Research

CrossMark
click for updates

Author for correspondence:
Hau-tieng Wu
e-mail: hauwu@math.toronto.edu

THE ROYAL SOCIETY
PUBLISHING

ConceFT: concentration of frequency and time via a multitapered synchrosqueezed transform

Ingrid Daubechies[1], Yi (Grace) Wang[2] and Hau-tieng Wu[3]

[1]Department of Mathematics, Duke University, Box 90320, Durham, NC 27708-0320, USA
[2]Department of Mathematics, Syracuse University, 215 Carnegie Building, Syracuse, NY 13244-1150, USA
[3]Department of Mathematics, University of Toronto, 40 St George Street, Toronto, Ontario, Canada M5S 2E4

A new method is proposed to determine the time–frequency content of time-dependent signals consisting of multiple oscillatory components, with time-varying amplitudes and instantaneous frequencies. Numerical experiments as well as a theoretical analysis are presented to assess its effectiveness.

1. Introduction

Oscillatory signals occur in a wide range of fields, including geophysics, biology, medicine, finance and social dynamics. They often consist of several different oscillatory components, the nature, time-varying behaviour and interaction of which reflect properties of the underlying system. In general, we want to assess the number, strength and rate of oscillation of the different components constituting the signal, to separate noise from signal, and to isolate individual components; efficient and robust extraction of this information from an observed signal will help us better describe and quantify the underlying dynamics that govern the system. For each of the quantities of interest listed, we thus want an estimator that is consistent, that has (ideally) small variance and that produces results robust to different types of noise.

If the observed signal f can be written as a finite sum of the so-called harmonic components, i.e. $f(t) = \sum_\ell a_\ell \cos(2\pi \xi_\ell t + \delta_\ell)$, where $a_\ell > 0$ (respectively, $\xi_\ell > 0$) represents the strength or amplitude (respectively, frequency) of the ℓth component, then one can recover the a_ℓ and ξ_ℓ from time samples of $f(t)$ via the Fourier transform $\hat{f}$ of f, defined by $\hat{f}(\xi) := \int f(t)\, e^{-i2\pi \xi t} dt$. (If the ξ_ℓ are all integer multiples of a common $1/t_0$, then the integral can be taken over an interval of length t_0; when this is not the case, one can resort to integrals over long time intervals and average. Typically, only discrete samples $f(t_n)$, $n \in \mathbb{Z}$, are known, rather than the continuous-time course $f(t)$, $t \in \mathbb{R}$, and the integrals are estimated by quadrature.) However, oscillatory signals of interest often have more complex behaviour. We shall be interested in particular in signals that are still the combination of 'elementary' oscillations, but in which both the amplitude and the frequency of the components are no longer constant; they can be written as

$$f(t) = \sum_{k=1}^{K} A_k(t) \cos(2\pi \varphi_k(t)), \tag{1.1}$$

where $K \in \mathbb{N}$, $A_k(t) > 0$ and $\varphi_k'(t) > 0$ for all k, but $A_k(t)$ and $\varphi_k'(t)$ are not constants. One can compute the Fourier transform $\hat{f}$ of such signals, and recover f from $\hat{f}$ (this can be validly done for a much wider class of functions), but it is now less straightforward to determine the time-varying amplitudes $A_k(t)$ and the so-called 'instantaneous frequencies' $\varphi_k'(t)$ from $\hat{f}$. Although the time-local behaviour of the oscillations, and their deviation from perfect periodicity, cannot be captured by the Fourier transform in an easily 'readable' way, an accurate description of this instantaneous behaviour is nevertheless important in many applications, both to understand the system producing the signal and to predict its future behaviour. Examples in the medical field include studies of the circadian [1,2] and cortical rhythms [3], or of heart rate [4,5] and respiratory variability [6,7], all widely studied to understand physiology and predict clinical outcomes.

The last 50 years have seen many approaches, in applied harmonic analysis and signal processing, to develop useful analysis tools for signals of this type; this is the domain of *time–frequency (TF) analysis*. Several algorithms and associated theories have been developed and widely applied (see e.g. the overview [8]); well-known examples include the short-time Fourier transform (STFT), continuous wavelet transform (CWT) and Wigner–Ville distribution (WVD). The main idea is often to 'localize' a portion of the signal in time, and then 'measure' the oscillatory behaviour of this portion. For example, given a function $f \in L^2$, the *windowed transform* or STFT associated with a window function $h(t)$ can be defined as

$$V_f^{(h)}(t, \eta) := \int f(s) h(t - s)\, e^{-i2\pi \eta(t-s)} ds, \tag{1.2}$$

where $t \in \mathbb{R}$ is the time, $\eta \in \mathbb{R}^+$ is the frequency and h is the window function chosen by the user—a commonly used choice is the Gaussian function with kernel bandwidth $\sigma > 0$, i.e. $h(t) = (2\pi\sigma)^{-1/2}\, e^{-t^2/(2\sigma^2)}$. (The overall phase factor $e^{-i2\pi \eta t}$ is not always present in the STFT, leading to the name *modified short-time Fourier transform* (mSTFT) for this particular form in [9].)

Other, more specialized methods, targeting in particular signals of type (1.1), include the empirical mode decomposition [10], ensemble empirical mode decomposition [11], the sparsity approach [12], iterative convolution filtering [13,14], the approximation approach [15], non-local mean approach [16], time-varying autoregression and moving average approach [17] as well as the synchrosqueezing transforms (SSTs) introduced and studied by some of us [9,18–21].

All TF methods that target reasonably large classes of functions (as opposed to functions with such specific models that complete characterization requires only fitting a small number of parameters) must face the Heisenberg uncertainty principle, limiting how accurately oscillatory information can be captured over short time intervals; for toy signals specially designed to have precise TF properties (e.g. chirps), this typically expresses itself by a 'blurring' or 'smearing out' of their TF representation, regardless of the analysis tool used. *Reassignment methods* [22–24], introduced in 1978 and recently attracting more attention again, were proposed to analyse and possibly counter this. Their main idea is to analyse the local behaviour in the TF plane of

3

rsta.royalsocietypublishing.org *Phil. Trans. R. Soc. A* **374**: 20150193

portions of the representation, and determine nearby possible TF concentration candidates that best explain it; each small portion is then reallocated to its 'right' place in the TF plane, to obtain a more concentrated TF representation that, one hopes, gives a faithful and precise rendering of the TF properties of the signal. Reassignment methods can be applied to very general TF representations [8,23]; they can be adaptive as well [25]. It has been argued recently [16] that reassignment methods can be viewed as analogues of 'non-local means' techniques commonly applied in image processing; this provides an intuitive explanation for their robustness to noise.

The SST can be viewed as a special reassignment method [23–25]. In SST, the STFT or CWT coefficients are reassigned *only* in the frequency 'direction' [19,21,26]; this preserves causality, making it possible to reconstruct each component with real-time implementation [27]. The STFT-based SST of f is defined as

$$S_f^{(h)}(t,\xi) := \lim_{\alpha \to 0} \int V_f^{(h)}(t,\eta) g_\alpha(\xi - \omega_f^{(h)}(t,\eta))\, \mathrm{d}\eta, \tag{1.3}$$

where g_α is an 'approximate δ-function' (i.e. g is smooth and has fast decay, with $\int g(x)\,\mathrm{d}x = 1$, so that $g_\alpha(t) := (1/\alpha)g(t/\alpha)$ tends weakly to the delta measure δ as $\alpha \to 0$), and with $\omega_f^{(h)}$ defined by

$$\omega_f^{(h)}(t,\eta) := \frac{-\mathrm{i}\partial_t V_f^{(h)}(t,\eta)}{2\pi V_f^{(h)}(t,\eta)} \quad \text{if} \quad V_f^{(h)}(t,\eta) \neq 0, \quad \text{and} \quad \omega_f^{(h)}(t,\eta) := -\infty \text{ otherwise.} \tag{1.4}$$

The SST for CWT is defined similarly; see [19,28], or §2. SST was proposed originally for sound signals [18,29]; its theoretical properties have been studied extensively [19,26–28,30–32], including its stability to different types of noise [28,33]. Several variations of the SST have been proposed [31,34–37]; in particular, the SST approach can also be used for other TF representations, such as the wave packet transform [35] and S-transform [38], and it can be extended to two-dimensional signals (such as images) [39,40]. In addition, its practical usefulness has been demonstrated in a wide range of fields, including medicine [5–7,41–44], mechanics [37,45,46], finance [47,48], geography [38,49–51], denoising [31], atomic physics [52–54] and image analysis [55,56].

The SST approach can extract the instantaneous frequency and reconstruct the constitutional oscillatory components of a signal of type (1.1) in the presence of noise [28,33]. However, its performance suffers when the signal-to-noise ratio (SNR) becomes low: as the noise level increases, and even before it completely obscures the main concentration in the TF plane of the signal, spurious concentration areas appear elsewhere in the TF plane, caused by correlations introduced by the overcomplete STFT or CWT analysis tool. The effect of these misleading perturbations, which downgrade the quality of the results, can be countered, to some extent, by *multitapering*.

Multitapering is a technique originally proposed to reduce the variance and hence stabilize power spectrum estimation in the spectral analysis of stationary signals [57–59]. Sampling the signal during only a finite interval leads to artefacts, traditionally reduced by tapering; an unfortunate side effect of tapering is to diminish the impact of samples at the extremes of the time interval. Thomson [57] showed that one can nevertheless exploit optimally the information provided by the samples at the extremities, by using several orthonormal functions as tapers: the average of the corresponding power spectra is a good estimator with reduced variance. This technique has since been applied widely [58,60–63]. Multitapering was later extended to non-stationary TF analysis by combining it with reassignment [64–66]: a more robust 'combined' reassigned TF representation is obtained by picking orthonormal 'windows' (used to isolate portions of the TF representation when working with a reassignment method), and averaging the reassigned TF representations determined by each of the individual windows. Heuristically, the concentration for a 'true' constituting component of the signal will be in similar locations in the TF plane for each of the individual reassigned TF representations, whereas the spurious concentrations, artefacts of correlations between noise and the windowing function, tend not to be co-located and have a diminished impact when averaged. In the SST context, a similar multi-taper idea was used by one of us in a study of anaesthesia depth [5,44], in which J different

window functions h_j, $j = 1, \ldots, J$, are considered, and the multitaper SST (MTSST) is computed as the average of the individual $S_f^{(h_j)}$: $MS_f(t, \xi) := (1/J) \sum_{j=1}^{J} S_f^{(h_j)}(t, \xi)$. Using multiple tapers reduces artefacts, and the MTSST remains 'readable' at higher noise levels than a 'simple' SST [5,44]. To suppress noise artefacts better, it is tempting to consider increasing J. However, the area in the TF plane over which the signal TF information is 'smeared out' also increases (linearly) with J, and a balance needs to be observed; in the multitaper reassignment method of [64], for instance, six Hermite functions were used (i.e. $J = 6$).

