates, we employ a block bootstrap procedure which preserves local temporal dependence in the data. As described in Lahiri (2003) and because we assume time blocks to be quasi weakly stationary, we resample entire blocks of consecutive observations (rather than single observations) keeping the correspondence between the conditioning variable Y_1 and the associated variables $Y_2, \ldots, Y_D$, and then refit the marginal and H&T dependence models. From the bootstrap samples, we compute confidence intervals for the parameters that evolve with time and histograms for those that are constant over time. Notice that the bootstrap blocks differ from the B time blocks used in the penalized piecewise model formulation. For more details on the performance of the block bootstrap and other types of bootstrap schemes for data with complex dependence, see Davison and Hinkley (1997).

Specifically, we split each of these B time blocks into C nonoverlapping (sub)blocks, of size s, which are resampled with replacement to generate M bootstrap samples. The block size s is determined with the aid of the sample (partial) autocorrelation function and extremograms (i.e., the extreme-value analog of the autocorrelation function) (Davis and Mikosch (2009), Davis, Mikosch and Cribben (2012)) to make sure it is large enough to keep the overall temporal dependence structure; see the Supplementary Material (Guerrero, Huser and Ombao (2023a)) for details.

Simultaneously, for each bootstrap sample, the nonexceedance probability for both marginal and H&T models is also randomly sampled from a uniform distribution given a range of reasonable values to account for thresholding uncertainty: $\tau_d \in \mathcal{T}_d \subset (0, 1)$, $d \in \{1, \ldots, D\}$,

and $\tilde{\tau} \in \tilde{\mathcal{T}} =\subset (0, 1)$. In this way, one may produce diagnostic plots, such as parameter stability plots, to select the best threshold for each channel. Also, from the bootstrap samples, quantile-quantile (Q-Q) plots and residual plots can be drawn to evaluate the goodness-of-fit. For the original fit, that is, not the bootstrapped ones, the medians of the nonexceedance probability intervals are used.

3. Extremal spectral analysis of seizure EEG data.

3.1. *Application of the Conex–Connect model.* We examine the changes in connectivity (or dependence between channels) using the proposed Conex–Connect method. In this paper the scalp EEG recording is from a female patient diagnosed with left temporal lobe epilepsy. The EEG data are recorded from a 10–20 system, and $D = 19$ channels are available for analysis. Figure 1 shows the electrode placement and EEG traces for some channels. In this specific case the physician knew in advance that the seizure onset would be on the left temporal lobe region which would be captured by the recording at the T3 channel ($d = 1$). Thus, T3 is set to be the reference channel (i.e., the conditioning variable). The remaining 18 channels are treated as the associated variables ($d = 2, \ldots, 19$). The record has a duration of 500 seconds collected at the sampling rate of 100 Hz; thus, there are 50,000 time points per channel. Furthermore, remember that in our analysis, we consistently use absolute values of the EEG signals or their (spectrally decomposed) waveforms.

The first stage of the Conex–Connect method is to model the marginal extremes for each channel d independently. Each channel d is split into pre- and postseizure onset phases at time $t = 350$ seconds, and thus, $T = 15{,}000$ points for each of the postseizure onset and preseizure phase. Note that we analyze the pre- and postseizure phases independently because of the sharp, nonsmooth transition around the seizure onset. Moreover, we split the data at this specific time point based on the neurologist's analysis of the EEG record which is corroborated by change point detection methods previously applied to the same dataset (Ombao, von Sachs and Guo (2005), Schröder and Ombao (2019)). Each phase is segmented into $B = 12$ time blocks of 1250 data points each. With this setup, the GP distribution is fitted jointly for all time blocks, specifying block-specific scale parameters and a constant shape parameter, as indicated in Section 2.2, and the fitted model is then used to transform the data to the standard Laplace scale. Moreover, to assess estimation uncertainty, $M = 500$ block bootstrap samples are generated within each time block $b = 1, \ldots, B = 12$, with bootstrap block size $s = 25$ data points. For all channels $d = 1, \ldots, D$, a common nonexceedance probability interval, $\mathcal{T}_d = (0.90, 0.95)$, is used. We chose this common interval because it is simpler than choosing d different intervals, while it guarantees a good model fit for all marginals. The results and diagnostics of this marginal estimation stage are provided in the Supplementary Material (Guerrero, Huser and Ombao (2023a)).

The second stage of the Conex–Connect method is to model the conditional extremal dependence of brain connectivity. The strength of the relationship between the associated variables and large values of the conditioning variable, T3, is estimated. After transforming the data from all channels into the standard Laplace scale, we choose the nonexceedance probability interval for the reference channel to be $\tilde{\mathcal{T}} = (0.88, 0.92)$, both for the pre- and postseizure onset phases. Here, we use a smaller threshold to ensure enough data points for all associated channels while conditioning on the reference channel. Estimation, uncertainty assessment, and goodness-of-fit details are thoroughly reported in the Supplementary Material (Guerrero, Huser and Ombao (2023a)). For the sake of brevity, we here only present and discuss the results from the associated frontal channels, namely, the left frontal F7 and right frontal F8 (see Section 3.2), while dashboards for the other channels can be found in the Supplementary Material (Guerrero, Huser and Ombao (2023a)) for completeness.

It is noteworthy that previous works analyzing this dataset (Ombao, von Sachs and Guo (2005), Ombao et al. (2001), Schröder and Ombao (2019)) concentrate on signal representation, spectral estimation, change-point detection, and dimensionality reduction, identifying the most relevant channels during the epileptic event. By contrast, the current EEG analysis, which is an application of the Conex–Connect method, focuses on modeling the entire brain network from an extreme-value perspective, contrasting it with widespread nonextreme-value models in the EEG literature (e.g., Acharya et al. (2013) and Bowyer (2016)).

Regarding computational cost, approximately 40 minutes were needed to complete a full analysis for all channels, pre- and postseizure phases, using an iMac with four cores (Intel i5, 3 GHz) and 16 GB of memory.

