Identifying Ocean Swell Generation Events from Ross Ice Shelf Seismic Data𝒪

MOMME C. HELL, BRUCE D. CORNUELLE, SARAH T. GILLE, ARTHUR J. MILLER,
AND PETER D. BROMIRSKI

Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California

(Manuscript received 7 June 2019, in final form 1 October 2019)

ABSTRACT

Strong surface winds under extratropical cyclones exert intense surface stresses on the ocean that lead to upper-ocean mixing, intensified heat fluxes, and the generation of waves, that, over time, lead to swell waves (longer than 10-s period) that travel long distances. Because low-frequency swell propagates faster than high-frequency swell, the frequency dependence of swell arrival times at a measurement site can be used to infer the distance and time that the wave has traveled from its generation site. This study presents a methodology that employs spectrograms of ocean swell from point observations on the Ross Ice Shelf (RIS) to verify the position of high wind speed areas over the Southern Ocean, and therefore of extratropical cyclones. The focus here is on the implementation and robustness of the methodology in order to lay the groundwork for future broad application to verify Southern Ocean storm positions from atmospheric reanalysis data. The method developed here combines linear swell dispersion with a parametric wave model to construct a time- and frequency-dependent model of the dispersed swell arrivals in spectrograms of seismic observations on the RIS. A two-step optimization procedure (deep learning) of gradient descent and Monte Carlo sampling allows detailed estimates of the parameter distributions, with robust estimates of swell origins. Median uncertainties of swell source locations are 110 km in radial distance and 2 h in time. The uncertainties are derived from RIS observations and the model, rather than an assumed distribution. This method is an example of supervised machine learning informed by physical first principles in order to facilitate parameter interpretation in the physical domain.

1. Introduction

Strong winds associated with extratropical cyclones act on the ocean surface and generate surface gravity waves. These waves propagate long distances and are observed as swell (Snodgrass et al. 1966). Long swell waves (in the range between 0.03 and 0.8 Hz) can travel across ocean basins with minimal attenuation (Snodgrass et al. 1966). Because wave dispersion depends on frequency, swell observed at distant locations contains information about its position and time of generation. The idea of tracking storms using swell was first shown by Munk (1947) and by Barber and Ursell (1948). The objective of this study is to establish a methodology to use modern swell observations to learn about conditions at the swell's source region as well as the travel path of the swell.

The locations where swell waves originate experience intense atmosphere–ocean interaction. Some of the strongest events occur in the Southern Ocean, where the observing system is sparse and storm systems are not well characterized by direct observation (e.g., Bourassa et al. 2013). Strong surface winds lead to intense air–sea fluxes of heat, momentum and CO_2, with potential implications for ocean circulation changes and the ocean uptake of anthropogenic heat and CO_2 in the Southern Ocean (SO; Swart et al. 2018; Rintoul 2018; Marshall et al. 2016; Munday and Zhai 2017; Gruber et al. 2019).

Ocean swell spectra are routinely generated from autonomous wave buoy observations, GPS sensors, or seafloor pressure sensors (Munk et al. 1963; Collard et al. 2009; Delpey et al. 2010; O'Reilly et al. 2016), and they have also been observed by land-based seismic stations when swell interacts with the coast (Bromirski et al. 1999). The time series of swell arrivals at an observation site can be converted to a time evolving power spectrum, known as a spectrogram. Each set of swell arrivals detected in a spectrogram is related to the surface wind at the storm (Pierson and Moskowitz 1964; Hasselmann et al. 1973; Elfouhaily et al. 1997) and can,

as we will show, be interpreted as a *remote* observation of the storm itself.

This study adopts a unique approach by using seismic data collected not on land, but instead on a floating ice shelf, as part of the Ross Ice Shelf (RIS) Vibration Project (Wiens et al. 2014; Bromirski et al. 2017). The data from the RIS allow us to compute an extensive series of high-resolution spectrograms of surface gravity waves, similar to conventional wave observations. We use these data as a training set to develop a new method to characterize ocean swell observations.

Feature comparison in geophysical data is often challenging because the observations are noisy, and the models are too simple. As we outline below, the combination of optimization and Monte Carlo methods enables us to improve our model understanding of the data, while we use the model to identify the relevant data. This is a ''machine learning'' approach that is constrained by physical laws, with the benefit that it generates uncertainties based on the data and the model, rather than assuming an a priori uncertainty distribution (Marone 2018).

We present a method to compare characteristic patterns in seismic spectrograms with a parametric model that is constrained by the physics of ocean gravity waves. We first briefly describe the physical background that motivates the model (section 2) and introduce the dataset (section 3). Then, we introduce the governing cost function (section 4), the model (section 5), and the data preparation (section 6). The actual fitting procedure is explained in section 7, and its performance is shown in section 8 and discussed in section 9. The developed code for this analysis will be publicly available in a github repository (https://github.com/mochell/stormy_forerunners) after completing the project.

2. Waves across the Pacific: Physical background

Observations on the RIS record storm-induced swell events (Fig. 1), much like previously reported observations along coastlines (Munk and Snodgrass 1957; Snodgrass et al. 1966). The gestalt of these coherent packages of swell energy is shaped by three processes:

1) The dispersion of deep water waves means longer waves travel faster, such that the longest wave generated by a storm arrives first (Munk 1947; Barber and Ursell 1948; Snodgrass et al. 1966; Gallet and Young 2014). At any point in the ocean, an observer who records the arrival time of waves of different frequencies can estimate both the time of origin and the distance traveled, assuming all waves come from the same source. For continuous observations, like those provided by the RIS seismometers, the succession of wave arrivals results in a sloped line in the wave spectra (Fig. 1). The sloped line of these dispersed wave events is an indirect measure of the radial distance to the origin of the waves.

2) The spectrogram and its shape are related to winds in the wave generation region. There is extensive literature about ocean wave spectra. See, for example, the compendial overviews of Massel (1996, chapter 3.2) or Elfouhaily et al. (1997). The most commonly used parametric models are the Pierson–Moskowitz (hereafter PM) spectrum for a fully developed sea or the Joint North Sea Wave Project (JONSWAP) spectrum (Fig. 2, Pierson and Moskowitz 1964; Phillips 1985; Hasselmann et al. 1973, 1976). Both models are possible functional forms for this optimization problem. The advantage of the JONSWAP spectrum is that it is more flexible and is not limited to fully developed seas. It also relates the peak frequency f_m and the amplitude parameter α to the nondimensional fetch $\tilde{X}$

$$f_m = \frac{3.5g}{U_{10}} \tilde{X}^{-0.33}, \qquad (1)$$

$$\alpha = 0.033 \left(\frac{f_m U_{10}}{g} \right)^{0.67} = 0.076\, \tilde{X}^{-0.22}, \qquad (2)$$

with

$$\tilde{X} = \frac{gX}{U_{10}^2}, \qquad (3)$$

where X is the fetch in meters (defined as the horizontal distance over which wave-generating winds are able to act), U_{10} is the 10-m wind speed over that area, and g is the acceleration due to Earth's gravity (Hasselmann et al. 1973). The JONSWAP relations can be inverted to infer speed and fetch at the location of the storm from the wave spectrum parameters α and f_m, detected at a remote location (Pierson and Moskowitz 1964; Hasselmann et al. 1976). Both PM and JONSWAP spectra are based on theories of wave generation by winds (Phillips 1957; Miles 1957, 1960), and their only difference stems from the JONSWAP model's inclusion of additional parameters that vanish under the assumption of a fully developed sea (see section 5, Fig. 2, and Massel 1996, chapter 3.2).