In this paper, we introduce a new approach to obtain better concentrated time–frequency representations, which we call *ConceFT*, for 'concentration of frequency and time'. It is based on STFT- or CWT-based SSTs, but the approach could be applied to yet other TF decomposition tools. The ConceFT algorithm will be defined precisely below, in §2. Like MTSST, ConceFT starts from a multilayered time–frequency representation, but instead of averaging the SST results obtained from STFT or CWT for orthonormal windows, which can be viewed as elements in a vector space of time–frequency functions, it considers many different projections in this same vector space, and averages the corresponding SSTs; for more details, see §2; figures illustrating the result of STFT-based ConceFT can be found in the Introduction section of the electronic supplementary material. Section 3 studies the theoretical properties of ConceFT, and explains how it can provide reliable results under challenging SNR conditions; finally, in §4, we provide several numerical results.

2. The ConceFT algorithm

We start by briefly reviewing SST. In the Introduction, we defined STFT-based SST, discussed in more detail in [21,26]; to show that the situation is very similar for CWT-based SST, we discuss that case here; see [19,28] for details. We start with the wavelet ψ with respect to which the CWT will be computed, which must necessarily have mean zero; that is, $\int \psi(t) \, dt = 0$; let us also pick it to be a Schwartz function. We shall assume that we are dealing with real signals f; in this case, the symmetry in ξ of $\hat{f}(\xi)$ makes it possible to consider only the 'positive-frequency part' of f, by picking ψ so that its Fourier transform $\hat{\psi}$ is supported on $\mathbb{R}_+$. (The approach can be extended easily to handle complex signals as well, but notation becomes a bit heavier.) Then the *continuous wavelet transform* $W_f^{(\psi)}(a, b)$ of a tempered distribution f, with the variables a, b standing for scale and time location, is defined as the inner product of f with $\psi^{(a,b)}(t) = |a|^{-1/2} \psi((t - b)/a)$. Even if the Fourier transform $\hat{f}$ is very concentrated around some frequency ω_0, the magnitude $|W_f^{(\psi)}(a, b)|$ of the CWT will be spread out over a range of scales a, corresponding to a neighbourhood of ω_0. However, the phase information of $W_f^{(\psi)}$ will still hold a 'fingerprint' of ω_0 on that whole neighbourhood, in that $W_f^{(\psi)}(a, b)$ will show oscillatory behaviour in b, with frequency ω_0, for a range of different a. This is the motivation for the SST, which shifts the CWT coefficients 'back', according to certain reassignment rules determined by the phase information. More concretely, we set a *threshold* $\Gamma > 0$, and then define

$$\Omega_f^{(\psi, \Gamma)}(a, b) := \frac{-i \partial_b W_f^{(\psi)}(a, b)}{2\pi W_f^{(\psi)}(a, b)} \quad \text{when } |W_f^{(\psi)}(a, b)| > \Gamma$$

and

$$\Omega_f^{(\psi, \Gamma)}(a, b) = -\infty \quad \text{when } |W_f^{(\psi)}(a, b)| \leq \Gamma,$$

where ∂_b is the partial derivative with respect to b (see the electronic supplementary material for a remark concerning robust numerical implementation). The hard threshold Γ can be adjusted for best reduction of the numerical error and noise influence. For example, when the noise level is known or estimated with reasonable accuracy, Γ can be chosen so as to reduce the influence of the noise most effectively; see [33] for a discussion. The CWT-based SST then moves the CWT

coefficient $W_f^{(\psi)}(a,b)$ to the 'right' frequency slot, using $\Omega_f^{(\psi,\Gamma)}(a,b)$ as guideline:

$$S_f^{(\psi,\Gamma,\alpha)}(b,\xi) := \int_{\{a:|W_f^{(\psi)}(a,b)|>\Gamma\}} W_f^{(\psi)}(a,b)\frac{1}{\alpha}g_\alpha(\xi - \Omega_f^{(\psi,\Gamma)}(a,b))a^{-3/2}\,\mathrm{d}a, \tag{2.1}$$

where $0 < \alpha \ll 1$ is chosen by the user, g is a smooth function so that $g_\alpha(\cdot) := (1/\alpha)g(\cdot/\alpha) \to \delta$ in the weak sense as $\alpha \to 0$, and the factor $a^{-3/2}$ is introduced to ensure that the integral of $S_f^{(\psi,\Gamma,\alpha)}(b,\xi)$ over ξ yields a close approximation to the original $f(b)$. For more details, we refer the reader to [19,28].

Although both the CWT $W_f^{(\psi)}$ and its derived SST $S_f^{(\psi,\Gamma,\alpha)}$ depend on the choice of the reference wavelet ψ, this is much less pronounced for the SST; CWT-based SST corresponding to different reference wavelets lead to different but very similar TF representations. (Theoretical reasons for this can be found in [19,28].) In particular, the dominant components in the TF representations are very similar. Moreover, even when the signal is contaminated by noise, these dominant components in the TF representations are not significantly disturbed [28]. However, the distribution of artefacts across the TF representation, induced by the noise, as seen in e.g. the middle left panel of figure S.1, vary from one reference wavelet to another; this can be intuitively explained by observing that the CWT is essentially a convolution with (scaled versions of) the reference wavelet, so that the wavelet transforms of independent and identically distributed (i.i.d.) noise based on different *orthogonal* reference wavelets are *independent*. These observations lead to the idea of a multitaper SST algorithm [5,44]. In brief, given J orthonormal reference wavelets ψ_j, $j = 1,\ldots,J$, one determines the reassignment rules $\Omega_f^{(\psi_j,\Gamma)}(a,b)$, as well as the corresponding $S_f^{(\psi,\Gamma,\alpha)}(b,\xi)$, and then defines the MTSST by

$$\mathrm{MS}_f^{\Gamma,\alpha}(b,\xi) := \frac{1}{J}\sum_{j=1}^{J} S_f^{(\psi_j,\Gamma,\alpha)}(b,\xi). \tag{2.2}$$

This suggests that averaging over a large number of orthonormal reference wavelets would smooth out completely the TF artefacts induced by the noise, as originally discussed for the reassignment method [64]. However, in order for reassignment to make sense, the reference function, whether it is the window h for STFT or the wavelet ψ for CWT, must itself be fairly well concentrated in time and frequency, so that inner products with modulated window functions or scaled wavelets do not mix up different components and behaviours of the signal. On the other hand, there is a limit to how many orthonormal functions can be 'mostly' supported in a concentrated region in the TF plane—by a rule of thumb generalizing the Nyquist sampling density one can find, for a region $\mathcal{R}$ in the TF plane, only $\mathrm{Area}(\mathcal{R})/(2\pi)$ orthonormal functions that are mostly concentrated on $\mathcal{R}$ [67]. This limits how many different orthonormal ψ_j can be used in MTSST.

ConceFT uses the different TF 'views' provided by the CWT transforms $W_f^{(\psi_j)}$ in a different way, exploiting the *nonlinearity* of the SST operation. (See the electronic supplementary material for a sketch of an alternative way in which one could extend multitaper CWT; not pursued in this paper, however.) For each choice of ψ, the collection of CWT $W_f^{(\psi)}$, where f ranges over the class of signals of interest, span a subspace of $\mathcal{F}$, the space of all reasonably smooth functions of the two variables a, b. Different orthonormal ψ_j generate different subspaces in $\mathcal{F}$; combined, they generate a larger subspace, in which one can define an infinite number of 'sections', each corresponding to the collection of CWT generated by one reference wavelet. Each linear combination of the ψ_j defines such a CWT space, in which one can carry out the corresponding SST operation. For $\psi = \sum_{j=1}^{J} r_j\psi_j$, where $r_j \in \mathbb{R}$, one has $W_f^{(\psi)} = \sum_{j=1}^{J} r_j W_f^{(\psi_j)}$; because synchrosqueezing is a highly nonlinear operation, the corresponding $S_f^{(\psi,\Gamma,\alpha)}$ are however not linear combinations of the $S_f^{(\psi_j,\Gamma,\alpha)}$. In practice, the artificial concentrations in the TF plane, triggered by fortuitous correlations between the noise and the (overcomplete) $\psi^{(a,b)}$, occur

at locations sufficiently different, for different choices of the vector $r = (r_1, \ldots, r_J)$, that averaging over many choices of r successfully suppresses noise artefacts.

More precisely, the CWT-based ConceFT algorithm proceeds as follows:

— Take J orthonormal reference wavelets, $\psi_1, \ldots, \psi_J$, in the Schwartz space, with good concentration in the TF plane.
— Pick N random vectors r_n, $n = 1, \ldots, N$, of unit norm, in $\mathbb{R}^J$; that is, uniformly select N samples in S^{J-1}.
— For each $n = 1, \ldots, N$, define $\psi_{[n]} := \sum_{j=1}^{J} (r_n)_j \psi_j$, and $W_f^{(\psi_{[n]})} = \sum_{j=1}^{J} (r_n)_j W_f^{(\psi_j)}$.
— Select the threshold $\Gamma > 0$ and the approximation parameter $\alpha > 0$, and evaluate, for each n between 1 and N, the corresponding CWT-based SST of f by computing the reassignment rule $\Omega_f^{(\psi_{[n]}, \Gamma)}(a, b)$, and hence $S_f^{(\psi_{[n]}, \Gamma, \alpha)}(b, \xi)$, as defined above, with the minor adjustment that when the expression $\sum_{j=1}^{J} (r_n)_j W_f^{(\psi_j)}(a, b)$ in the reassignment rule denominator has a negative real part, we switch to the vector $-r_n$.
— The final ConceFT representation of f is then the average

$$C_f^{\Gamma, \alpha}(b, \xi) := \frac{1}{N} \sum_{n=1}^{N} S_f^{(\psi_{[n]}, \Gamma, \alpha)}(b, \xi). \tag{2.3}$$

In practice, J could be as small as 2, while N could be chosen as large as the user wishes.