3.2. *Results and discussion*. Figure 4 displays a dashboard with the results of the Conex–Connect method for channel F7, given high values of the reference channel T3, both for pre- and postseizure onset phases. Panel I provides the individual behavior of the parameter estimates $\{\hat{\alpha}_{2,b}\}_{b=1}^{B}$, $\hat{\beta}_2$, $\hat{\mu}_2$, and $\hat{\sigma}_2$ *in the time-domain*. Here, the subscript $d = 2$ refers to channel F7, and $b \in \{1, \ldots, 24\}$ denotes the time block index. The 95% confidence bands for $\hat{\alpha}_{2,b}$ and the histograms for $\hat{\beta}_2$, $\hat{\mu}_2$, and $\hat{\sigma}_2$ are drawn out from the block bootstrap procedure; see Section 2.5. In the preseizure phase the estimates of the first-order dependence parameter, $\hat{\alpha}_{2,b}$, $b = 1, \ldots, 12$, are very high and stable (close to 1). This indicates a strong positive linear relationship between F7 and large values of T3. This result from the Conex–Connect method makes sense because both channels F7 and T3 are located on the same side of the scalp topography (left hemisphere). Moreover, since T3 (more than any other channel) captures the abnormal behavior on the seizure foci, which then spreads to neighboring channels, this result from the Conex–Connect analysis confirms this extremal dependence between the T3 and F7 channels, that is, abnormally large fluctuations in the left temporal T3 are also accompanied by abnormally large fluctuations in nearby channels, such as F7. Note that the confidence bands are narrow, suggesting a high level of certainty concerning this dependence structure. The histogram of $\hat{\beta}_2$ for the preseizure phase is centered around 0.4 with practically no values close to 0, indicating asymptotic independence yet with strong subasymptotic dependence (i.e., at finite levels). In the postseizure onset moment, we can see a sudden change immediately after the seizure, with $\hat{\alpha}_{2,13}$ being much smaller than $\hat{\alpha}_{2,12}$, indicating weaker extremal dependence. Furthermore, this reduction in the extremal dependence is observed throughout the entire postseizure phase, indicating that after the seizure onset, abnormally large fluctuations of T3 tend to be unaccompanied by abnormally large fluctuations of F7. Also, the behavior of $\hat{\alpha}_{2,b}$, $b = 13, \ldots, 24$, are much more erratic with wider confidence bands, suggesting a higher level of uncertainty after the seizure onset. This is interesting because it reveals the more chaotic nature of the seizure process. For some blocks the estimates are close to 0.2. Thus, there is a weakening in the linear relationship between F7 and large values of T3, though it still remains positive. In terms of β_2, there is a shift to the left in the histogram with a high concentration around 0, pointing to a possible change toward independence. In the Supplementary Material (Guerrero, Huser and Ombao (2023a)), we present similar dashboards for the tail correlation coefficient, χ, where some of its limitations are evident. In the case of this data analysis, χ dashboards present wider confidence intervals, most of the time containing zero. On the other hand, the first-order dependence parameter of Conex–Connect can capture the underlying behavior of χ throughout the time blocks but with tighter confidence intervals that do not contain zero. Also, jointly with the second-order dependence parameter, they can easily distinguish between asymptotic independence and dependence.

By contrasting the findings above to Panel I of Figure 5, it is clear that a different story prevails on the right side of the brain. For the right frontal channel F8 ($d = 3$), given high values