3) When swell travels into sea ice, it can be damped or reflected (Fox and Squire 1994; Squire 2007; Vaughan et al. 2009). While low-frequency swell waves travel through sea ice and are detected in seismic records on the RIS (typical periods of about 15 s, Fig. 1 shading; Collard et al. 2009; Cathles et al. 2009; Bromirski et al. 2010; MacAyeal et al. 2009), higher frequencies are
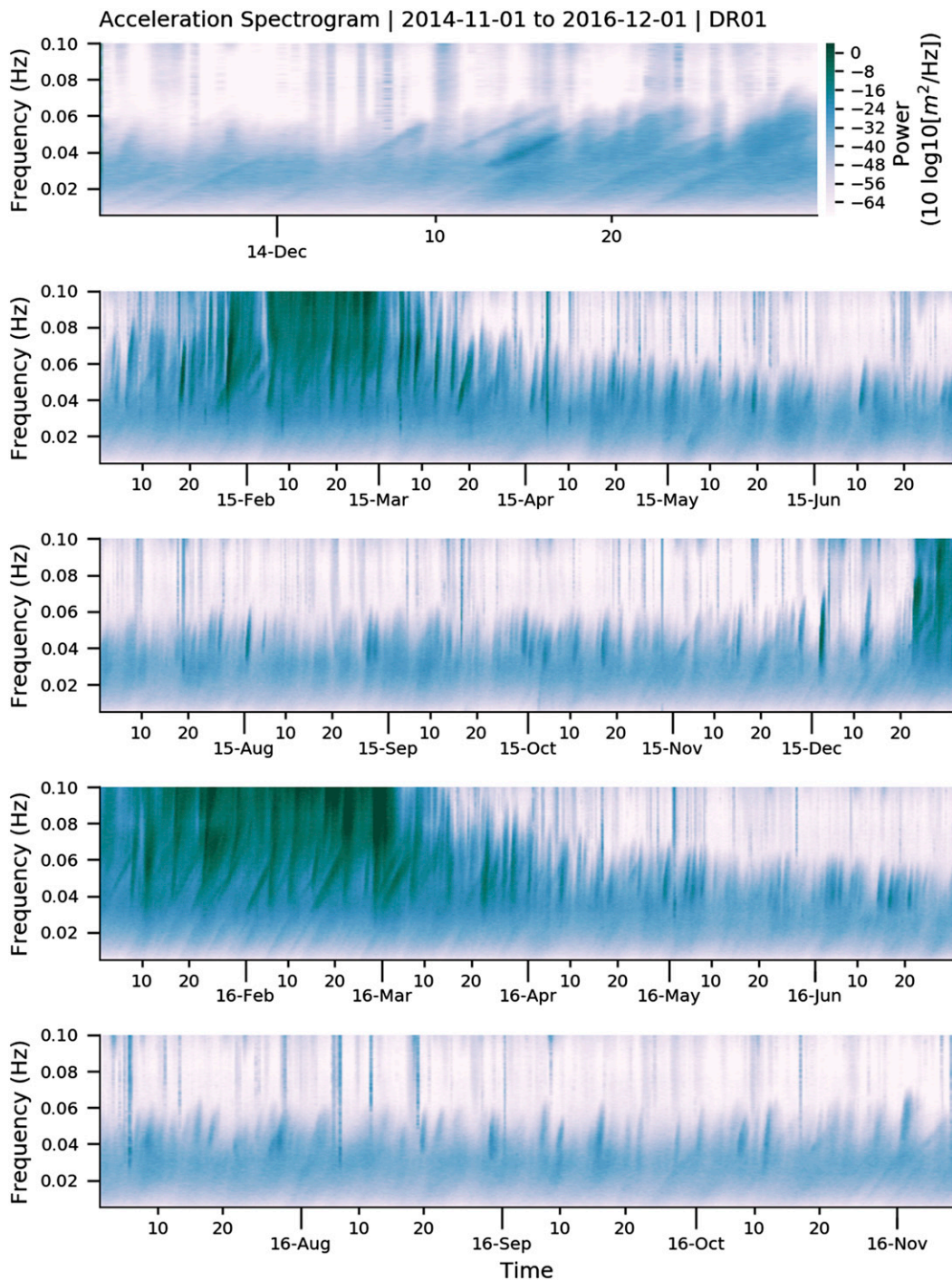
FIG. 1. Spectrogram of the vertical acceleration in DR01 between November 2014 and December 2016. The spectrogram is expressed as a power spectra with a basic segment length of 20 min. The spectral estimate at each time step is sampled from 12 of these segments.

strongly damped (periods of about 10 s and shorter; Kohout et al. 2014; Collins et al. 2015; Ardhuin et al. 2016). It is hypothesized that damping of incident swell energy by sea ice helps to maintain the overall

ice shelf stability (Squire et al. 1994; MacAyeal et al. 2006; Robinson and Haskell 1990; Lipovsky 2018; Massom et al. 2018). However, a validated parametric model of sea ice induced damping does not exist and
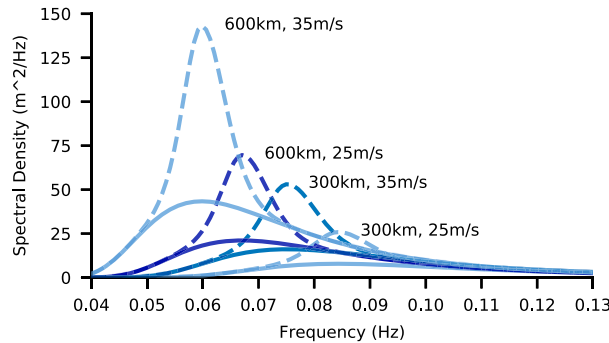
FIG. 2. Pierson–Moskoviz (solid lines) and JONSWAP spectra (dashed) for a variety of fetches and wind speeds. The fetch length χ and wind speed U_{10} are indicated [see (1) and (3)].

is therefore not included in our formulation of the model (section 5).

The first two processes, wave dispersion and spectral shape, are used to construct our model for the optimization procedure. The model is constrained by prior physical knowledge about the processes that we aim to investigate. If the residual differences between model and data share common features between multiple events, then the misfit can potentially be attributed to physical processes that are not represented in the model, such as the attenuation due to sea ice.

Most ocean swell observations show a superposition of locally and remotely generated waves (Rapizo et al. 2015; O'Reilly et al. 2016). In contrast, the swell spectra observed on the RIS are only due to remotely generated waves because swell generation in the proximity of the ice shelf is suppressed by sea ice. Even in summer, when melting may produce open water areas close to the RIS, any locally generated waves are shorter than the remotely generated swell. Wave generation at the observation site is not possible because the observations are made on the ice shelf rather than in the ocean. The fact that RIS data highlight the impacts from remote storm activity in the Southern Ocean makes them unique.

On the other hand, RIS seismic records may be the result of processes that are absent in open ocean observations, such as interactions with sea ice, topography and currents, or the ice shelf itself. The intent of this discussion of the method is to first set these additional complexities aside and, in a second step, assess whether RIS-specific processes can explain the deviation of the observations compared to the model function. This model function represents a physical hypothesis for the evolving ocean wave spectra; however, we do not expect it to apply exactly in each individual case.
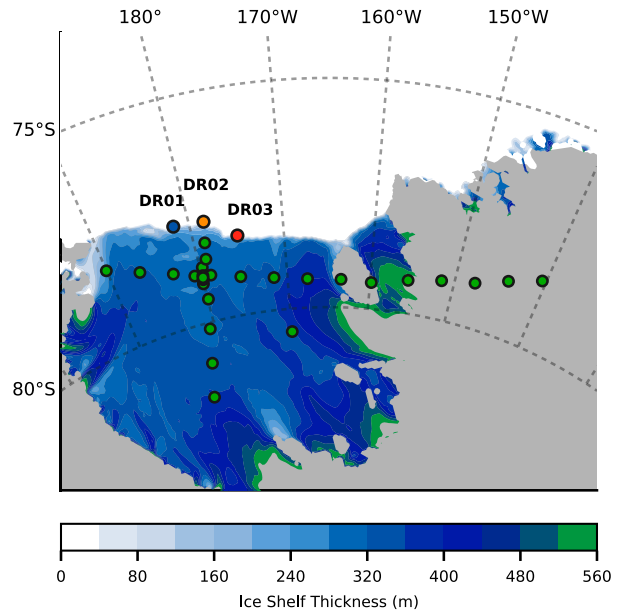

FIG. 3. Map of the Ross Ice Shelf Vibration Project. The position of the front Stations DR01, DR02, and DR03 are indicated by the blue, orange, and red dots. Other seismic stations are shown as green dots. The ice-shelf thickness is shown as shading. The ice-shelf edge migrated northward since the ice-shelf thickness was derived (Haran et al. 2005; Haran and Bohlander 2014). The front stations were about 2 km away from the ice-shelf front when the data was recorded. The ice shelf thickness data are taken from Basemap2 (Fretwell et al. 2013).