The square of the magnitude of $C_f^{\Gamma, \alpha}(b, \xi)$, $\tilde{P}_f(b, \xi) := |C_f^{\Gamma, \alpha}(b, \xi)|^2$, can be of interest in its own right, as an estimated *time-varying power spectrum* (tvPS) of f.

STFT-based ConceFT representations are defined entirely analogously, based on the STFT reassignment rule given in §1.

3. Theoretical results

In this section, we list and explain theoretical results about CWT-based ConceFT. The detailed mathematical computations and proofs can be found in the electronic supplementary material. Entirely similar results hold for STFT-based ConceFT; as they are established by the same arguments, we skip those details. We start by recalling the structure of our signal space, as introduced in [19,28]. We emphasize that this is, to a large extent, a purely phenomenological model, constructed so as to reflect the fairly (but not exactly) periodic nature of many signals of interest, in particular (but not only) those of a physiological origin (see the discussion in [6]).

A single-component or *intrinsic-mode type* (IMT) function has the following form:

$$F(t) = A(t) \cos(2\pi \varphi(t)), \tag{3.1}$$

where both the amplitude modulation $A(t)$ and the phase function $\varphi(t)$ are reasonably smooth; in addition, both $A(t)$ and the derivative $\varphi'(t)$ (or the 'instantaneous frequency') are strictly positive at all times as well as bounded; finally, we assume that A and φ' vary in time at rates that are slow compared with the instantaneous frequency of F itself. For the precise mathematical formulation of these conditions, we refer to the electronic supplementary material; this precise formulation invokes a few parameters, one of which, ϵ, bounds the ratio of the rate of change of A and φ'. This parameter will play a role in our estimates below. (Although we are assuming the signal to be real-valued here, all this can easily be adapted to the complex case by replacing the cosine with the corresponding complex exponential; the discussion in the remainder of this section can be adapted similarly.) We also consider signals that contain several IMT components, that is, functions of the type

$$G(t) = \sum_{\ell=1}^{L} F_\ell(t) = \sum_{\ell=1}^{L} A_\ell(t) \cos(2\pi \varphi_\ell(t)), \tag{3.2}$$

where each F_ℓ is an IMT function, and we assume in addition that the instantaneous frequencies $\varphi'_\ell(t)$ are ordered (higher ℓ corresponding to larger φ'_ℓ) and well separated,

$$\varphi'_{\ell+1}(t) - \varphi'_\ell(t) > d(\varphi'_{\ell+1}(t) + \varphi'_\ell(t)) \tag{3.3}$$

for all $\ell = 1, \ldots, L - 1$, for some d with $0 < d < 1$. We denote by $\mathcal{A}$ the set of all such functions G; it provides a flexible *adaptive harmonic model* space for a wide class of signals of interest. (Strictly speaking, they are not 'truly' harmonic, if harmonicity is interpreted—as it often is—as 'having components with frequencies that are integer multiples of a fundamental frequency'.)

Next, we turn to the noise model for which we prove our main theoretical result. For the purposes of this theoretical discussion, we use a simple additive Gaussian white noise (even though the approach works for much more challenging noise models as well—see the electronic supplementary material). That is, we consider our noisy signals to be of the form

$$Y(t) = G(t) + \sigma \Phi(t) = \sum_{\ell=1}^{L} F_\ell(t) + \sigma \Phi(t) = \sum_{\ell=1}^{L} A_\ell(t) \cos(2\pi \varphi_\ell(t)) + \sigma \Phi(t), \tag{3.4}$$

where $G = \sum_{\ell=1}^{L} F_\ell$ is in $\mathcal{A}$, Φ is a Gaussian white noise with standard deviation 1 and $\sigma > 0$ is the noise level. Y is a generalized random process, since by definition G is a tempered distribution. We could extend this, introducing also the trend and a more general noise model as in [28], the wave-shape function used in [30], or the generalized IMT functions that model oscillatory signals with fast varying instantaneous frequency of [68]. Since none of these generalizations would significantly affect the mathematical analysis, we restrict ourselves to the model (3.4).

Finally, we describe the wavelets $\psi_1, \ldots, \psi_J$ with respect to which we compute the CWT of Y. For the sake of convenience of the theoretical analysis, we assume that they are smooth functions with fast decay, and that their Fourier transforms $\widehat{\psi}_j$ are all real functions with compact support, supp $\widehat{\psi}_j \subset [1 - \Delta_j, 1 + \Delta_j]$, where $0 < \Delta_j < 1$. We also assume that the $\psi_1, \ldots, \psi_J$ form an orthonormal set, that is, $\int \psi_i(x) \bar{\psi}_j(x) \, dx = \delta_{i,j}$, where $\delta_{i,j}$ is the Kronecker delta. To build appropriate linear combinations of the ψ_j, we define, for any unit-norm vector $\mathbf{r} = (r_1, \ldots, r_J)$ in $\mathbb{R}^J$, the corresponding combination as $\psi^{[r]} := \sum_{j=1}^{J} r_j \psi_j$. It is convenient to characterize intervals for the scale a such that the support of $\widehat{\psi_j^{(a,b)}}$ overlaps $\varphi'_\ell(b)$, where $\psi_j^{(a,b)}(t) := (1/\sqrt{a})\psi_j((t - b)/a)$; we thus introduce the notation $Z_\ell^{(j)}(b) = [(1 - \Delta_j)/\varphi'_\ell(b), (1 + \Delta_j)/\varphi'_\ell(b)]$. It then follows from the definition of the CWT as the inner product between the signal and scaled, translated versions of the wavelets that [19,28]

$$W_{F_\ell}^{(\psi_j)}(a,b) = \begin{cases} e^{i2\pi\varphi_\ell(b)} Q_{j,\ell}(a,b) + \epsilon_j(a,b) & \text{when } a \in Z_\ell^{(j)}(b), \\ \epsilon_j(a,b) & \text{otherwise,} \end{cases} \tag{3.5}$$

where $Q_{j,\ell}(a,b) := A_\ell(b)\sqrt{a}\, \overline{\widehat{\psi}_j(a\varphi'_\ell(b))} \in \mathbb{R}$ and ϵ_j is of order ϵ for all $j = 1, \ldots, J$. Here ϵ_j depends on the first three absolute moments of ψ_j and ψ'_j and the model parameters. It follows that the CWT of Y, with respect to ψ_j, is given by

$$W_Y^{(\psi_j)}(a,b) = \sum_{\ell=1}^{L} e^{i2\pi\varphi_\ell(b)} Q_{j,\ell}(a,b) \chi_{Z_\ell^{(j)}(b)} + \epsilon_j(a,b) + \sigma \Phi(\psi_j^{(a,b)}), \tag{3.6}$$

where $\chi_{Z_\ell^{(j)}(b)}$ is the indicator function of the set $Z_\ell^{(j)}(b)$; note that the $\epsilon_j(a,b)$ term, again of order ϵ, need not be the same as before. As shorthand notations, we will use bold symbols to regroup quantities indexed by $j = 1, \ldots, J$ into one J-dimensional vector, e.g. $\boldsymbol{\epsilon}(a,b) = [\epsilon_1(a,b), \ldots, \epsilon_J(a,b)]^\mathsf{T}$ (which has norm of order ϵ), $\boldsymbol{\Phi}(a,b) = [\Phi(\psi_1^{(a,b)}), \ldots, \Phi(\psi_J^{(a,b)})]^\mathsf{T}$ (a complex Gaussian random vector [69], with mean $[0, \ldots, 0]^\mathsf{T} \in \mathbb{R}^J$, and covariance as well as relation matrix equal to $I_{J \times J}$—see the electronic supplementary material), $\mathbf{W}_Y^{\boldsymbol{\psi}}(a,b) = [W_Y^{(\psi_1)}(a,b), \ldots, W_Y^{(\psi_J)}(a,b)]^\mathsf{T}$, and $\mathbf{Q}_\ell(a,b) :=$

$[Q_{1,\ell}(a,b),\dots,Q_{J,\ell}(a,b)]^{\mathsf{T}}$. Finally, $W_Y^{(\psi^{[r]})}(a,b):=\boldsymbol{r}^{\mathsf{T}}\boldsymbol{W}_Y^{\psi}(a,b)$ or, more explicitly,

$$W_Y^{(\psi^{[r]})}(a,b)=\sum_{\ell=1}^{L}\sum_{j=1}^{J}r_j\,\mathrm{e}^{\mathrm{i}2\pi\varphi_j(b)}Q_{j,\ell}(a,b)\chi_{Z_\ell^{(j)}}(a,b)+\boldsymbol{r}^{\mathsf{T}}[\boldsymbol{\epsilon}(a,b)+\sigma\boldsymbol{\Phi}(a,b)]. \tag{3.7}$$

Under the general assumptions for our model, a similar calculation as for (3.7) leads to

$$-\mathrm{i}\partial_b W_Y^{(\psi^{[r]})}(a,b)=2\pi\left(\sum_{j=1}^{J}\sum_{\ell=1}^{L}r_j\varphi_j'(b)\,\mathrm{e}^{\mathrm{i}2\pi\varphi_j(b)}Q_{j,\ell}(a,b)\chi_{Z_\ell^{(j)}}(a,b)+\boldsymbol{r}^{\mathsf{T}}[\tilde{\boldsymbol{\epsilon}}(a,b)+\sigma\tilde{\boldsymbol{\Phi}}(a,b)]\right),$$

where $\tilde{\boldsymbol{\epsilon}}(a,b)$ is again a J-dimensional vector of order ϵ, and $\tilde{\boldsymbol{\Phi}}(a,b):=[\Phi(\psi_1^{(a,b)\prime}),\dots,\Phi(\psi_J^{(a,b)\prime})]^{\mathsf{T}}$, where $\psi_i^{(a,b)\prime}(t):=\partial_t\psi_i^{(a,b)}(t)$ is again a complex Gaussian random vector. The scalar products $\boldsymbol{r}^{\mathsf{T}}\boldsymbol{\Phi}(a,b)$ and $\boldsymbol{r}^{\mathsf{T}}\tilde{\boldsymbol{\Phi}}(a,b)$ are independent complex Gaussian random variables, with mean 0 and variance $\|\boldsymbol{r}\|^2$ and $\sum_{j=1}^{J}r_j^2\|\psi_j'\|^2/(2\pi a)^2$, respectively. (See the electronic supplementary material.) Set now $Z_\ell(b)=\bigcap_{j=1}^{J}Z_\ell^{(j)}(b)$. Then it follows that for $a\in Z_\ell(b)$, we get the following reassignment for the CWT $W_Y^{(\psi^{[r]})}$:

$$\omega_Y^{(\psi^{[r]})}(a,b)=\frac{-\mathrm{i}\partial_b W_Y^{(\psi^{[r]})}(a,b)}{2\pi W_Y^{(\psi^{[r]})}(a,b)}=\frac{\boldsymbol{r}^{\mathsf{T}}[\varphi_\ell'(b)\,\mathrm{e}^{\mathrm{i}2\pi\varphi_\ell(b)}\boldsymbol{Q}_\ell(a,b)+\tilde{\boldsymbol{\epsilon}}(a,b)+\sigma\tilde{\boldsymbol{\Phi}}(a,b)]}{\boldsymbol{r}^{\mathsf{T}}[\mathrm{e}^{\mathrm{i}2\pi\varphi_\ell(b)}\boldsymbol{Q}_\ell(a,b)+\boldsymbol{\epsilon}(a,b)+\sigma\boldsymbol{\Phi}(a,b)]}, \tag{3.8}$$

which is a ratio random variable of two dependent complex Gaussian random variables with non-zero means. Next, we consider, for each *fixed* realization of the random noise, the unit-norm vector $\boldsymbol{r}\in\mathbb{R}^J$ as a random vector, picked uniformly from $S_\kappa:=\{\boldsymbol{r}\in S^{J-1};\ |\boldsymbol{r}^{\mathsf{T}}\boldsymbol{W}_Y^{(\psi)}(a,b)|>2\kappa$ and $\mathrm{Re}(\boldsymbol{r}^{\mathsf{T}}\boldsymbol{W}_Y^{(\psi)}(a,b))>0\}\subset S^{J-1}$. Restricting the choice of $\boldsymbol{r}$ to the subset of S^{J-1} for which the inner product of $\boldsymbol{r}$ and $\boldsymbol{W}_Y^{\psi}(a,b)$ has magnitude larger than 2κ reflects the threshold used in the SST algorithm (see §2); restricting $\boldsymbol{r}$ so that the inner product has positive real part means that we sample $\boldsymbol{r}$ from a half-sphere rather than the whole sphere. (See the electronic supplementary material for more details.)

Assuming that the bound on the noise is such that $\|\boldsymbol{\epsilon}+\sigma\boldsymbol{\Phi}\|_2^2<\kappa$, the expectation of $\omega_Y^{(\psi_r)}(a,b)$ over S_κ is then given by

$$\mathbb{E}_{\boldsymbol{r}}\,\omega_Y^{(\psi_r)}(a,b)=\varphi_\ell'(b)+\mathrm{e}^{-\mathrm{i}2\pi\varphi_\ell(b)}\mathfrak{p}_{\boldsymbol{Q}_\ell(a,b)}(\boldsymbol{V}_\ell(a,b))+E_1, \tag{3.9}$$

where $\boldsymbol{V}_\ell(a,b):=\tilde{\boldsymbol{\epsilon}}(a,b)+\sigma\tilde{\boldsymbol{\Phi}}(a,b)-\varphi_\ell'(b)[\boldsymbol{\epsilon}(a,b)+\sigma\boldsymbol{\Phi}(a,b)]$, $\mathfrak{p}_{\boldsymbol{v}}$ denotes 'taking the component along' a vector $\boldsymbol{v}$, that is, $\mathfrak{p}_{\boldsymbol{v}}(\boldsymbol{u})=\boldsymbol{v}^{\mathsf{T}}\boldsymbol{u}/\|\boldsymbol{v}\|$, and E_1 is bounded by

$$|E_1|\le\frac{1}{2}\left(\left[1-\frac{c}{J-1}\right]|\mathfrak{p}_{\boldsymbol{Q}_\ell(a,b)}(\boldsymbol{V}_\ell(a,b))|^2+c\frac{\|\boldsymbol{V}_\ell(a,b)\|^2}{J-1}\right)^{1/2}. \tag{3.10}$$

Furthermore, the variance is bounded by

$$\mathrm{Var}_{\boldsymbol{r}}\,\omega_Y^{(\psi^{[r]})}(a,b)\le\frac{5}{2}\left(\left[1-\frac{c}{J-1}\right]|\mathfrak{p}_{\boldsymbol{Q}_\ell(a,b)}(\boldsymbol{V}_\ell(a,b))|^2+c\frac{\|\boldsymbol{V}_\ell(a,b)\|^2}{J-1}\right). \tag{3.11}$$

A detailed derivation, and an explicit expression for the constant c, is given in the electronic supplementary material; if J becomes large, we have $c\approx 2\sqrt{2}/[\kappa\sqrt{\pi J}]$. The quantity $\mathfrak{p}_{\boldsymbol{Q}_\ell(a,b)}(\boldsymbol{V}_\ell(a,b))=\mathfrak{p}_{\boldsymbol{Q}_\ell(a,b)}(\tilde{\boldsymbol{\epsilon}}(a,b)+\sigma\tilde{\boldsymbol{\Phi}}(a,b)-\varphi_\ell'(b)[\boldsymbol{\epsilon}(a,b)+\sigma\boldsymbol{\Phi}(a,b)])$ or its absolute value occur in several of these estimates. Estimates in the electronic supplementary material show that, although $\|V\|^2$ itself is expected to be of order $J[\bar{\epsilon}^2+\sigma^2]$, the expectation of $|\mathfrak{p}_Q(V)|^2$ is bounded above independently of J. This does not explain our observation that ConceFT estimates seem to be *more* concentrated that those obtained by SST or MTSST; we will return to this in further work.

The detailed estimates given in section ESM-3 in the electronic supplementary material are derived under the restrictive conditions listed at the start of this section for the signals and the

wavelets used. However, as noted above, these conditions can be relaxed significantly (at the price of more intricate estimates). In practice, we observe similar behaviour in our numerical examples even for more complex situations; in particular, the method can handle noise models that are much more challenging, as illustrated in the next section as well as by figure S.1.

4. Numerical experiments

In this section, we demonstrate the results of the CWT-based ConceFT algorithm on examples; we also discuss different choices for some of the different parameters involved. The results of the STFT-based ConceFT algorithm are shown in the electronic supplementary material. The ConceFT Matlab code and the codes leading to the figures in this paper can be found at https://sites.google.com/site/hautiengwu/home/download.

The first choice to be made, when applying CWT- or STFT-based ConceFT, concerns the family of orthonormal reference functions (wavelets or window functions) for the underlying wavelet or windowed Fourier transform. In both cases, we pick a family of eigenfunctions for a time–frequency localized operator designed for the CWT or STFT framework; as shown in [70,71] these can provide 'optimal' localization within a restricted region of time–frequency space, where the size of the region depends solely on the number of functions used. More precisely, we use orthonormal Hermite functions for the STFT case [64,70] (see also figure S.1), and Morse wavelets for the CWT case [71,72]. In both cases, the shape of the localization domain in the TF plane is not completely fixed, but can be adjusted by varying some parameters; for details, see the electronic supplementary material. Once the family of orthonormal reference functions is fixed, we need to decide how many ψ_j, $j = 1, \ldots, J$, we pick; this corresponds to choosing the size of the corresponding domain of concentration in the TF plane. Flexibility in the choices of shape and size of the TF localization domain makes it possible to adapt ConceFT, to some extent, to the family of signals under consideration. Finally, ConceFT also depends on the number N of random projections chosen (see §§2 and 3). In principle, the larger N, the closer the results are to the expected value of the random process, and the more we expect accidental correlations between reference function and the noise to cancel out in regions of the TF plane where the signal does not reside; in practice, increasing N beyond a certain value does not appreciably improve the results. In what follows, we explore these different choices for the CWT case, on a simple family of challenging examples, with noise of different types (white Gaussian, Poisson and autoregressive moving-average (ARMA)), and of different strengths.

For our test data, we restrict ourselves to simulated signals only, so as to be able to quantify the deviation from the 'ground truth', usually not available in real-life applications. (ConceFT results on concrete signals will appear elsewhere [73].) On the other hand, we want to avoid parametric models, so as to be sufficiently general. Accordingly, we generate a class $\mathcal{C}$ of non-stationary data via a random process described in §4a; each realization provides us not only with a (simulated) clean signal, but also with the exact 'ground truth' for the time-dependent instantaneous frequency and amplitude of the components of that signal. The same subsection also describes in detail three different noise models (white Gaussian, Poisson and ARMA(1,1)) for which the approach is tested. After applying ConceFT to signals in $\mathcal{C}$, we want to compare the ConceFT results with the optimal, ground-truth TF representation; to quantify their (dis)similarity, we use an optimal transport (OT) distance, as described in §4b. In §4c, we discuss how choices of the parameters and of the number of orthogonal Morse wavelets impact the ConceFT results, for this family of examples; §4d illustrates the effect of the number N of random projections. Finally, in §4e, we explore the effect on CWT-based ConceFT of different noise levels, for each of the three noise types we consider; §4f briefly discusses the STFT case.