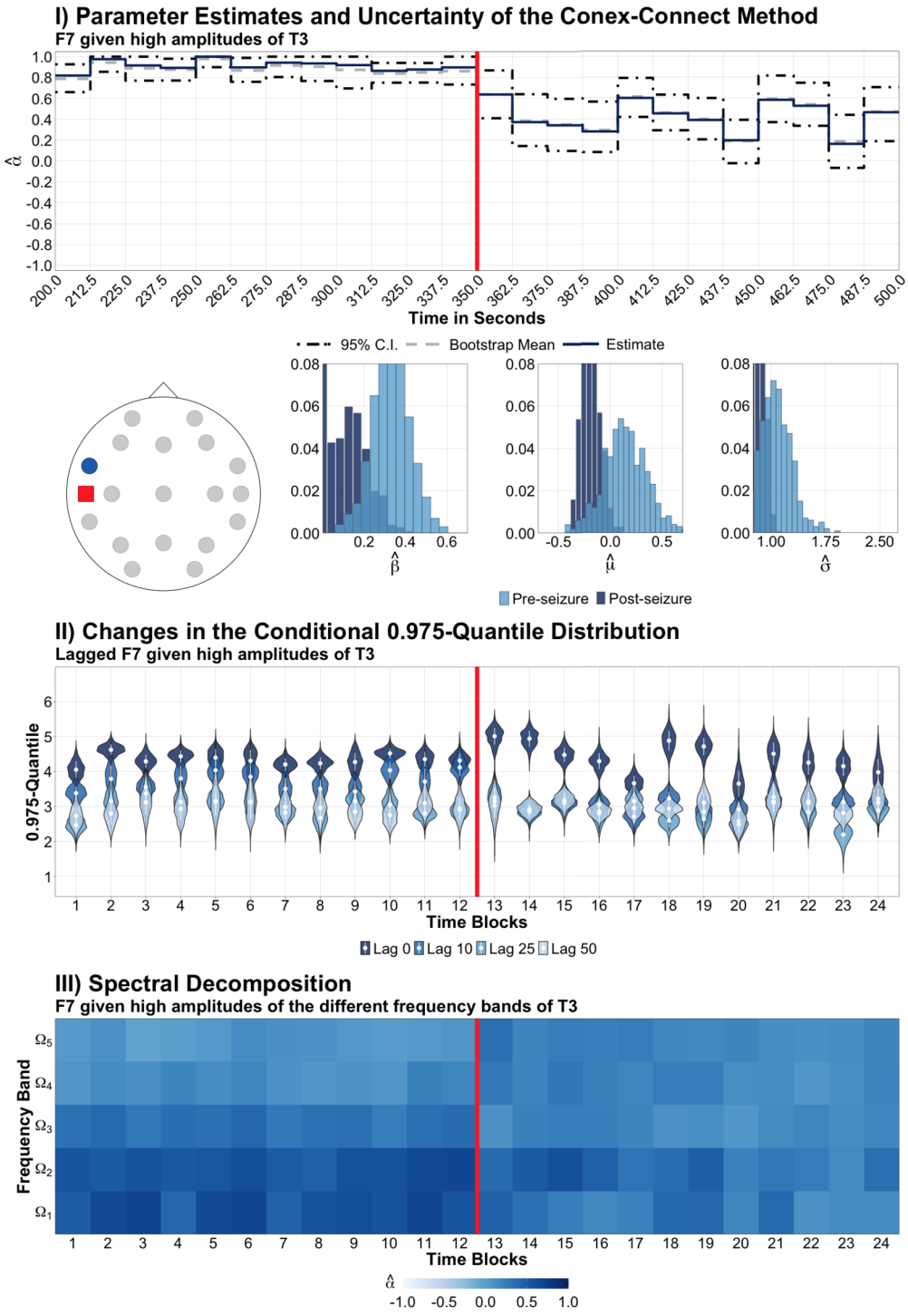

FIG. 4. *A dashboard with results of the Conex–Connect method for pre- and postseizure onset phases. Channel F7 (darker circle) given high values of T3 (square) are highlighted in the EEG scalp cartoon.* Panel I) *In the first line, the evolution of the estimated first-order dependence parameter* α_{db} *(solid line) through time with its bootstrap mean (dashed gray line) and 95% confidence bands (dash-dotted lines). In the second line, the histograms of the bootstrap estimates for scale exponent parameter* β_d *and for residual mean* μ_d *and scale* σ_d. Panel II) *Bootstrap violin plots for the* 0.975*-quantile of the conditional distribution of F7, given that T3 reaches its own* 0.975*-quantile. Different colors represent estimates of the conditional quantile for different lag values of the associate channel F7.* Panel III) *Effect of the different frequency bands* (Ω_1: 1–4, Ω_2: 4–8, Ω_3: 8–12, Ω_4: 12–30, *and* Ω_5: 30–50 *Hz) on the first-order dependence parameter. Darker pixels indicate higher dependence.*

of the reference channel T3, it appears that seizure does not affect the extremal dependence structure as much, since there is practically no discernible change in the estimates $\hat{\alpha}_{3,b}$ before and after the seizure. Also, in the preseizure phase, the linear relationship between F8 and large values of T3 is weaker when compared to F7. This is partly because F7 is closer to the seizure focus location which is around the reference channel T3. In addition, the histograms for $\hat{\beta}_3$, $\hat{\mu}_3$, and $\hat{\sigma}_3$, both for pre- and postseizure onset phases, are relatively stable, indicating the absence of changes in the second-order extremal dependence characteristics.

Back to Figure 4, Panel II displays, *in the time-domain*, the time evolution of the conditional distribution of the conditional 0.975-quantile in (3), given that T3 reaches its own 0.975-quantile, estimated semiparametrically, as explained in Section 2.3. In this plot we can jointly analyze the model parameters. Also, with the different violin plots, we investigate how the effect of large values of T3 at time t impacts channel F7 at time $t + \ell$, for time lags $\ell = 0$, 0.10, 0.25, and 0.50 seconds. The Conex–Connect method shows that the extremal dependence is stronger at lag $\ell = 0$ and weakens for other lags, indicating a stronger contemporaneous extremal dependence than lagged extremal dependence. Moreover, when contrasting lag $\ell = 0$ to the other lags, the discrepancy between the violins, both in terms of medians and shapes, is more pronounced after the seizure onset. This suggests that seizure has an impact on the conditional extremal dependence of the brain network. The same panel of Figure 5, in the case of channel F8, the opposite side from the seizure focus, shows that the extremal dependence structure at lag $\ell = 0$ is indistinguishable from those of higher lags, and at a lower level overall than for channel F7. This denotes less synchrony (in the extremal dependence) between the channel corresponding to the seizure foci (reference channel) and the channel on the contra-lateral side of the foci.

The Conex–Connect method produces interesting results regarding how the extreme values of the different oscillations of the reference channel T3 impact brain connectivity. Henceforth, working *in the frequency domain*, Figure 4, Panel III, shows the estimated first-order dependence coefficient $\hat{\alpha}_{2,b}$ $(b = 1, \ldots, B)$ between the (absolute) Ω_k-waveform in channel F7 given high (absolute) amplitudes of the same waveform in channel T3. Values of $\hat{\alpha}_{2,b}$ closer to 1 are darker. In the preseizure phase the extreme values in the high-frequency Gamma-band exhibit the lowest level of extremal dependence. This seems to be consistent across the entire preseizure phase. However, the dependence pattern changes in the postseizure phase. First, the extremes of *Gamma-band* from T3 shows the highest level of extremal dependence with the Gamma-band from F7. This is quite interesting because sudden outbursts of high-frequency oscillations typically characterize seizure onset, as shown by Medvedev, Murro and Meador (2011). Moreover, since the postseizure onset is typically nonstationary, we see that this dependence structure also evolves over time blocks. In the right side of the brain, Figure 5, Panel III shows that prior to seizure onset, the higher values of the low-frequency Delta-band lead the changes in the extremal dependence structure. In the postseizure phase we notice that, immediately after the seizure onset, the high-frequency Gamma-band becomes more prominent, similar to the left side of the brain (channel F7), the same side as T3.

Figure 6 presents a dashboard with the results of classical methods, based on cross-correlation and cross-coherence for comparison. The left column shows the results for F7, while the right column shows the results for F8. In all panels we display both pre- and post-seizure onset phases. Panel I displays the evolution of the cross-correlation over time blocks *in the time-domain*. Panel II shows how the cross-correlation between T3 at time t evolves when computed for future values of channel F7 and F8 at time $t + \ell$, for time lags $\ell = 0$, 0.10, 0.25, and 0.50 seconds; here, again, *in the time-domain*. Finally, Panel III, now *in the frequency-domain*, exhibits the impact of high values of the different frequency bands of T3 in its cross-coherence with F7 and F8.