3. Seismic observations in the Ross Ice Shelf

The Ross Ice Shelf Vibration Project was a field campaign carried out from October 2014 to December 2016 with the goal of recording the Ross Ice Shelf response to gravity wave impacts for geophysical, glaciological and oceanographic purposes (Wiens et al. 2014). To investigate the RIS response to gravity wave forcing, a network of 28 seismic stations recorded 2 years of continuous vertical and horizontal displacements at each station (Fig. 3). The sampling rate was either 100 or 200 Hz depending on the instrument configurations at each station (Bromirski et al. 2015). Data were archived in accordance with IRIS (Incorporated Research Institutions for Seismology; www.iris.edu) standards for seismic data. The three stations closest to the front (DR01, DR02, and DR03) recorded the highest amplitude response for swell waves, and are thus used for the analysis presented in this paper, because they are expected to have the largest signal-to-noise ratios for swell waves.

The processing is as follows. First, the 100- or 200-Hz time series are averaged to 1 Hz, because the time scales of interest (waves with frequencies less than 0.1 Hz) are perfectly resolved by 1-Hz sampling, and the much smaller data volume makes processing more efficient.

Second, the recorded time series are corrected for the frequency-dependent response function of the seismometer. The 1-Hz time series is deconvolved with the instrument response function, which is a cosine window ($f_1 = 10^{-4}\,\text{Hz}, f_2 = 2 \times 10^{-4}\,\text{Hz}, f_3 = 0.4\,\text{Hz}, f_4 = 0.5\,\text{Hz}$). The resulting displacement time series are prewhitened by taking the second derivative in time to generate a time series of acceleration in meters per second squared (m s^{-2}). Peaks in the acceleration time series are removed if they deviate from the mean by more than 10 standard deviations; any resulting gaps are filled by linear interpolation. Less than 0.1% of the data are removed from the time series, and the removals have no effect on the model estimate.

After these preliminary adjustments, a spectrogram is calculated at each station using the 1-Hz averaged time series for a shifting window discrete Fourier transformation. To calculate spectrograms, data are first split into segments of 20-min duration, with 50% overlap. Each segment is detrended and fast Fourier transformed to produce periodograms. Spectral estimates are computed at hourly time increments, by averaging periodograms from eleven 20-min segments centered around 1-h time steps (i.e., spanning a total time period of 120 min), while a Hanning window was applied to each segment. The 20-min segment length determines the frequency resolution with a lowest frequency of 1/1200 Hz. The resulting 2-yr spectrogram for DR01 is shown in Fig. 1. (Spectrograms for DR02 and DR03 appear indistinguishable from DR01.)

4. Cost function definition

Our next objective is to optimally fit the swell arrivals detected in the spectrograms to a model based on the JONSWAP spectrum by adjusting the free parameters in the model. We do this via a nonlinear minimization method performed on a global cost function

$$\Phi = J_d + J_m, \qquad (4)$$

which is the sum of the data cost function J_d and the model cost function J_m (known as ridge or lasso regularization). The data cost function J_d is the sum of the squared difference between data **D** and model **M(p)**, with **p** being the model parameter, multiplied by the weight function **w** at each point

$$J_d = \sum_{i,j} [(D_{i,j} - M_{i,j})\, w_{i,j}]^2. \qquad (5)$$

The model cost function J_m is the sum of squares of the normalized parameter values

$$J_m = \sum_k \left(\frac{p_0^i - p^i}{p_\sigma^i}\right)^2, \qquad (6)$$

where **p** is a set of function parameters for optimization, $\mathbf{p}_0$ represents the initial guesses of the parameter vector, and $\mathbf{p}_\sigma$ is the corresponding prior error estimate (see the appendix).

The model cost function allows us to optimize function parameters **p**, while taking account of prior estimates of uncertainty in the parameters $\mathbf{p}_\sigma$. To allow a wider range of parameter values, the prior uncertainty is artificially set to be large. Too small values for $\mathbf{p}_\sigma$ result in an overweighting of the costs due to the parameters, resulting in overly conservative model behavior that is more likely to remain close to the initial conditions. The following sections explain the parametric model (section 5), and the data preparation and weight function (section 6).

5. Model description

The model $\mathbf{M}_S \equiv M_S(f, t)$ is compared against the data $\mathbf{D}(f, t)$ at each iteration of the minimization procedure. The model has a time component $M_T(t)$ and a spectral component $S(f)$ that are both described here. We assume a separable model $M_s(f, t) = S(f)M_T(t)$.

1) The spectral part of the model $S(f)$ is based on open-ocean swell spectra of a fully developed sea (section 2). The JONSWAP spectrum [Hasselmann et al. 1973; Massel 1996, p. 94, Eq. (3.81)] is reformulated as

$$S(f) = \hat{\alpha}(2\pi f)^{-5} \exp\left[-\chi\left(\frac{f}{f_m}\right)^{-4}\right]\gamma^\delta, \qquad (7)$$

$$\delta = \exp\left[-\frac{1}{2}\left(\frac{f - f_m}{\sigma_0 f_m}\right)^2\right], \qquad (8)$$

where f is the frequency, $\hat{\alpha}$ the amplitude parameter in units of acceleration squared, χ the nondimensional stretching term ($\chi = 5/4$ in the standard JONSWAP model), f_m the position of the peak frequency, γ the measure of the height of the peak function, and σ_0 the width of the peak function. We define an amplitude parameter $\hat{\alpha} = g^2\alpha$ such that the first guess of $\hat{\alpha}$ is of order one, while the initial value of α is inferred from the data (see below). This model reverts to the original PM-spectrum when $\hat{\alpha}$ and χ are set to constant values taken from JONSWAP (Hasselmann et al. 1973). In total, there are five free parameters in (7) and (8): the conventional peak parameter f_m and four more parameters (χ, $\hat{\alpha}$, γ, σ_0) to allow for the additional complexity in the seismic data due to the interaction with sea ice and the RIS (section 2). Other parameters of the JONSWAP