(a) Data simulation

To generate a typical multicomponent signal, we use smoothened Brownian path realizations to model the non-constant amplitudes and the instantaneous frequencies of the components; more
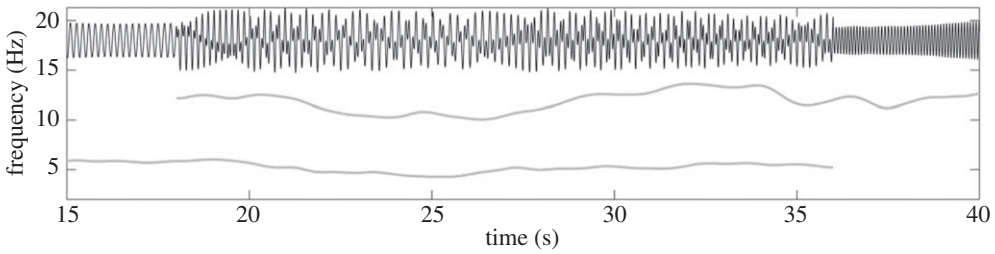
Figure 1. The signal *s* (top, in black) and the corresponding instantaneous frequencies (below, in grey) of the two components, restricted to the time interval [15, 40].

precisely, if W is the standard Brownian motion defined on $[0, \infty)$, then we define the *smoothened Brownian motion with bandwidth* $B > 0$ as $\Phi_B := W \star K_B$, where K_B is the Gaussian function with standard deviation $B > 0$ and $\star$ denotes the convolution operator. Given $T > 0$ and parameters $\zeta_1, \ldots, \zeta_6 > 0$, we then define the following family of random processes on $[0, T]$:

$$\Psi_{[\zeta_1, \ldots, \zeta_6]}(t) := \zeta_1 + \zeta_2 t + \zeta_3 \frac{\Phi_{\zeta_4}(t)}{\|\Phi_{\zeta_4}\|_{L^\infty[0,T]}} + \zeta_5 \int_0^t \frac{\Phi_{\zeta_6}(s)}{\|\Phi_{\zeta_6}\|_{L^\infty[0,T]}} \, \mathrm{d}s. \tag{4.1}$$

For the amplitude $A(t)$ of each IMT, we set $\zeta_2 = \zeta_5 = 0$; every realization then varies smoothly between ζ_1 and $\zeta_1 + \zeta_3$. In the examples shown below and in the electronic supplementary material, the signal consists of two components (i.e. $L = 2$) on $[0, 60]$; their two amplitudes are independent realizations of $\Psi_{[2,0,1,200,0,0]}(t)$. To simulate a phase function, we set $\zeta_1 = \zeta_3 = 0$; $\Psi_{[0,\zeta_2,0,0,\zeta_5,\zeta_6]}(t)$ is then, appropriately, a monotonically increasing process. In the examples we consider, we take for $\varphi_1(t)$ a realization of $\Psi_{[0,10,0,0,6,400]}(t)$ for $t \in [0, 60]$, and for $\varphi_2(t)$ a realization of $\Psi_{[0,2\pi,0,0,2,300]}(t)$. We constrain each component to 'live' on only part of the interval, by setting

$$s(t) = A_1(t) \cos(2\pi \varphi_1(t)) \chi_{[18,60]}(t) + A_2(t) \cos(2\pi \varphi_2(t)) \chi_{[0,36]}(t) =: s_1(t) + s_2(t),$$

where $\chi_{[\tau_1, \tau_2]}$ is the indicator function of $[\tau_1, \tau_2]$; that is, $\chi_{[\tau_1, \tau_2]}(t) = 1$ if $\tau_1 \leq t \leq \tau_2$, $\chi_{[\tau_1, \tau_2]}(t) = 0$ otherwise. We shall denote the resulting class of two-component signals by $\mathcal{C}$. In our examples, signals in $\mathcal{C}$ are sampled uniformly at rate 160 Hz, corresponding to 9600 samples. Figure 1 plots $s(t)$ for one example $s \in \mathcal{C}$, as well as the instantaneous frequencies (IFs) of its two components, all restricted to the subinterval $[15, 40] \subset [0, 60]$.

Note that the signal s should not be viewed as a random process itself—we use the random processes $\Psi_{[\zeta_1, \ldots, \zeta_6]}$ as a means to generate signals consisting of several components for which the amplitudes and instantaneous frequencies are not easily expressed analytically, but we will not consider or compute expectations with respect to these processes—once $s \in \mathcal{C}$ is generated, we consider it fixed when we apply ConceFT to it. (In further subsections, we shall encounter other elements of $\mathcal{C}$.)

To study the performance of ConceFT in the presence of noise, we add noise to $s(t)$, setting $Y(t_k) = s(t_k) + \sigma \xi(t_k)$, where t_k is the kth sampling time and ξ is a stationary random process. We shall consider three different noise models; in each case, we set the value of σ so that the SNR, defined as $\mathrm{SNR} := 20 \log[\mathrm{std}(s)/(\sigma \, \mathrm{std}(\xi))]$, equals 0 dB. The three noise models we consider are Gaussian white noise, an ARMA noise and Poisson noise. For the ARMA case, we consider an ARMA(1, 1) model determined by autoregression polynomial $a(z) = 0.5z + 1$ and moving-averaging polynomial $b(z) = -0.5z + 1$; for the innovation process we use i.i.d. Student t_4 random variables. (Note that this ARMA(1, 1) noise is not white, because of the time dependence; in addition, the Student t_4 random variable has a 'fat-tailed' distribution, resulting in possibly spiky realizations.) For the Poisson noise, we pick the $\xi(t_k)$ to be independent and identically sampled from the Poisson distribution with parameter $\lambda = 1$. Figure 2 plots
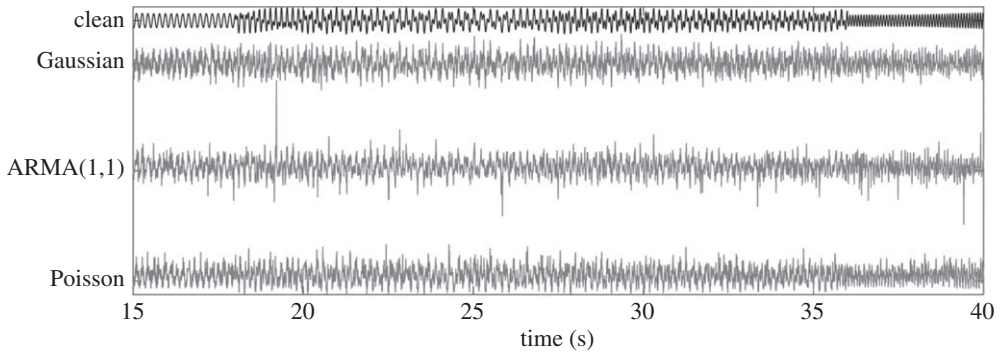
Figure 2. The restrictions to [15, 40] of the clean signal s (top) and of the noisy signal $Y = s + \sigma\xi$, where the added noise is Gaussian, ARMA(1,1), or Poisson noise (below, in order); in each case σ is picked so that the noisy signal has 0 dB SNR. All signals are plotted at the same scale.

a realization of $Y(t) = s(t) + \sigma\xi(t)$ for each of these three noise processes, restricted to the subinterval [15,40].

(b) Performance evaluation

To evaluate the performance of ConceFT, we propose comparing the time-varying power spectrum (tvPS, defined at the end of §2) of the results of the ConceFT analysis of Y with the ideal time-varying power spectrum (itvPS) of our simulated signal s, which can easily be defined explicitly (because our construction was designed accordingly) as follows [68]:

$$P_s(t, \omega) := \sum_{k=1}^{2} A_k^2(t)\delta_{\varphi_k'(t)}(\omega). \tag{4.2}$$

In order to quantify the (dis)similarity between the ConceFT-estimated tvPS $\tilde{P}_Y$ and the itvPS P_s, we use the OT distance (also called the earth mover distance) between pairs of *probability* distributions, μ and ν, on $\mathbb{R}$. Denote $f_\mu(x) = \int_{-\infty}^{x} d\mu$ (analogously for f_ν). Then the OT distance between μ and ν is defined as

$$d_{OT}(\mu, \nu) := \int_S |f_\mu(x) - f_\nu(x)|\, dx. \tag{4.3}$$

For a more general definition and discussion of the OT distance, see the electronic supplementary material. The principle of ConceFT is to 'reassign' content in the TF plane, keeping the time variable fixed (see §2); we therefore choose to keep t fixed for our OT distance. Because the positive functions $\tilde{P}_Y(t, \cdot)$ and $P_s(t, \cdot)$ might not have integral 1 for all t, we normalize them before computing their OT distance. After normalization, we interpret, at each time t, $\tilde{P}_Y(t, \omega)$ and $P_s(t, \omega)$ as (probability) distributions in ω and compute the OT distance between them, which essentially measures how much one distribution needs to be 'deformed' in order to coincide with the other; this is repeated for all t, and the average of the t-dependent individual OT distances then indicates the quality of the estimator $\tilde{P}_Y$ for P_s. Figure 3 displays four examples in which two delta measures localized on curves in the TF plane (similar to the itvPS defined above) lie at similar OT distances of each other—although in each example the distance indicates a different type of 'distortion'. Together, these examples give an intuitive understanding of the way in which OT distances capture the difference between the TF distributions of interest to us here.
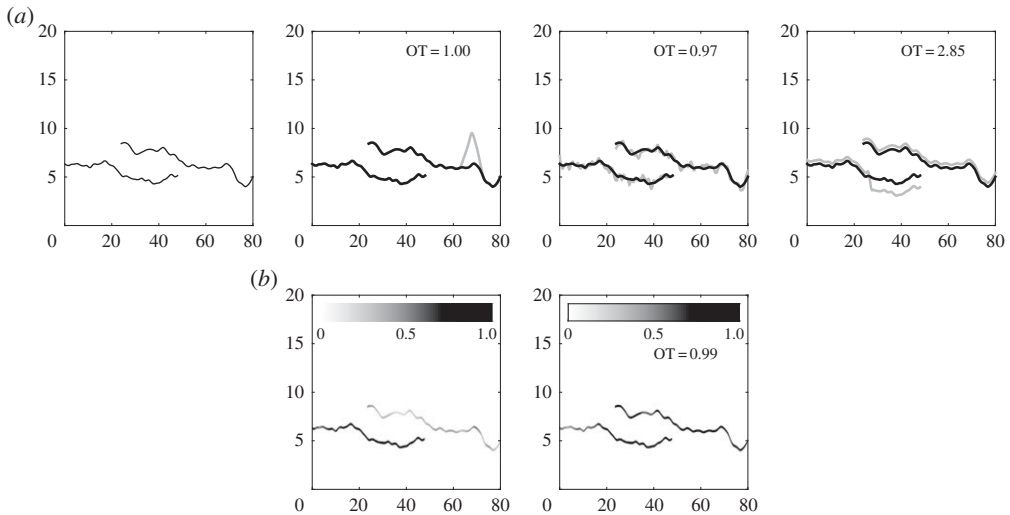
Figure 3. (*a*) Left: the TF localization of the itvPS of a two-component simulated signal s_a (not showing the amplitude modulation (AM)). Other itvPS shown in the top row are for signals s_b, s_c and s_d that have a fairly small different OT distance with respect to s_a; the two components have the same time-dependent amplitudes as for s_a, but the instantaneous frequency curves have been moved (in order from left to right) by a narrow bump (left), a random dither (middle) and a shift (right). (*b*) Illustration of amplitude change. Left: the original itvPS of s_a with the AM values indicated by grey scale level; right: an itvPS example with the same IF but different AM. In all panels, the horizontal axis is time and the vertical axis is frequency. The image for each 'deformed' itvPS indicates its OT distance to the original itvPS (shown in the leftmost image on each row).