To contrast the results of our method Conex–Connect, based on extreme-value theory with classical results, we notice that the first-order dependence parameters of Conex–Connect have

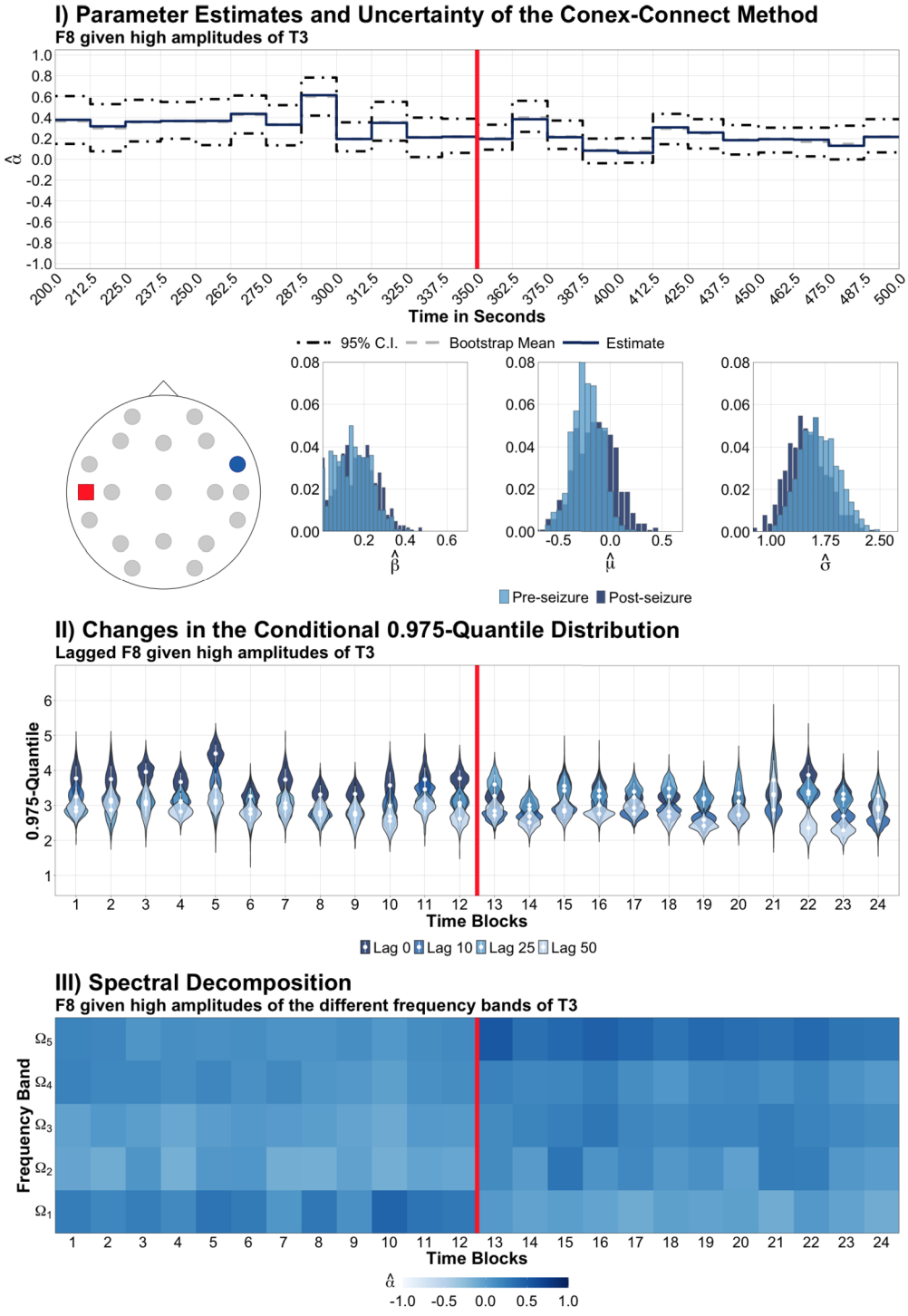

FIG. 5. *A dashboard with results of the Conex–Connect method for pre- and postseizure onset phases. Channel F*8 (*darker circle*) *given high values of T*3 (*square*) *are highlighted in the EEG scalp cartoon.* Panel I) *In the first line, the evolution of the estimated first-order dependence parameter* α_{db} (*solid line*) *through time with its bootstrap mean* (*dashed gray line*) *and* 95% *confidence bands* (*dash-dotted lines*). *In the second line, the histograms of the bootstrap estimates for scale exponent parameter* β_d *and for residual mean* μ_d *and scale* σ_d. Panel II) *Bootstrap violin plots for the* 0.975*-quantile of the conditional distribution of F*8 *given that T*3 *reaches its own* 0.975*-quantile. Different colors represent estimates of the conditional quantile for different lag values of the associate channel F*8. Panel III) *Effect of the different frequency bands* (Ω_1: 1–4, Ω_2: 4–8, Ω_3: 8–12, Ω_4: 12–30, *and* Ω_5: 30–50 *Hz*) *on the first-order dependence parameter. Darker pixels indicate higher dependence.*