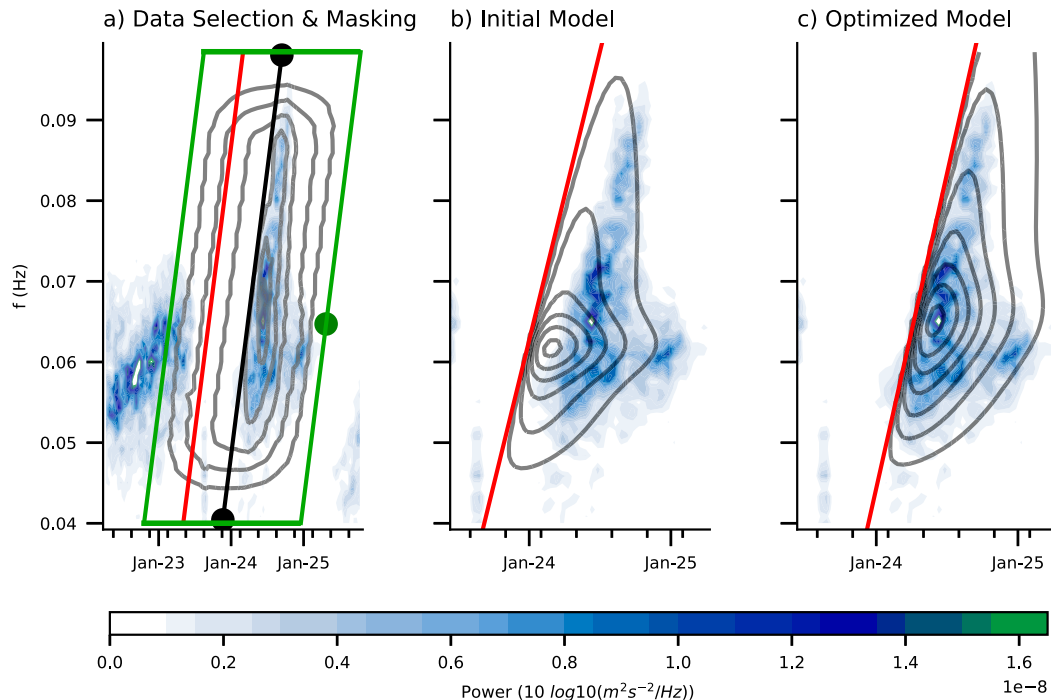**Event in Station DR01 | 2015-01-24T00**

FIG. 4. Spectrogram of a single swell event in January 2015 with the three stages of the fitting procedure. (a) Derived spectrogram (shading; section 3). Black and green dots indicate manually identified edge points of the parallelogram-shaped date mask (green line; section 6a). Red lines are used as the model initial condition, and gray contours indicate the data weighting function **w** in (15). (b) Masked data (shading; section 6a) with initial slope (red) and model (black contours). (c) As in (b), but with the optimized slope and model.

spectrum, such as U_{10} and $\tilde{X}$ are not used directly in the model (7), but can be inferred using the estimated parameters f_m and $\hat{\alpha}$ in (1)–(3).

2) In the time domain, visual inspection of the spectrogram suggests that swell arrivals generally have a relatively sharp leading edge (Fig. 4), while their decay varies (Munk and Snodgrass 1957). This behavior is approximated by the Γ distribution

$$M_T(\tilde{t}) = \frac{1}{\sigma_t \Gamma(\tilde{t})} \left(\frac{\tilde{t} - \tilde{t}_{\text{peak}}}{\sigma_t} \right)^{b-1} e^{-\tilde{t}}, \qquad (9)$$

with the Γ function as

$$\Gamma(\tilde{t}) = \int_0^\infty x^{\tilde{t}-1} e^{-x} \, dx, \qquad (10)$$

where $\tilde{t} \equiv (t - t_{\text{start}})/(t_{\text{end}} - t_{\text{start}}) = (t - t_{\text{start}})/\Delta t$ is the normalized time, with t_{start} being the lower left and t_{end} the upper right corner of the parallelogram (Fig. 4a, described in section 6a). The dimensionless parameter σ_t is set to 0.07, so that $M_T(\tilde{t})$ in (9) has a

maximum value of order one, such that the only parameter that determines the amplitude is $\hat{\alpha}$. The term $\tilde{t}_{\text{peak}}$ represents the location of the peak in time, and b controls the width of the Γ distribution; both are used for parameter optimization (section 7). Figure 5a illustrates $M_T(\tilde{t})$ for default values of $\tilde{t}_{\text{peak}}$ and b and for the maximum and minimum values for b, set as constraints for the optimization (see the appendix).

3) The time-evolving peak frequency is expressed as a linear function that is informed by the deep-water dispersion relation:

$$f_{\text{peak}}(\tilde{t}) = (\tilde{t} - \tilde{t}_0) \, m_{\tilde{t}}, \qquad (11)$$

where f_{peak} is the peak frequency, $\tilde{t}_0$ is the center of the nondimensional normalized time axis, and $m_{\tilde{t}}$ is the rate of change in units of hertz. Note that f_{peak} is different from the maximum peak frequency f_m: f_{peak} is peak frequency at each nondimensional time $\tilde{t}$, while f_m is the maximum of the peak frequencies, that is, the peak frequency over the whole event.
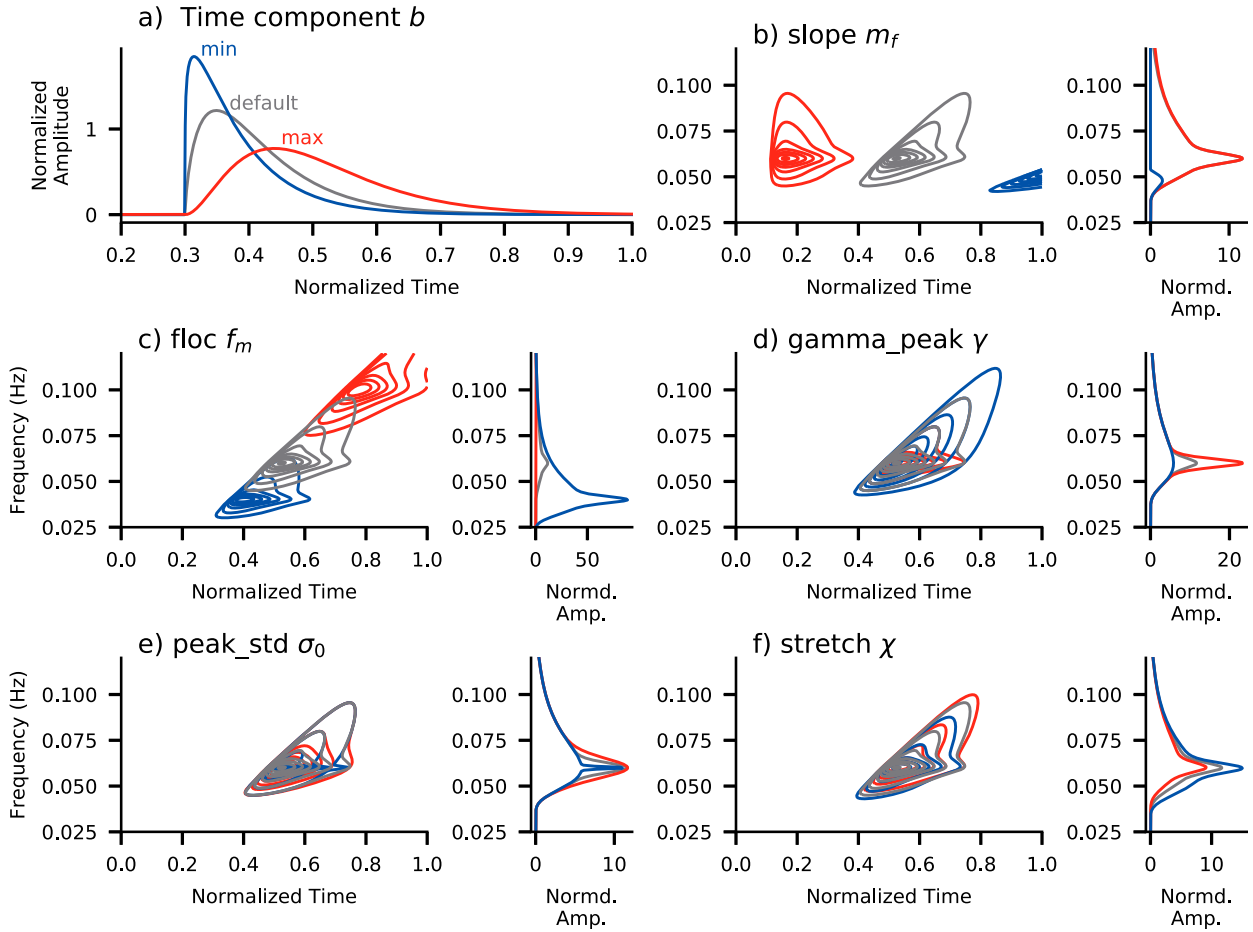
Default Model Paramters

FIG. 5. Default (gray), minimum (blue), and maximum (red) model parameters with the Γ distribution for the (a) shape parameter b, (b) the slope parameter m_f, (c) the peak frequency f_m, (d) the peak parameter γ, (e) the peak width σ_0, and (f) the stretching χ.