(c) Parameter selection

As described in [72], generalizing the construction in [71], orthonormal families of Morse wavelets can be defined for different values of two parameters, β and γ; different choices correspond to different shapes of the domain in the TF plane on which they are mostly localized (see the electronic supplementary material). Once the values of β and γ are chosen, determining the family of ψ_j, one also needs to select J, the total number of orthonormal reference wavelets used in the ConceFT method. For signals in $\mathcal{C}$ (see §4a), we explored systematically a range of (β, γ) pairs, as well as different values of J, to find the choice that, under different types of noise, with SNR of 0 dB, gave rise to the smallest OT-based distance (as described above) between the itvPS and the ConceFT-estimated tvPS. Surprisingly, the optimal choice depended very little on the type of noise; the optimal values we found are $\beta = 30$, $\gamma = 9$ and $J = 2$. (Detailed results are given in the electronic supplementary material.)

(d) Effect of the number of random projections

The ConceFT algorithm averages the SST results computed with N randomly picked reference wavelets (or windows, for STFT) from the linear span of the ψ_j, $j = 1, \ldots, J$. It is expected that the concentration in the TF plane observed with ConceFT kicks in only when N is sufficiently large; on the other hand, the larger N, the more expensive the computation. To explore the trade-off, we applied ConceFT to the three noisy versions of the signal $s \in \mathcal{C}$ (see §4a), with N ranging from 1 to 200. In all cases, the ConceFT algorithm uses the optimal parameters as described in §4c, i.e. it uses the first two Morse wavelets with parameters $\beta = 30$, $\gamma = 9$. In this simulation, each ConceFT computation was repeated 300 times and the mean and standard deviation of the OT distances of the ConceFT tvPS to the itvPS were computed. Figure 4 plots the results. For each of the three noise types, the graph of the average OT distance shows an 'elbow' shape, i.e. a regime in which the decrease is faster, as N increases, followed by one in which the decrease is
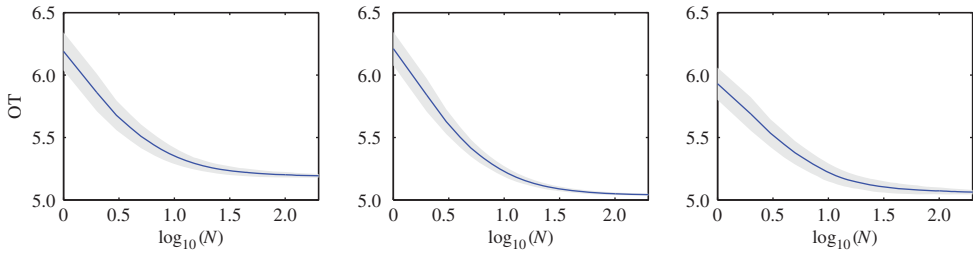
Figure 4. The OT distance as a function of the number N of random projections. The shaded band indicates the standard deviation of the OT distance at the corresponding number of random projections. From left to right, the noise types are Gaussian, ARMA(1,1) and Poisson, respectively. For all three experiments, $\beta = 30$, $\gamma = 9$ and the first two Morse wavelets are used. (Online version in colour.)
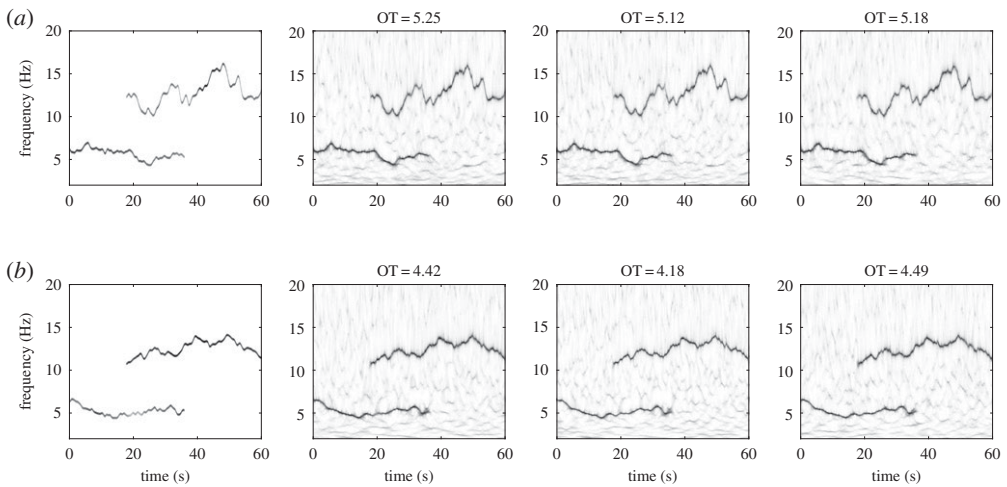

Figure 5. (*a*) Results for the signal s; (*b*) results for a new example s^*. Left to right: ideal time-varying TF power spectrum (itvPS) for the clean signal, followed by results of ConceFT with Morse wavelets after (in order) Gaussian, ARMA(1,1) or Poisson noise was added, with SNR of 0 dB. Clearly, even for an SNR as low as 0 dB, the results approximate the truth with high precision. For each of the tvPS panels, the header gives the OT distance to the corresponding itvPS.

less marked. The elbow is located around $N = 20$; the standard deviation is also quite small for this N. We accordingly decided to set $N = 20$ in our further experiments.

(e) ConceFT results for noisy signals

We now show the result of using ConceFT with the calibrated parameter choices. We illustrate the performance of ConceFT on signals of the simulation class $\mathcal{C}$ (see §4a), for a range of SNR, as well as on deterministic signals.

As a warm-up, we start with the signal s seen before. The top row of figure 5 plots the tvPS $\tilde{P}_Y$ of the three noisy versions of s next to the itvPS P_s. To compress the dynamical range of the tvPS plots, we carry out the following procedure. We first normalize the discretized version $\tilde{\mathbb{P}}_Y \in \mathbb{R}^{m \times n}$ of $\tilde{P}_Y$ (where m and n stand for the number of discrete frequencies and the number of time samples, respectively) by multiplying it by a constant so that the total weight of all entries equals the same number for all cases—i.e. for some $\theta > 0$ (to be picked—see below), $(1/nm) \sum_{k=1}^{m} \sum_{l=1}^{n} (\tilde{\mathbb{P}}_Y)_{k,l} = \theta$. We then plot a grey-scale visualization of $\boldsymbol{R} \in \mathbb{R}^{m \times n}$ rather than the (normalized) $\tilde{\mathbb{P}}_Y \in \mathbb{R}^{m \times n}$ itself, where $\boldsymbol{R}_{k,l} := \log(1 + \min\{\tilde{\mathbb{P}}_{k,l}, q\})$, $k = 1, \ldots, m$, $l = 1, \ldots, n$, and q is
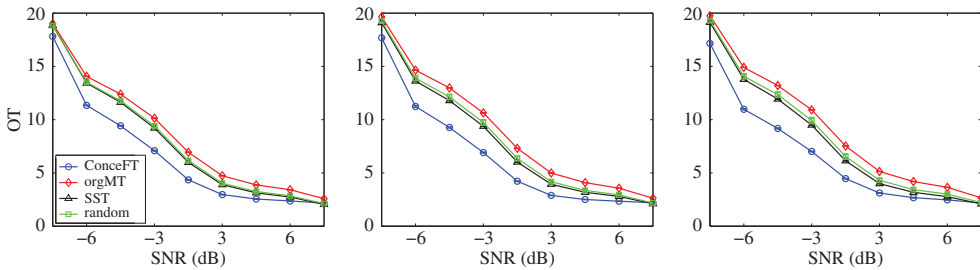
Figure 6. OT distance of ConceFT tvPS results against SNR of the signal $s^*(t)$, and comparison with standard SST and standard multitaper SST (see text). Noise type (left to right): Gaussian, ARMA(1,1) and Poisson. The standard deviation is smaller, at the scale of this figure, than the height of the markers, and has not been plotted. (Online version in colour.)

a (very high) cut-off to downplay the effect of far-off outliers. We choose q to be the same for all three tvPS, so that comparable grey levels on the different tvPS panels indicate comparable values of R (see §4f in the electronic supplementary material, for a more extensive discussion of choosing q and grey-scale plotting of tvPS). For the figures, we choose $\theta = 5$ and $q = 5.718$; this value for q is the minimum of the 99.8% quantiles of the different tvPSs. The second row of figure 5 gives the results for s^*, a signal of the simulation class $\mathcal{C}$ that was not used (in contrast to s) to calibrate parameters of ConceFT. The results are similarly highly accurate.

Next, we study the effect on the ConceFT performance of the noise level, as quantified by SNR. To this end, we revisit the analysis of the signal s^* (and s in the electronic supplementary material). For each signal, each type of noise (Gaussian, ARMA(1,1) or Poisson) and each SNR (SNR $= x$ dB, where $x \in \{-7, -6, \ldots, 6, 7\}$), we considered 20 independent realizations of the noise process; for each of the resulting noisy signals we carried out the ConceFT analysis and computed the OT distance of the tvPS to the itvPS of the clean signal; we then computed the mean and the standard deviation for each. The results are shown in figure 6. The same figure also compares the ConceFT results with those of simple SST (using either the first Morse wavelet with parameters $\beta = 30$, $\gamma = 9$ as reference wavelet, or *one* random linear combination of the two first Morse wavelets) and of multitaper SST (denoted as orgMT), using the same ψ_j as ConceFT. For each of these alternative methods, we likewise computed the mean OT distance of the tvPS to the itvPS for 20 noise realizations. It is striking that the ConceFT method outperforms the other methods in all cases.