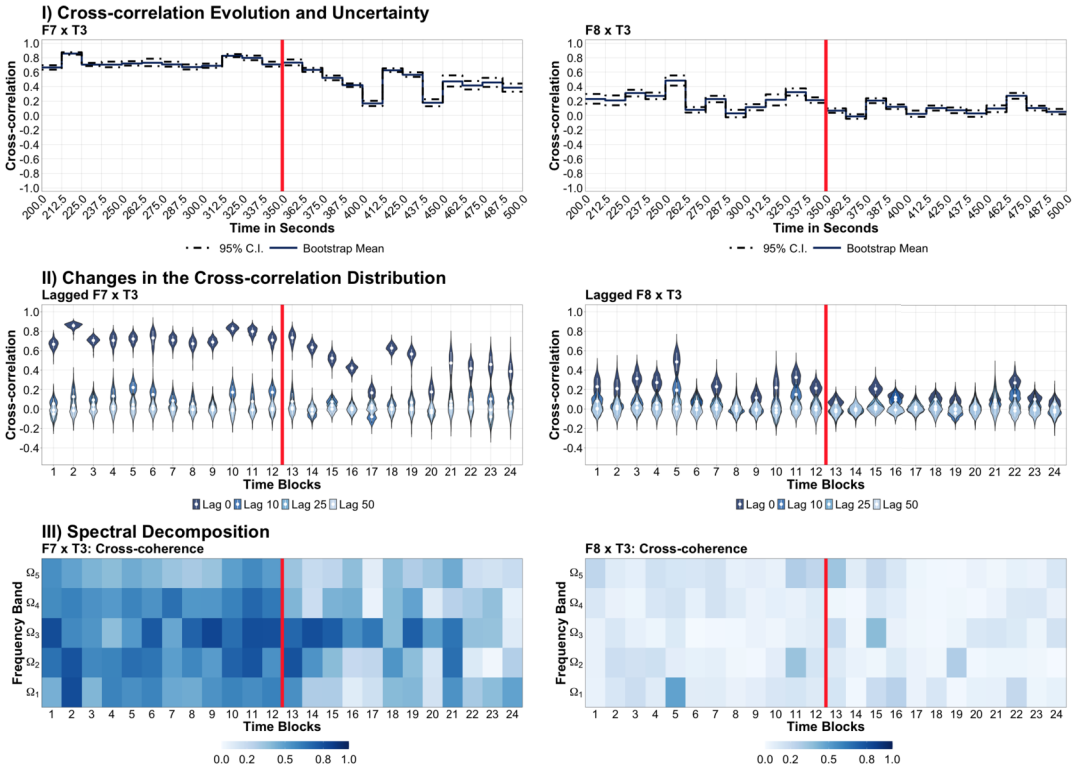

FIG. 6. *A dashboard with results of the classical analysis for pre- and postseizure onset phases. Left column displays results for channel F7 (brain left size) while the right column show the results for F8 (brain right side).* Panel I) *Evolution of the cross-correlation estimates (solid line), along with its confidence bands (dash-dotted lines), between channels F7 and T3 (reference channel) and between channels F8 and T3.* Panel II) *Violin plots to access the changes in the distribution of the cross-correlation both over time and for different lag values of the associated channels.* Panel III) *Effect of the different frequency bands on the cross-coherence. Darker pixels indicates higher values of cross-coherence. For all panels, uncertainty is obtained trough 500 bootstrap samples.*

a similar overall temporal pattern as that of the classical cross-correlation, both for F7 and F8, pre- and postseizure, as can be seen when comparing Figures 4, 5, and 6. In addition, comparing the changes in the conditional 0.975-quantile with the changes in the cross-correlation (Panel II, Figure 6), a similar pattern is observed across time blocks both for pre- and postseizure onset phases. However, regarding correlation, we notice a more striking discrepancy between contemporaneous dependence (lag $\ell = 0$) against other higher lags. These similarities suggest either (a.) that the conditional extremal dependence dominates the global dependence (as measured by correlation) or (b.) that the phenomenon that we see in the joint tail is similar to other less extreme quantiles of the distribution. Regarding the spectral decomposition, our method shows that, for both sides of the brain, immediately after the seizure, the high-frequency Gamma-band becomes the most relevant frequency in explaining the conditional extremal dependence. This finding does not agree with the results in terms of the classical coherence; see Figure 6, Panel III.

We further investigate the conditional extremal dependence in terms of frequency oscillations by decomposing the associated channels in their canonical frequency bands. We refit the model for all pairs of Ω_k-waveforms. Figure 7 displays the impact of extreme values of the different frequency bands of the reference channel T3 on the different frequency bands of the associated channels F7 and F8. Note that the heatmaps in Panel III of Figures 4 and 5 correspond to the diagonals of Figure 7. Here, beyond the previous findings in terms of seizure phases and sides of the brain, we notice that the medium-frequency bands, Beta and Alpha,