The rate of change $m_{\tilde{t}}$ and nondimensional initial time $\tilde{t}_0$ are directly related to the distance and time estimates (Snodgrass et al. 1966; Munk and Snodgrass 1957; Barber and Ursell 1948). The inversion of (11) gives a relation for $\tilde{t}_{\text{peak}}(f)$, which is inserted into (9).

The Γ distribution M_T, the JONSWAP spectrum S, and the linear slope equation (11) yield a two-dimensional model of swell arrivals:

$$M_S(f, \tilde{t}) = S(f)\, M_T(\tilde{t}) \tag{12}$$

$$= \hat{\alpha}\frac{2}{0.7(2\pi)^5}\frac{1}{\Gamma(\tilde{t})f^5}\gamma^\delta\left(\frac{\tilde{t} - f\,m_t^{-1} + \tilde{t}_0}{0.07}\right)^{b-1}$$

$$\times \exp\left[-\chi\left(\frac{f}{f_m}\right)^{-4} - \tilde{t}\right], \tag{13}$$

$$\text{with} \quad \delta = \exp\left[-\frac{1}{2}\left(\frac{f - f_m}{\sigma_0 f_m}\right)^2\right]. \tag{14}$$

Equation (12) has $N_{\text{var}} = 8$ fitting parameters: $\mathbf{p} = \{\hat{\alpha}, \chi, f_m, m_f, \tilde{t}_0, b, \gamma, \sigma_0\}^{\text{T}}$. These parameters are the basis for the nonlinear optimization procedure described in section 7, and the sensitivity of the model (12) to these parameters is shown in Fig. 5.

6. Prehandling the data

Achieving optimal agreement between the data and the model function requires selection and preliminary correction of the data. This section explains how events are selected and corrected to facilitate nonlinear optimization. First, the shape (section 6a) and amplitude (sections 6b and 6c) of individual events are used to ensure similar signal-to-noise levels. Subsequently the

model is fitted to the adjusted data, given a customized set of initial parameters, but without further individual tuning of the model (section 7).

a. Selection procedure and masking

The vertical acceleration spectrograms from stations DR01 to DR03 show about 250 wave events during the 2-yr RIS measurement period. (Fig. 1 shows the full record for DR01, and Figs. S2 and S3 in the online supplemental material show the same for DR02 and DR03.) These events are common features in the spectrograms of other stations across the RIS array (Fig. 3). Wave events are strongest near the ice shelf front and decay with distance toward the interior of the shelf (Bromirski et al. 2017).

Each event has a characteristic slope, indicating that low-frequency energy arrives before higher frequencies (section 2). In this analysis, the slopes are identified using an interactive hand picking procedure. An example of this is shown in Fig. 4a, in which the wave event (blue shading) is identified by its low- and high-frequency limits (black dots in Fig. 4a) and its estimated time width (green dot in Fig. 4a).

The data mask is a parallelogram defined by three values, as follows (Fig. 4a, green perimeter). The upper and lower limits are the corresponding frequencies of the black dots, and the tilted sides are twice the temporal separation between the green point and the black middle line. Initial parameters for slope and intersect (section 5) are taken from a line centered between the left boundary and the middle line (Fig. 4a, red line).

b. Data weighting

There is additional prior information about the usefulness of the data within the data mask. High amplitudes close to the mask boundary are typically attributed to noise, while data in the center of the domain are likely associated with the selected event. The geometry derived in the previous section is also used to construct a data weighting function defined as

$$\mathbf{w} = (\mathbf{w}_G + \mathbf{w}_G \circ \mathbf{w}_D)/2 + w_{\text{floor}}, \tag{15}$$

where $\mathbf{w}_G$ is a geometrical weight that decays from 1 in the center to 0 at the boundary using a Hanning window (Fig. 4a, gray contours), $\mathbf{w}_D(f, \tilde{t})$ is the spatially smoothed[1] data divided by its maximum value, such that $\mathbf{w}_D$ is a matrix that weights high-amplitude data points

more strongly, and ($\circ$) is the Schur product. The minimum is $w_{\text{floor}} = 10^{-6}$. The total weight $\mathbf{w}$ can vary between w_{floor} and $w_{\text{floor}} + 1$ and is constructed such that data points at the boundary, especially of high amplitude, are downweighted, while data points in the center with high amplitudes are upweighted. The noise floor value w_{floor} represents the general uncertainty in the data that is estimated from the uncertainty in the spectral estimate. The uncertainty of the spectral estimate is derived from subsampling described in section 3.

c. Noise handling

Within the parallelogram-shaped mask used to select data from the spectrogram, higher noise levels often occur at lower frequencies (Fig. 4a, below 0.05 Hz and in Fig. S4). Here, noise is accounted for by fitting a noise model prior to fitting the actual model. The noise model follows an exponential decay of the form

$$M_n(f) = \beta_n e^{-\tau_n f}, \tag{16}$$

where β_n and τ_n are free parameters. The difference between the masked data $\tilde{D}$ and $\mathbf{M}_n \equiv M_n(f)$ defines the noise cost function

$$\Phi = \sum_{i,j} \left[\left(\frac{\tilde{D}^{i,j}}{\sigma_D} - M_n^{i,j} \right) w_{\text{noise}}^{i,j} \right]^2, \tag{17}$$

with $\tilde{D}$ as the masked acceleration spectrum, normalized by its standard deviation

$$\sigma_D = \left[\frac{1}{N-1} \sum_i^N \left(\tilde{D}_i - \overline{\tilde{D}} \right)^2 \right]^{1/2}. \tag{18}$$

The model weighting function $\mathbf{w}_{\text{noise}} = 1 - \mathbf{w}_G + w_{\text{floor}}$ is the opposite of the geometric weight from (15) and downweights data points with high signal-to-noise ratios, such that (16) fits to the background noise rather than the data. The noise cost function (17) is minimized for each event individually using the gradient descent methods described in section 7. The resulting noise-reduced data matrix

$$\mathbf{D} = \frac{\tilde{D}}{\sigma_D} - \mathbf{M}_n \tag{19}$$

is used for the actual model fitting. It contains the noise-corrected and normalized data for each event. The geometric data selection and the constructed weight function focus the nonlinear optimization on individual dispersed wave events, while downweighting neighboring events and the seasonally changing low-frequency noise due to sea ice (Fig. 1). The data selection process

---

[1] The data are smoothed by using a running mean with a width that is 0.2% of the size of the data matrix.

generates a collection of 250 similar events that can be well characterized by the model function (section 4).

7. Nonlinear fitting

The optimization method changes the parameters **p** of the model function M to minimize the cost function $\Phi[M(\mathbf{p}), \mathbf{p}]$ equation in (6). The smallest value of Φ that the method finds represents the best fit between model and data and is only dependent on a set of parameters **p**.

a. Choice of initial values

The initial values of these parameters are either set to a standard value or are informed by the geometrical form of the data mask (Fig. 4b). The first guess dispersion slope m_f and nondimensional initial time $\tilde{t}_0$ are taken from the masking procedure (red line in Fig. 4a), and the peak frequency f_m is initially set to the peak frequency of the masked data. The initial f_m is also used to calculate $\hat{\alpha}(\alpha)$ from (2) assuming a wind speed of $10 \, \mathrm{m \, s^{-1}}$. Other parameters that modify the spectral shape are initialized from the JONSWAP spectrum standard values (Hasselmann et al. 1973), which are estimated based on open ocean observations.