Finally, to address possible concerns that the randomness in the generation and plots of $\varphi'(t)$ and $A(t)$ somehow 'help' ConceFT in these estimations, we show in figure 7 the results for yet another signal, which (in contrast to s and s^*) is completely deterministic; it consists of three components, each given by an explicit, analytic formula (again for $t \in [0, 60]$):

$$s^\circ(t) = \chi_{[10,48]}(t) \left(1 + 0.3 \cos\left(\frac{\pi(t-10)}{20}\right)^2\right) \cos\left(\frac{\pi}{3} + 5t + \frac{t^2}{50}\right)$$

$$+ \left(0.4 + 0.9 \sin\left(\frac{\pi t}{60}\right)^2\right) \cos\left(12t + \sin\left(\frac{\pi t}{6}\right)\right)$$

$$+ 1.2 \chi_{[15,60]}(t) \cos\left(17t + \frac{(t-35)^3}{800}\right).$$

Figure 7 shows that the results are of a quality similar to those in figure 5.

(f) ConceFT with short-time Fourier transform

As described earlier, the ConceFT approach can be carried out for STFT-based SST as well as for CWT-based SST. Figure S.1 already showed the results of STFT ConceFT on one example. Other
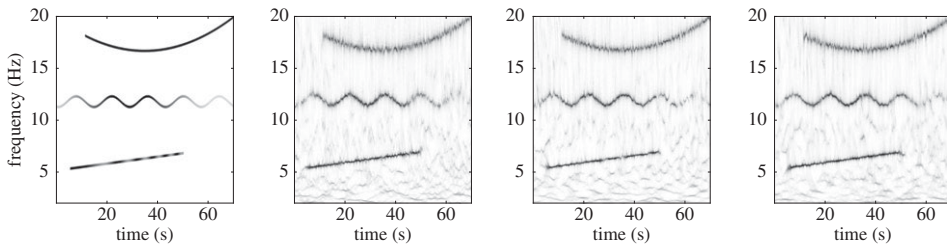
Figure 7. Results for the three-component deterministic signal s°. Left: ideal time-varying TF power spectrum (itvPS) for the clean signal, followed by results of ConceFT with Morse wavelets after (in order) Gaussian, ARMA(1,1) or Poisson noise was added, with SNR of 0 dB.

examples are shown in the electronic supplementary material, together with values of the OT distance of the STFT ConceFT estimated tvPS to the itvPS, and other discussions.

5. Conclusion

We consider signals that are the linear combination of a small number of 'intrinsic-mode functions', each of which can be reasonably viewed as an oscillatory function with well-defined but time-varying amplitude and 'instantaneous frequency'. We have introduced a new approach, called ConceFT, to determine the time–frequency representation of such signals, combining multi-taper estimation ideas and averaging over random projections with synchrosqueezing. Numerical results show that this leads to improved estimation of the time-varying characteristics of the signals of interest, even when the signals are corrupted by significant and challenging noise.

We also introduced two tools to evaluate the effectiveness of this method (or other similar methods), which may be of interest in their own right to others working in the TF field. On the one hand, we introduced a class of explicit, easy-to-construct signals with explicit time-varying characteristics, even though the signals themselves are not given by explicit formulae; the explicit time-varying amplitude and instantaneous frequency give a 'ground truth' with which estimations can be compared. On the other hand, we introduced a distance between time–frequency representations that can be useful in comparing results obtained by different methods, by computing for each the distance to the 'ground-truth' time–frequency representation.

References

1. Golombek D, Rosenstein R. 2010 Physiology of circadian entrainment. *Physiol. Rev.* **90**, 1063–1102. (doi:10.1152/physrev.00009.2009)
2. Takeda N, Maemura K. 2011 Circadian clock and cardiovascular disease. *J. Cardiol.* **57**, 249–256. (doi:10.1016/j.jjcc.2011.02.006)
3. Wang X. 2010 Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol. Rev.* **90**, 1195–1268. (doi:10.1152/physrev.00035.2008)
4. Lewis GF, Furman SA, McCool MF, Porges SW. 2012 Statistical strategies to quantify respiratory sinus arrhythmia: are commonly used metrics equivalent? *Biol. Psychol.* **89**, 349–364. (doi:10.1016/j.biopsycho.2011.11.009)

5. Lin YT, Wu HT, Tsao J, Yien HW, Hseu SS. 2014 Time-varying spectral analysis revealing differential effects of sevoflurane anaesthesia: non-rhythmic-to-rhythmic ratio. *Acta Anaesthesiol. Scand.* **58**, 157–167. (doi:10.1111/aas.12251)

6. Wu HT, Hseu SS, Bien MY, Kou YR, Daubechies I. 2013 Evaluating physiological dynamics via synchrosqueezing: prediction of ventilator weaning. *IEEE Trans. Biomed. Eng.* **61**, 736–744. (doi:10.1109/TBME.2013.2288497)

7. Baudin F, Wu HT, Bordessoule A, Beck J, Jouvet P, Frasch M, Emeriaud G. 2014 Impact of ventilatory modes on the breathing variability in mechanically ventilated infants. *Front. Pediatr. Sect. Neonatol.* **2**, 132. (doi:10.3389/fped.2014.00132)

8. Flandrin P. 1999 *Time–frequency/time-scale analysis*. San Diego, CA: Academic Press.

9. Thakur G, Wu HT. 2011 Synchrosqueezing-based recovery of instantaneous frequency from nonuniform samples. *SIAM J. Math. Anal.* **43**, 2078–2095. (doi:10.1137/100798818)

10. Huang NE, Shen Z, Long SR, Wu M, Shih H, Zheng Q, Yen NC, Tung CC, Liu HH. 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A* **454**, 903–995. (doi:10.1098/rspa.1998.0193)

11. Wu Z, Huang NE. 2009 Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **1**, 1–41. (doi:10.1142/S1793536909000047)

12. Tavallali P, Hou T, Shi Z. 2014 Extraction of intrawave signals using the sparse time–frequency representation method. *Multiscale Model. Simul.* **12**, 1458–1493. (doi:10.1137/140957767)

13. Lin L, Wang Y, Zhou H. 2009 Iterative filtering as an alternative for empirical mode decomposition. *Adv. Adapt. Data Anal.* **1**, 543–560. (doi:10.1142/S179353690900028X)

14. Huang C, Wang Y, Yang L. 2009 Convergence of a convolution-filtering-based algorithm for empirical mode decomposition. *Adv. Adapt. Data Anal.* **1**, 561–571. (doi:10.1142/S1793536909000205)

15. Chui CK, Mhaskar H. 2016 Signal decomposition and analysis via extraction of frequencies. *Appl. Comput. Harmon. Anal.* **40**, 97–136. (doi:10.1016/j.acha.2015.01.003)

16. Galiano G, Velasco J. 2014 On a non-local spectrogram for denoising one-dimensional signals. *Appl. Math. Comput.* **244**, 1–13. (doi:10.1016/j.amc.2014.06.092)

17. De Livera AM, Hyndman RJ, Snyder RD. 2011 Forecasting time series with complex seasonal patterns using exponential smoothing. *J. Am. Stat. Assoc.* **106**, 1513–1527. (doi:10.1198/jasa.2011.tm09771)

18. Daubechies I, Maes S. 1996 A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models. In *Wavelets in medicine and biology* (eds A Aldroubi, M Unser), pp. 527–546. Boca Raton, FL: CRC Press.

19. Daubechies I, Lu J, Wu HT. 2011 Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. *Appl. Comput. Harmon. Anal.* **30**, 243–261. (doi:10.1016/j.acha.2010.08.002)

20. Wu HT, Flandrin P, Daubechies I. 2011 One or two frequencies? The synchrosqueezing answers. *Adv. Adapt. Data Anal.* **3**, 29–39. (doi:10.1142/S179353691100074X)

21. Wu HT. 2011 Adaptive analysis of complex data sets. PhD thesis, Princeton University. See https://sites.google.com/site/hautiengwu/home.

22. Kodera K, Gendrin R, Villedary C. 1978 Analysis of time-varying signals with small *BT* values. *IEEE Trans. Acoust. Speech, Signal Process.* **26**, 64–76. (doi:10.1109/TASSP.1978.1163047)

23. Auger F, Flandrin P. 1995 Improving the readability of time–frequency and time-scale representations by the reassignment method. *IEEE Trans. Signal Process.* **43**, 1068–1089. (doi:10.1109/78.382394)

24. Chassande-Mottin E, Auger F, Flandrin P. 2003 Time–frequency/time-scale reassignment. In *Wavelets and signal processing* (ed. L Debnath), pp. 233–267. Applied and Numerical Harmonic Analysis Series. Cham, Switzerland: Springer International. (doi:10.1007/978-1-4612-0025-3_8)

25. Auger F, Chassande-Mottin E, Flandrin P. 2012 Making reassignment adjustable: the Levenberg–Marquardt approach. In *2012 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March*, pp. 3889–3892. Piscataway, NJ: IEEE. (doi:10.1109/ICASSP.2012.6288767)

26. Thakur G. 2014 The synchrosqueezing transform for instantaneous spectral analysis. In *Excursions in harmonic analysis*, vol. 4 (eds R Balan, M Begué, JJ Benedetto, W Czaja, KA Okoudjou), pp. 397–406. Applied and Numerical Harmonic Analysis Series. Cham, Switzerland: Springer International. (doi:10.1007/978-3-319-20188-7_15)

27. Chui CK, Lin YT, Wu HT. In press. Real-time dynamics acquisition from irregular samples—with application to anesthesia evaluation. *Anal. Appl.* (doi:10.1142/S0219530515500165)

28. Chen YC, Cheng MY, Wu HT. 2014 Nonparametric and adaptive modeling of dynamic seasonality and trend with heteroscedastic and dependent errors. *J. R. Stat. Soc. B* **76**, 651–682. (doi:10.1111/rssb.12039)

29. Maes S. 1995 Synchrosqueezed representation yields a new reading of the wavelet transform. In *Wavelet Applications II, Orlando, FL, 17–21 April*, Proc. SPIE, vol. 2491, p. 532. Bellingham, WA: SPIE. (doi:10.1117/12.205417)

30. Wu HT. 2013 Instantaneous frequency and wave shape functions (I). *Appl. Comput. Harmon. Anal.* **35**, 181–199. (doi:10.1016/j.acha.2012.08.008)

31. Meignen S, Oberlin T, McLaughlin S. 2012 A new algorithm for multicomponent signals analysis based on synchrosqueezing: with an application to signal sampling and denoising. *IEEE Trans. Signal Process.* **60**, 5787–5798. (doi:10.1109/TSP.2012.2212891)

32. Lin YT, Flandrin P, Wu HT. In press. When interpolation-induced reflection artifact meets time–frequency analysis. *IEEE Trans. Biomed. Eng*. (doi:10.1109/TBME.2015.2510580)