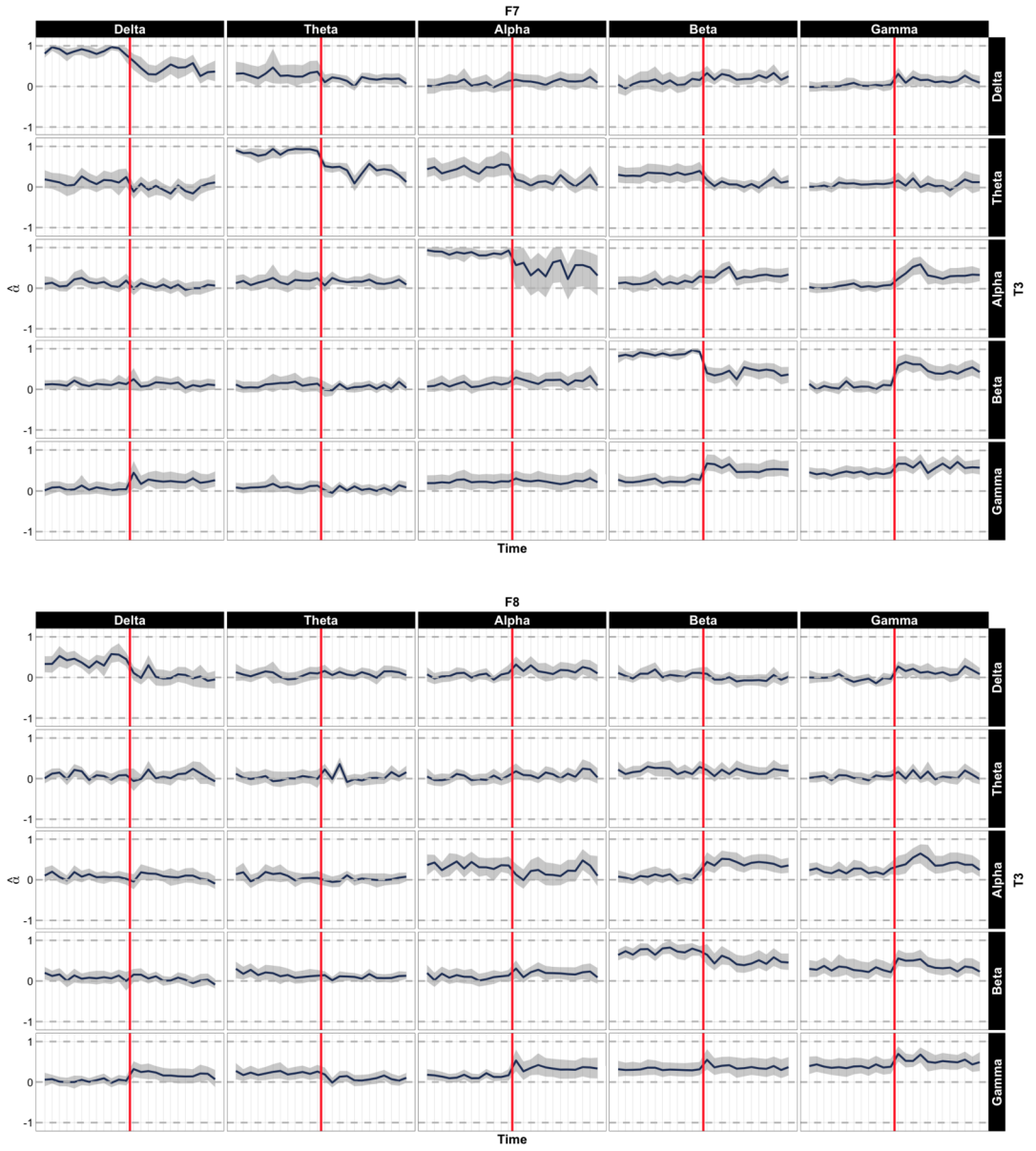

FIG. 7. *Effect of the different frequency bands of the reference channel T3 on the different frequency bands of the associated channels, F7 and F8, pre- and postseizure onset phases. The solid lines represents the evolution of the estimated first-order dependence parameter α through time with its bootstrap 95% confidence bands (shaded area). The vertical solid line is the seizure onset. The frequency bands are: Delta-band* (Ω_1: 1–4 *Hz*), *Theta-band* (Ω_2: 4–8 *Hz*), *Alpha-band* (Ω_3: 8–12 *Hz*), *Beta-band* (Ω_4: 12–30 *Hz*), *and Gamma-band* (Ω_5: 30–50 *Hz*).

of the reference channel T3, impact the first-order dependence parameter estimates, mainly on the Gamma-band of the associated channels. This may indicate that the Gamma-band can be used for feature engineering to improve the performance of machine learning algorithms for epilepsy detection.

Finally, Figure 8 presents the results of a first attempt to analyze how the extreme signal amplitudes triggered in channel T3 propagate through the brain (i.e., in space and time) after the seizure onset. These results are obtained using the following procedure. As usual, for the first time block we condition on T3 and identify the channel with the highest α estimate; in this case, F7. Then, for the second block, we refit the model, conditioning on F7, and we

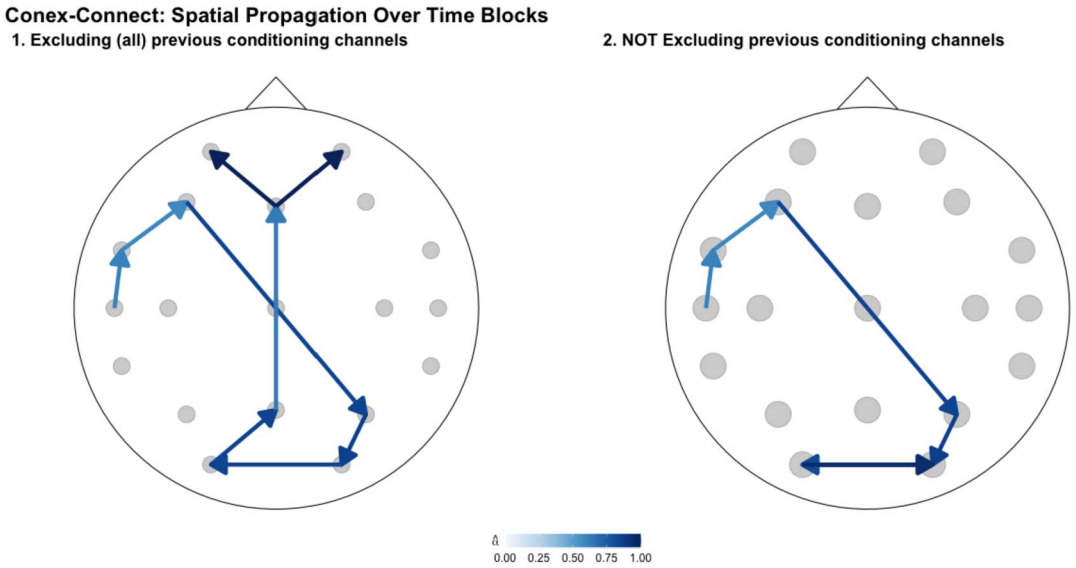

FIG. 8. *Results of a first attempt to study the "spatial propagation" of extreme signal amplitudes over different time blocks after the seizure onset, based on the conditional extremes model. The darker the arrow, the stronger the conditional extremal dependence captured by the first-order dependence parameter of the Conex–Connect method. For this specific analysis, we considered eight time blocks.*

identify the channel with highest $\hat{\alpha}$ and so forth. In this way we can draw a possible path of channels that may trigger further successive extreme values within the brain network. In Panel 1 of Figure 8, we refit the model excluding all previously identified conditioning channels, while, in Panel 2, we always use all available 19 channels. In both panels we can see that, just after the seizure onset, the conditional extremal dependence is stronger in neighboring areas of T3 on the left side of the brain and that it then jumps to the occipital region, where stronger dependence prevails. Also, when not excluding the previous conditioning channels, the path ends in channels O1 and O2. In the Supplementary Material (Guerrero, Huser and Ombao (2023a)), we present the same analyses but within each time block.

4. Conclusion. This paper presents the novel Conex–Connect method, the first extreme-value model-based approach for learning patterns in the extremal dependence during periods of high volatility in brain signals. The method extends the conditional multivariate extremes model (Heffernan and Tawn (2004), Ross et al. (2018)) to capture time-varying extremal dependence features, both from time-domain and frequency-domain perspectives, while adequately assessing estimation uncertainty using a block bootstrap procedure accounting for the autocorrelation in the data. We here give a full characterization of the conditional extremal dependence of brain connectivity, shedding light on how the brain network responds to an epileptic seizure event. We also study the extremal brain connectivity based on the spectral decomposition of the different channels. To the best of our knowledge, our method is a pioneer in linking the association between extreme values of frequency oscillations in a reference channel with oscillations in other channels of the brain network, and we foresee further research in this direction because of the mathematically elegant and natural use of extreme-value theory in this context.

Essentially, our methodology relies on "dissecting" the original brain signal into various components at different frequency bands and studying the extremal dependence structure of each frequency component separately, based on distinct conditional extreme-value models. While our statistical analysis is motivated by a very concrete applied problem and our

empirical results reveal salient features of brain connectivity during an epileptic seizure, it would also be interesting in future research to further investigate the theoretical link between the (conditional) extremal dependence structure of each frequency component to that of the overall signal. Specifically, we expect that the frequency component with the strongest tail dependence would dominate the overall tail dependence structure and dictate the occurrence of joint extremes in the original time series. However, this conjecture still needs to be rigorously validated and mathematically formalized. Conversely, the overall extremal dependence structure imposes certain constraints on the joint behavior of each frequency component, and it would be interesting to study this in more detail.