An overview of the sensitivity to parameter values is shown in Fig. 5. The initial parameters and their limits are set to physically plausible ranges (see the appendix, Table 1, and Massel 1996), such that they allow a wide range of possible values, and equally importantly, also adjust to the noise level. In high noise cases, the model is often unrealistic and results in a poor fit characterized by a large fractional error (section 8a). These cases can be identified and are not considered for further analysis (section 8c).

b. Optimization method and estimations of uncertainty

The nonlinear model (12) is optimized using a two-stage fitting algorithm to minimize the cost function Φ in (6). In the first stage, the model is initialized with $\mathbf{p}_0$ and then changed using the Levenberg–Marquardt (LM) algorithm (damped least squares; Newville et al. 2014) to find a local minimum of the cost function. The LM algorithm calculates the local gradient in parameter space and moves its next guess of parameters in the direction of the gradient. The iteration terminates if the change of the cost function is small ($< 10^{-15}$), if the change in the independent variables is small, or if the number of iterations exceeds its limit defined as $100(n + 1)n$, with n being the size of **D**. We used a gradient method first, rather than a nonlinear search, because of its faster convergence to a (local) minimum for a relatively smooth cost function.

TABLE 1. Table of model fitting parameters, their initial conditions, and priors. An initial value of "adjusted" indicates that the initial value is inferred from the data.

Parameter	Initial value $(\cdot)^0$	Min	Max	Prior std $(\cdot)^\sigma$
$\hat{\alpha}$	αg^2	$0.01 \, \hat{\alpha}^0$	$100 \, \hat{\alpha}^0$	0.2
f_m	Varying	0.04	0.1	0.002
γ	2	1	4	0.4
$\tilde{t}_0$	Varying	-0.5	0.2	0.04
σ_0	0.05	0.01	0.1	0.01
m_f	Varying	$0.5 \, m_f^0$	$2.5 \, m_f^0$	0.2
χ	5/4	1	3/2	0.05
b	2.1	1.2	3	0.08

In the second step, a parallel tempering Markov chain Monte Carlo (PTMCMC; Goodman and Weare 2010; Foreman-Mackey et al. 2013; Earl and Deem 2005) method is used to further minimize the cost function and to produce an a posteriori error distribution for all variables simultaneously. This process is similar to simulated annealing, where the progress toward an optimal solution can only be seen from the average of many iterations rather than from each single iteration (Kirkpatrick et al. 1983). This is a powerful tool in situations in which multiple optimal solutions could exist, as in this problem: even though one origin per wave event is assumed, the uncertainty estimate from PTMCMC is generally capable of capturing several wave events that arrive at the same time.

Each Markov chain is initialized with the optimal parameters from the steepest descent method, and its first guess is seeded from a random distribution. This chain, often called a walker, goes through 1000 function evaluations, with two different annealing temperatures, but only every second evaluation from the final 75% of this process contributes to the error distribution (750 function evaluations per walker). This is repeated 1000 times in a Monte Carlo sense to create a distribution with 7.5×10^5 data points in the eight-parameter space.

Figure 6 shows two examples of codistributions of two elements of **p** for the event in Fig. 4c. (All distributions are shown in Fig. S1.) The distributions have clear maxima, which are the optimal values for each parameter. We use the median (blue lines in Fig. 6) as the best model fit, while half the difference of the 15.87% and 84.13% quantiles (dashed lines, the width of one standard deviation in a normal distribution) is taken as a simple measure of uncertainty (Newville et al. 2014). The resulting best (median) model fit is shown in Fig. 4c.

The parameters are assumed to be uncorrelated (see the appendix and Fig. S1), with their width being sensitive to the choice of prior uncertainty values (section 4 and the appendix). However, the codistribution of the slope $\tilde{t}_0$ and intersect $m_{\tilde{i}}$ parameters shows a correlated error in
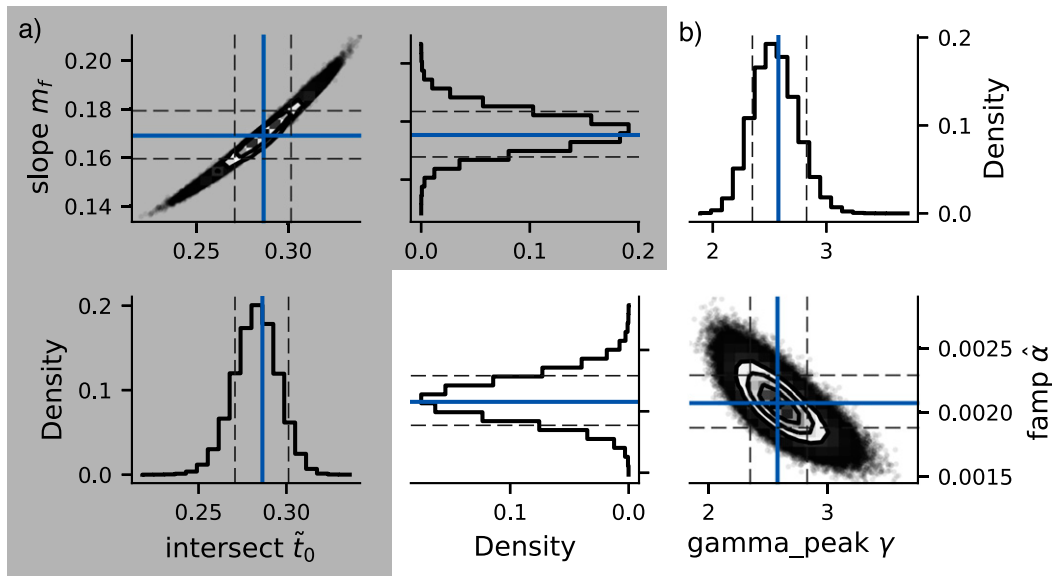
FIG. 6. PDFs of four parameters inferred in an example case using the PTMCMC algorithm. (a) The three panels in the top left show the joint PDF as well as the single PDF for the slope parameter m_f and intersect parameter $\tilde{t}_0$. (b) The three panels in the lower right show the joint PDF and single PDFs for the peak parameter γ and the amplitude parameter $\hat{\alpha}$. The blue lines indicate the median and the dashed lines the 15.87th and 84.13th percentiles.

all observed cases (Fig. 6a). This distribution is converted from the $\tilde{t}_0$–$m_{\tilde{t}}$ space to initial time

$$t_0 = \tilde{t}_0 \Delta T + t_{\text{start}}, \tag{20}$$

and radial distance

$$r_0 = \frac{g}{4\pi \Delta T} m_{\tilde{t}}, \tag{21}$$

with $\Delta t = t_{\text{end}} - t_{\text{start}}$ (section 5) and using the deep water dispersion relations (Barber and Ursell 1948; Munk and Snodgrass 1957; Snodgrass et al. 1966). The resulting distribution reveals probabilities of wave event origin in time and radial distance and allows us to create probability maps to quantify the likelihood of a specific origin. Figure 7 shows these maps of probability in the time and radial distance space (TR space). They are direct conversions of the observational scatter captured by the PTMCMC method (section 7). Smaller patches in the TR space (Fig. 7, bottom) correspond with very certain model estimates of events in the observed spectrogram (Fig. 7, top), while larger patches in the TR space correspond with less well defined wave events. This ambiguity between a recent nearby event and a distant event from further back in the past can be reduced by using other dependencies in the model and by drawing on extra information about wind events from atmospheric observations and models. The authors plan to address this in future work.

8. Performance of the optimization

a. Distribution of fitting parameters

To compare the eight model parameters consistently we express them as normalized distance computed relative to the prior (section 4). Figure 8 shows the decomposition of the model cost J_m in (6) into the cost introduced by each of the eight parameters for the three front stations. The initial cost of each parameter is zero (green line), while the median model cost of the parameter is indicated by the black line.