33. Thakur G, Brevdo E, Fuckar NS, Wu HT. 2013 The synchrosqueezing algorithm for time-varying spectral analysis: robustness properties and new paleoclimate applications. *Signal Process.* **93**, 1079–1094. (doi:10.1016/j.sigpro.2012.11.029)

34. Li C, Liang M. 2012 A generalized synchrosqueezing transform for enhancing signal time–frequency representation. *Signal Process.* **92**, 2264–2274. (doi:10.1016/j.sigpro.2012.02.019)

35. Yang H. 2015 Synchrosqueezed wave packet transforms and diffeomorphism based spectral analysis for 1D general mode decompositions. *Appl. Comput. Harmon. Anal.* **39**, 33–66. (doi:10.1016/j.acha.2014.08.004)

36. Oberlin T, Meignen S, Perrier V. 2015 Second-order synchrosqueezing transform or invertible reassignment? Towards ideal time–frequency representations. *IEEE Trans. Signal Process.* **63**, 1335–1344. (doi:10.1109/TSP.2015.2391077)

37. Xi S, Cao H, Chen X, Zhang X, Jin X. 2015 A frequency-shift synchrosqueezing method for instantaneous speed estimation of rotating machinery. *ASME. J. Manuf. Sci. Eng.* **137**, 031012. (doi:10.1115/1.4029824)

38. Huang Z, Zhang J, Zhao T, Sun Y. 2016 Synchrosqueezing *S*-transform and its application in seismic spectral decomposition. *Geosci. Remote Sensing, IEEE Trans.* **54**, 817–825. (doi:10.1109/TGRS.2015.2466660)

39. Yang H, Ying L. 2013 Synchrosqueezed wave packet transform for 2D mode decomposition. *SIAM J. Imaging Sci.* **6**, 1979–2009. (doi:10.1137/120891113)

40. Yang H, Ying L. 2014 Synchrosqueezed curvelet transform for two-dimensional mode decomposition. *SIAM J. Math. Anal.* **46**, 2052–2083. (doi:10.1137/130939912)

41. Iatsenko D, Bernjak A, Stankovski T, Shiogai Y, Owen-Lynch PJ, Clarkson PBM, McClintock PVE, Stefanovska A. 2013 Evolution of cardiorespiratory interactions with age. *Phil. Trans. R. Soc. A* **371**, 20110622. (doi:10.1098/rsta.2011.0622)

42. Wu HT, Chan YH, Lin YT, Yeh YH. 2014 Using synchrosqueezing transform to discover breathing dynamics from ECG signals. *Appl. Comput. Harmon. Anal.* **36**, 354–359. (doi:10.1016/j.acha.2013.07.003)

43. Wu HT, Talmon R, Lo YL. 2015 Assess sleep stage by modern signal processing techniques. *IEEE Trans. Biomed. Eng.* **62**, 1159–1168. (doi:10.1109/TBME.2014.2375292)

44. Lin Y-T. 2015 The modeling and quantification of rhythmic to non-rhythmic phenomenon in electrocardiography during anesthesia. PhD thesis, National Taiwan University. (http://arxiv.org/abs/1502.02764)

45. Li C, Liang M. 2012 Time–frequency signal analysis for gearbox fault diagnosis using a generalized synchrosqueezing transform. *Mech. Syst. Signal Process.* **26**, 205–217. (doi:10.1016/j.ymssp.2011.07.001)

46. Feng Z, Chen X, Liang M. 2015 Iterative generalized synchrosqueezing transform for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions. *Mech. Syst. Signal Process.* **52–53**, 360–375. (doi:10.1016/j.ymssp.2014.07.009)

47. Guharay S, Thakur G, Goodman F, Rosen S, Houser D. 2013 Analysis of non-stationary dynamics in the financial system. *Econ. Lett.* **121**, 454–457. (doi:10.1016/j.econlet.2013.09.026)

48. Vatter T, Wu HT, Chavez-Demoulin V, Yu B. 2015 Non-parametric estimation of intraday spot volatility: disentangling instantaneous trend and seasonality. *Econometrics* **3** (4), 864–887. (doi:10.3390/econometrics3040864)

49. Wang P, Gao J, Wang Z. 2014 Time–frequency analysis of seismic data using synchrosqueezing transform. *Geosci. Remote Sensing Lett. IEEE* **11**, 2042–2044. (doi:10.1109/LGRS.2014.2317578)

50. Herrera RH, Han J, van der Baan M. 2014 Applications of the synchrosqueezing transform in seismic time–frequency analysis. *Geophysics* **79**, V55–V64. (doi:10.1190/geo2013-0204.1)

51. Tary JB, Herrera RH, Han J, van der Baan M. 2014 Spectral estimation—What is new? What is next? *Rev. Geophys.* **52**, 723–749. (doi:10.1002/2014RG000461)

52. Li PC, Sheu YL, Laughlin C, Chu SI. 2013 Role of laser-driven electron-multirescattering in resonance-enhanced below-threshold harmonic generation in He atoms. *Phy. Rev. A* **90**, 041401(R). (doi:10.1103/PhysRevA.90.041401)

53. Sheu YL, Hsu LY, Wu HT, Li PC, Chu SI. 2014 A new time–frequency method to reveal quantum dynamics of atomic hydrogen in intense laser pulses: synchrosqueezing transform. *AIP Adv.* **4**, 117138. (doi:10.1063/1.4903164)

54. Li PC, Sheu YL, Laughlin C, Chu SI. 2015 Dynamical origin of near- and below-threshold harmonic generation of Cs in an intense mid-infrared laser field. *Nat. Commun.* **6**, 7178. (doi:10.1038/ncomms8178)

55. Yang H, Lu J, Ying L. 2015 Crystal image analysis using 2D synchrosqueezed transforms. *Multiscale Model. Simul.* **13** (4), 1542–1572. (doi:10.1137/140955872)

56. Yang H, Lu J, Brown WP, Daubechies I, Ying L. 2015 Quantitative canvas weave analysis using 2-D synchrosqueezed transforms: application of time–frequency analysis to art investigation. *Signal Process. Mag., IEEE* **32** (4), 55–63. (doi:10.1109/MSP.2015.2406882)

57. Thomson DJ. 1982 Spectrum estimation and harmonic analysis. *Proc. IEEE* **70**, 1055–1096. (doi:10.1109/PROC.1982.12433)

58. Percival DB, Walden AT. 1993 *Spectral analysis for physical applications: multitaper and conventional univariate techniques*. Cambridge, UK: Cambridge University Press.

59. Babadi B, Brown EN. 2014 A review of multitaper spectral analysis. *IEEE Trans. Biomed. Eng.* **61**, 1555–1564. (doi:10.1109/TBME.2014.2311996)

60. Fraser G, Boashash B. 1994 Multiple window spectrogram and time–frequency distributions. In *1994 IEEE Int. Conf. on Acoustics, Speech and Signal Processing* (ICASSP), *Adelaide, Australia, 19–22 April*, vol. 4, IV/293–IV/296. Piscataway, NJ: IEEE. (doi:10.1109/ICASSP.1994.389818)

61. Farry KA, Baraniuk RG, Walker ID. 1995 Nonparametric, low bias, and low variance time–frequency analysis of myoelectric signals. In *1995 IEEE 17th Annu. Conf. Engineering in Medicine and Biology Society, Montreal, Canada, 20–23 September*, vol. 2, pp. 993–994. Piscataway, NJ: IEEE. (doi:10.1109/IEMBS.1995.579397)

62. Bayram M, Baraniuk RG. 1996 Multiple window time–frequency and time-scale analysis. In *Advanced Signal Processing Algorithms, Architectures, and Implementations VI, Denver, CO, 4 August*, Proc. SPIE, vol. 2846, p. 174. Bellingham, WA: SPIE. (doi:10.1117/12.255431)

63. Xu Y, Haykin S, Racine RJ. 1999 Multiple window time–frequency distribution and coherence of EEG using Slepian sequences and Hermite functions. *IEEE Trans. Biomed. Eng.* **46**, 861–866. (doi:10.1109/10.771197)

64. Xiao J, Flandrin P. 2007 Multitaper time–frequency reassignment for nonstationary spectrum estimation and chirp enhancement. *IEEE Trans. Signal Process.* **55**, 2851–2860. (doi:10.1109/TSP.2007.893961)

65. Lin YT, Hseu SS, Yien HW, Tsao J. 2011 Analyzing autonomic activity in electrocardiography about general anesthesia by spectrogram with multitaper time–frequency reassignment. In *2011 4th Int. Conf. on Biomedical Engineering and Informatics* (BMEI), *Shanghai, China, 15–17 October,* vol. 2, pp. 630–634. Piscataway, NJ: IEEE. (doi:10.1109/BMEI.2011.6098432)

66. Orini M, Bailón R, Mainardi LT, Laguna P, Flandrin P. 2012 Characterization of dynamic interactions between cardiovascular signals by time–frequency coherence. *IEEE Trans. Biomed. Eng.* **59**, 663–673. (doi:10.1109/TBME.2011.2171959)

67. Daubechies I. 1992 *Ten lectures on wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics, no. 61. Philadelphia, PA: SIAM. (doi:10.1137/1.9781611970104)

68. Kowalski M, Meynard A, Wu HT. 2015 Convex optimization approach to signals with fast varying instantaneous frequency. (http://arxiv.org/abs/1503.07591)

69. Gallager R. 2008 Circularly-symmetric Gaussian random vectors. See http://www.rle.mit.edu/rgallager/documents/CircSymGauss.pdf.
70. Daubechies I. 1988 Time–frequency localization operators: a geometric phase space approach. *IEEE Trans. Inform. Theory* **34**, 605–612. (doi:10.1109/18.9761)
71. Daubechies I, Paul T. 1988 Time–frequency localization operators: a geometric phase space approach: II. The use of dilations. *Inverse Probl.* **4**, 661–680. (doi:10.1088/0266-5611/4/3/009)
72. Olhede SC, Walden AT. 2002 Generalized morse wavelets. *IEEE Trans. Signal Process.* **50**, 2661–2671. (doi:10.1109/TSP.2002.804066)
73. Lin YT, Wu HT. Submitted. ConceFT for time-varying heart rate variability analysis as a measure of noxious stimulation during general anesthesia.

19

rsta.royalsocietypublishing.org *Phil. Trans. R. Soc. A* **374**: 20150193