Our proposed approach demonstrates changes in the conditional extremal dependence of brain connectivity between pre- and postseizure onset phases. In general, before the seizure, the dependence is notably stable for all channels (conditioning on extreme values of the T3). On the other hand, during the postseizure phase (a period of very high volatility at T3), the dependence between channels is weaker. In general, the strength of dependence decreases at large lagged values of the associated channels when T3 is kept fixed. Also, after the seizure, the high values of the high-frequency Gamma-band are the most relevant features to explain the conditional extremal dependence of brain connectivity. Furthermore, our approach can be easily generalized to a multivariate setup where one may study the dependence between pairs (or groups) of relevant channels, conditional on the extremal behavior of the reference channel. We plan to investigate this aspect in future research.

The current state of the Conex–Connect method does not incorporate the underlying spatial structure of brain activity; in Figure 8 we present a first attempt to estimate a possible path of channels that may trigger successive extreme signal amplitudes throughout the brain network, based on the proposed (bivariate) conditional extremes model, but it would be interesting in future research to conduct a proper space–time (potentially causal) analysis combining all channels in a single joint model. This is a complex topic because the usual Euclidean distance is not adequate to characterize spatial dependence between brain signals at different channels. Hence, in a future research goal we aim to develop extreme-value-based models to deal not only with the time-varying features of brain connectivity but also with its spatiotemporal characteristics. A possibility would be to adapt the multivariate version of the H&T model, or the conditional spatial extremes model of Wadsworth and Tawn (2019), or a combination thereof, to the time-varying framework of brain signals, combined with an appropriate distance metric.

For reproducibility purposes and to boost the application of our proposed methodology in other contexts, we provide a minimal reproducible version of our `R` code as well as a working example on a subset of the EEG data from three different channels as Supplementary Material (Guerrero, Huser and Ombao (2023b)) of this paper. The code can also be accessed from the first author's GitHub account, that is, https://github.com/matheusguerrero/conex-connect.

Acknowledgments. The authors would like to thank the Editor, Associate Editor, and two referees for valuable suggestions that have improved the manuscript. The EEG dataset was collected by Dr. Beth Malow, attending neurologist from the University of Michigan Hospital, and made available by the authors of the paper Ombao et al. (2001).

Funding. This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Awards No. OSR-CRG2017-3434 and No. OSR-CRG2020-4394.

SUPPLEMENTARY MATERIAL

Supplement to "Conex–Connect: Learning patterns in extremal brain connectivity from multi-channel EEG data" (DOI: 10.1214/22-AOAS1621SUPPA; .pdf). We provide further details for each stage of the Conex–Connect method to model the conditional extremal dependence on brain connectivity. In this supplement, we discuss the model fitting and the goodness-of-fit of our method; we provide sensitivity analysis on the size of the temporal blocks and their independence; we study the extremal features of the data within each time block; and we also give extra details on the data application.

R **code for "Conex–Connect: Learning patterns in extremal brain connectivity from multi-channel EEG data"** (DOI: 10.1214/22-AOAS1621SUPPB; .zip). We provide the R code to apply the full Conex–Connect workflow to a subset of the EEG data from three different channels, namely, T3 (reference channel), F7 and F8.

REFERENCES

ACHARYA, U. R., VINITHA SREE, S., SWAPNA, G., MARTIS, R. J. and SURI, J. S. (2013). Automated EEG analysis of epilepsy: A review. *Knowl.-Based Syst.* **45** 147–165.

BOWYER, S. M. (2016). Coherence a measure of the brain networks: Past and present. *Neuropsychiatr. Electrophysiol.* **2** 1.

COHEN, M. X. (2014). *Analyzing Neural Time Series Data*. MIT Press, USA.

COLES, S. G., HEFFERNAN, J. and TAWN, J. A. (1999). Dependence measures for extreme value analyses. *Extremes* **2** 339–365.

DAS, B. and RESNICK, S. I. (2011). Conditioning on an extreme component: Model consistency with regular variation on cones. *Bernoulli* **17** 226–252. MR2797990 https://doi.org/10.3150/10-BEJ271

DAVIS, R. A. and MIKOSCH, T. (2009). The extremogram: A correlogram for extreme events. *Bernoulli* **15** 977–1009. MR2597580 https://doi.org/10.3150/09-BEJ213

DAVIS, R. A., MIKOSCH, T. and CRIBBEN, I. (2012). Towards estimating extremal serial dependence via the bootstrapped extremogram. *J. Econometrics* **170** 142–152. MR2955945 https://doi.org/10.1016/j.jeconom.2012.04.003

DAVISON, A. C. (2003). *Statistical Models*. *Cambridge Series in Statistical and Probabilistic Mathematics* **11**. Cambridge Univ. Press, Cambridge. MR1998913 https://doi.org/10.1017/CBO9780511815850

DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application*. *Cambridge Series in Statistical and Probabilistic Mathematics* **1**. Cambridge Univ. Press, Cambridge. MR1478673 https://doi.org/10.1017/CBO9780511802843

DAVISON, A. C. and HUSER, R. (2015). Statistics of extremes. *Annu. Rev. Stat. Appl.* **2** 203–235.

EWANS, K. and JONATHAN, P. (2014). Evaluating environmental joint extremes for the offshore industry using the conditional extremes model. *J. Mar. Syst.* **130** 124–130.