The distribution of costs due to the parameter adjustment is similar for all parameters and for all stations, with a clear maximum close to zero. Final parameter values close to the initial value suggest that small changes in the parameters are enough to reduce J_d substantially without introducing costs in J_m. However, there are cases for all parameters where the final value deviates from the initial value more substantially. In these cases, the costs in J_m introduced by large parameter adjustments are small compared to the reduction achieved in J_d such that the overall cost Φ is still minimized. This must be the case because the ratio of model to data cost, J_m/J_d, rarely exceeds 20% for all fitted cases (Fig. A1). This suggests that, based on gradients in Φ, an efficient minimization can often be effected via small changes in the model parameters, or sometimes through a few larger changes to a subset of parameters. Since the gradient descent method terminates if the number of
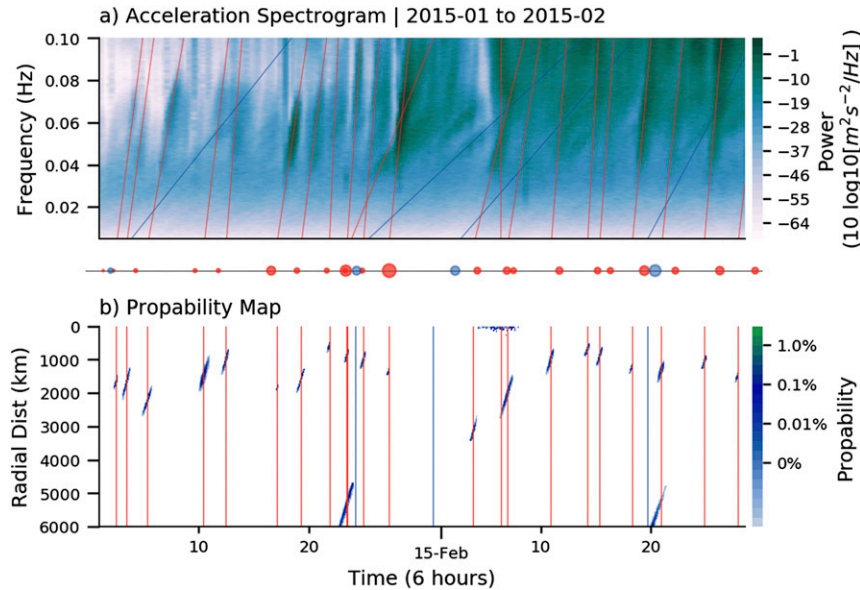
FIG. 7. Seismic spectrogram in the RIS (as Fig. 1). The lines show the optimized dispersion slopes for events from the Southern Ocean (red, south of 30°S) or north of it (blue).(b) Probability maps of wave origins in radial distance and time for the same time as (a). The red and blue lines indicate the best guess of the initial time, that is, when the lines in (a) cross the abscissa. Each elliptic pattern in (b) is the probability distribution of the corresponding wave event in (a). The smaller the probability patterns, the more certain the wave event origin is. The red and blue dots between both panels show the amplitude as observed on the RIS but position at their initial time.

iterations exceeds its limit, it is possible that regions of parameter space in the direction of small gradients are never explored. However, these are sampled later by the Monte Carlo method.

Parameters that determine the radial distance and initial time are optimized during the minimization procedure. The position of $\tilde{t}_0$ has, in the median, a larger contribution to J_m than other parameters (Fig. 8a). That is, in about 85% of all cases, changes in the modeled position were necessary to achieve minimum cost. In contrast, the model slope parameter m_f also adjusts (Fig. 8b), but introduces smaller costs, because the manual selection criteria better define its initial values than the position of the initial time (section 6a).

Two parameters of the JONSWAP spectrum (γ and σ_0) also introduce noticeable model costs (Figs. 8c,d). Their initial values were manually adjusted away from the standard JONSWAP values (Fig. A2), because trials that started with standard JONSWAP values reduced the overall quality of the fit (larger fractional errors). In the end, neither the standard JONSWAP values nor the chosen initial values (the appendix) are the best choices to capture the shape of the observed swell events well. However, an additional free parameter, χ (Fig. 8e), that changes the general shape of the spectrum, often remains close to its initial, theoretically

well-constrained value of 5/4 [see (7)], as predicted by Hasselmann et al. (1973).

b. Comparing fitting performance between front stations

The optimization algorithm found 225 events during the 24-month observational period. They occurred at each of the three "front" stations DR01, DR02, and DR03, each about 2 km away from the ice-shelf edge (blue, orange, and red dots in Fig. 3); common events at all three stations are identified by similar arrival times.

The seismic stations at the RIS ice shelf front are close together ($\approx$80-km separations) compared to the distance traveled by the waves ($\approx$1500 km). As a result, the incoming wave angles and amplitudes are assumed to be uniform along the ice shelf front. However, the event amplitudes observed at DR01 are systematically larger than at DR02 and DR03 (Fig. 9a). The difference in amplitude between the stations may be caused by the structure of the ice shelf, and affected by a major rift that separates DR03 and DR01, near DR02. If the amplitude difference were due to the incoming angle of the waves rather than ice-shelf rheology, one would expect more randomly scattered differences between the stations, because the incident waves are expected to come from a wider range of incident angles.

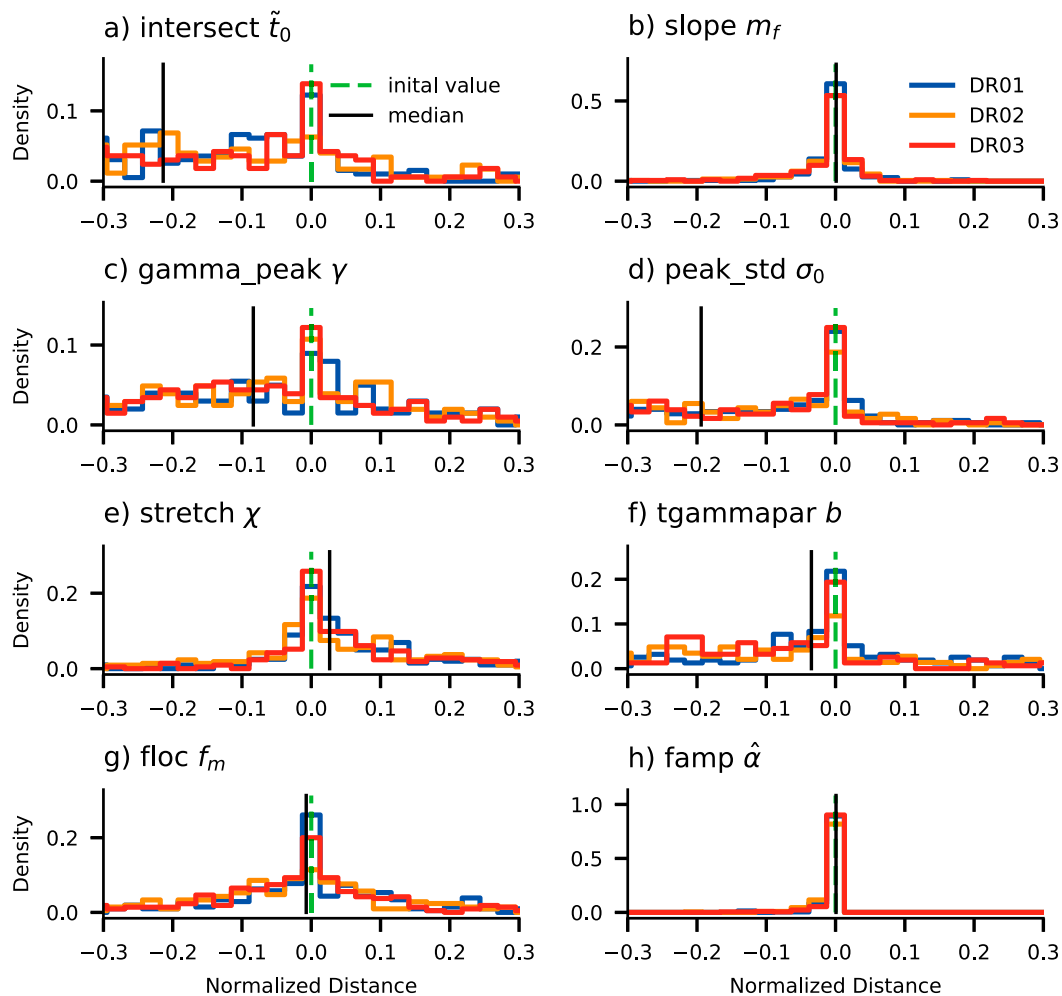## Fitting Parameter Statistics - Normalized

FIG. 8. PDFs of parameters from the 225 fitted parameters sets in the three front stations DR01 (blue), DR02 (orange), and DR03 (red) expressed as normalized distance to initial value following (6). Green dashed lines indicate the normalized initial values of the model and the black lines are the medians of all cases at all stations.