FIECAS, M. and OMBAO, H. (2016). Modeling the evolution of dynamic brain processes during an associative learning experiment. *J. Amer. Statist. Assoc.* **111** 1440–1453. MR3601700 https://doi.org/10.1080/01621459.2016.1165683

GAO, X., SHEN, W., SHAHBABA, B., FORTIN, N. J. and OMBAO, H. (2020). Evolutionary state-space model and its application to time-frequency analysis of local field potentials. *Statist. Sinica* **30** 1561–1582. MR4257545 https://doi.org/10.5705/ss.202017.0420

GUERRERO, M. B., HUSER, R. and OMBAO, H. (2023a). Supplement to "Conex–Connect: Learning patterns in extremal brain connectivity from multichannel EEG data." https://doi.org/10.1214/22-AOAS1621SUPPA

GUERRERO, M. B., HUSER, R. and OMBAO, H. (2023b). R code for "Conex–Connect: Learning patterns in extremal brain connectivity from multichannel EEG data." https://doi.org/10.1214/22-AOAS1621SUPPB

HEFFERNAN, J. E. and RESNICK, S. I. (2007). Limit laws for random vectors with an extreme component. *Ann. Appl. Probab.* **17** 537–571. MR2308335 https://doi.org/10.1214/105051606000000835

HEFFERNAN, J. E. and TAWN, J. A. (2004). A conditional approach for multivariate extreme values. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 497–546. MR2088289 https://doi.org/10.1111/j.1467-9868.2004.02050.x

HILAL, S., POON, S. H. and TAWN, J. (2011). Hedging the black swan: Conditional heteroskedasticity and tail dependence in S&P500 and VIX. *J. Bank. Financ.* **35** 2374–2387.

HYNDMAN, R. J. and ATHANASOPOULOS, G. (2019). *Forecasting*: *Principles and Practice*, 3rd ed. OTexts, Australia. Available at https://OTexts.com/fpp3. Accessed on 08/20/2020.

KEEF, C., PAPASTATHOPOULOS, I. and TAWN, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *J. Multivariate Anal.* **115** 396–404. MR3004566 https://doi.org/10.1016/j.jmva.2012.10.012

KRAFTY, R. T. and COLLINGE, W. O. (2013). Penalized multivariate Whittle likelihood for power spectrum estimation. *Biometrika* **100** 447–458. MR3068445 https://doi.org/10.1093/biomet/ass088

KRAFTY, R. T., HALL, M. and GUO, W. (2011). Functional mixed effects spectral analysis. *Biometrika* **98** 583–598. MR2836408 https://doi.org/10.1093/biomet/asr032

KRAFTY, R. T., ROSEN, O., STOFFER, D. S., BUYSSE, D. J. and HALL, M. H. (2017). Conditional spectral analysis of replicated multiple time series with application to nocturnal physiology. *J. Amer. Statist. Assoc.* **112** 1405–1416. MR3750864 https://doi.org/10.1080/01621459.2017.1281811

LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data. Springer Series in Statistics*. Springer, New York. MR2001447 https://doi.org/10.1007/978-1-4757-3803-2

LEDFORD, A. W. and TAWN, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83** 169–187. MR1399163 https://doi.org/10.1093/biomet/83.1.169

LUGRIN, T., DAVISON, A. C. and TAWN, J. A. (2019). Penultimate analysis of the conditional multivariate extremes tail model. Available at arXiv:1902.06972.

MEDVEDEV, A. V., MURRO, A. M. and MEADOR, K. J. (2011). Abnormal interictal gamma activity may manifest a seizure onset zone in temporal lobe epilepsy. *Int. J. Neural Syst.* **21** 103–114.

NUNEZ, P. L. and SRINIVASAN, R. (2007). Electroencephalogram. *Scholarpedia* **2** 1348.

OMBAO, H. and PINTO, M. (2021). Spectral dependence. Available at arXiv:2103.17240.

OMBAO, H. and VAN BELLEGEM, S. (2008). Evolutionary coherence of nonstationary signals. *IEEE Trans. Signal Process.* **56** 2259–2266. MR2516630 https://doi.org/10.1109/TSP.2007.914341

OMBAO, H., VON SACHS, R. and GUO, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Amer. Statist. Assoc.* **100** 519–531. MR2160556 https://doi.org/10.1198/016214504000001448

OMBAO, H. C., RAZ, J. A., VON SACHS, R. and MALOW, B. A. (2001). Automatic statistical analysis of bivariate nonstationary time series. *J. Amer. Statist. Assoc.* **96** 543–560. MR1946424 https://doi.org/10.1198/016214501753168244

OMBAO, H., LINDQUIST, M., THOMPSON, W. and ASTON, J. (2016). *Handbook of Neuroimaging Data Analysis*. CRC Press/CRC, USA.

OMBAO, H., FIECAS, M., TING, C.-M. and LOW, Y. F. (2018). Statistical models for brain signals with properties that evolve across trials. *NeuroImage* **180** 609–618. https://doi.org/10.1016/j.neuroimage.2017.11.061

ROSS, E., SAM, S., RANDELL, D., FELD, G. and JONATHAN, P. (2018). Estimating surge in extreme North Sea storms. *Ocean Engineering* **154** 430–444.

SCHEFFLER, A. W., DICKINSON, A., DISTEFANO, C., JESTE, S. and ŞENTÜRK, D. (2020). Covariate-adjusted hybrid principal components analysis. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems* (M. J. Lesot, S. Vieira, M. Z. Reformat, J. P. Carvalho, A. Wilbik, B. Bouchon-Meunier and R. R. Yager, eds.) Springer, Switzerland.

SCHRÖDER, A. L. and OMBAO, H. (2019). FreSpeD: Frequency-specific change-point detection in epileptic seizure multi-channel EEG data. *J. Amer. Statist. Assoc.* **114** 115–128. MR3941242 https://doi.org/10.1080/01621459.2018.1476238

TAWN, J., SHOOTER, R., TOWE, R. and LAMB, R. (2018). Modelling spatial extreme events with environmental applications. *Spat. Stat.* **28** 39–58. MR3887155 https://doi.org/10.1016/j.spasta.2018.04.007

WADSWORTH, J. L. and TAWN, J. (2019). Higher-dimensional spatial extremes via single-site conditioning. Available at arXiv:1912.06560.

WHO (2019). Epilepsy: A public health imperative—summary. Available at https://www.who.int/mental_health/neurology/epilepsy/report_2019/en/. Accessed on 05/20/2020.

ZHANG, Z. (2008). Quotient correlation: A simple based alternative to Pearson's correlation. *Ann. Statist.* **36** 1007–1030. MR2396823 https://doi.org/10.1214/009053607000000866

ZHANG, Z. (2021). On studying extreme values and systematic risks with nonlinear time series models and tail dependence measures. *Stat. Theory Relat. Fields* **5** 1–25. MR4232528 https://doi.org/10.1080/24754269.2020.1856590

ZHANG, Z., ZHANG, C. and CUI, Q. (2017). Random threshold driven tail dependence measures with application to precipitation data analysis. *Statist. Sinica* **27** 685–709. MR3674692