Comparison of other fitted parameters suggests that the same wave events lead to comparable observations at the three RIS-front stations. For example, peak frequencies of common events vary by about 0.02 Hz or less between the three stations (Fig. 9b). Austral summer events, that generally have higher peak frequencies, also have a similar observed frequencies at all three front stations. Differences between them are likely due to independent noise or local shelf structure. Peak frequencies lower than 0.04 Hz are mainly observed at DR01, while the same events at DR02 or DR03 rarely fall below 0.04 Hz. Reasons for this could be systematic differences in the ice shelf front geometry, or crevasses and rifts in the ice shelf between the stations, which are beyond the scope of this study. Aside from this discrepancy, observations show peak frequencies and spectral shapes similar to those in the open ocean, indicating that the RIS response to incident waves maintains properties from the open ocean waves.

The uncertainty in the radial distance estimate, and therefore also the uncertainty in initial time, can vary between the front stations. Figure 9c shows the radial distance uncertainty for all events (defined as half the distance between the 15.87 and 84.13 percentiles of the uncertainty distribution (section 7b, Fig. 6a black dashed lines). The difference in radial distance uncertainty between the stations is generally larger for larger uncertainties. For many events, DR02 and DR03 have smaller uncertainties than DR01. These events are often, but not always, selected by the criterion of smallest fractional error, which is explained in the next section (black half dots, Fig. 9c).
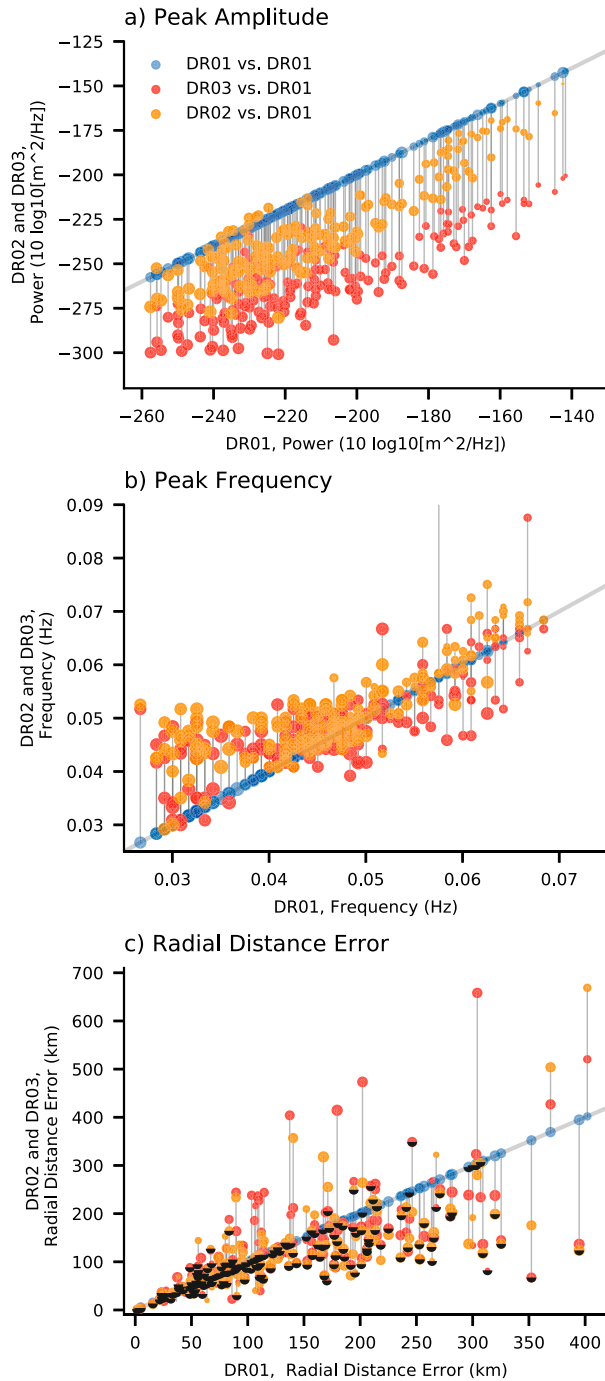
FIG. 9. (a) Peak amplitude of data under fitted events in DR01 (blue), DR02 (orange), and DR03 (red) against DR01. The dot sizes indicate the event amplitude. The gray lines connect the same event in the three stations. (b) As in (a), but for the peak frequency. (c) As in (a), but for the radial distance error. The half-black circles indicate the station with the minimal fractional error for each case (as in Fig. 10b).

c. Measures of fit

The performance of the model optimization is assessed quantitatively using the fractional error between data and model

$$e_{\text{frac}} = \frac{J_d}{\sum_i^N D_i^2 w_i^2}, \tag{22}$$

with J_d identifying the cost, D_i the data, and w_i the specific weight for each point. A good model fit, that is, low fractional errors, means that Φ is small compared to the data. The distributions of e_{frac} for DR01–DR03 are shown in Fig. 10. Events with a $e_{\text{frac}} \leq 0.6$ are defined as successfully fitted (gray area in Fig. 10a). This represents between 74% and 84% of all events (varying between stations) considered in this analysis. Values of $e_{\text{frac}} > 0.6$ are interpreted as unsuccessful fits and represent between 16% and 26% of all cases considered.

The signal-to-noise ratio is generally the same for all three stations (the median e_{frac} is 0.43 for DR01 and DR03 and 0.37 for DR02), with a large variance between individual events (Fig. 10b). For each swell event, the comparison of e_{frac} varies widely between stations, with no systematic preference for one station. Higher values of e_{frac} correspond to a less successful explanation of the data by the model, or higher noise levels.

Most events have a small fractional error at all stations (green area in Fig. 10b), but there are a number of events in which one station performs substantially better than its neighboring stations (gray areas). To optimize the use of multiple observations, we compile a set of 208 events taken from DR01–03 (88% of all initial events, black half dots) based on events when e_{frac} is smallest. Events with a fractional error larger than 0.6 at all three stations are not considered (upper right area in Fig. 10b). By considering all three stations, we identify 35 additional events that would not meet our fractional error criterion for DR01 alone. The use of the three stations reduces the mean fractional error from 0.42 to 0.26, with the disadvantage that no common attenuation transfer function can be determined for the events (Figs. 10a and 9a).

As shown in Fig. 10b, the fractional error for the same swell event varies between the different observation sites, although differences between the three spectrograms are often difficult to distinguish by eye (cf. Fig. 1 with Figs. S2 and S3). The fractional error criterion provides a single metric for evaluating all observed events, and when we adopt a conservative threshold for the fractional error, it allows rigorous quality control of the data.

The optimization procedure identified swell arrivals at the RIS based on physical constraints that are parameterized in the model (section 5). If data and model are
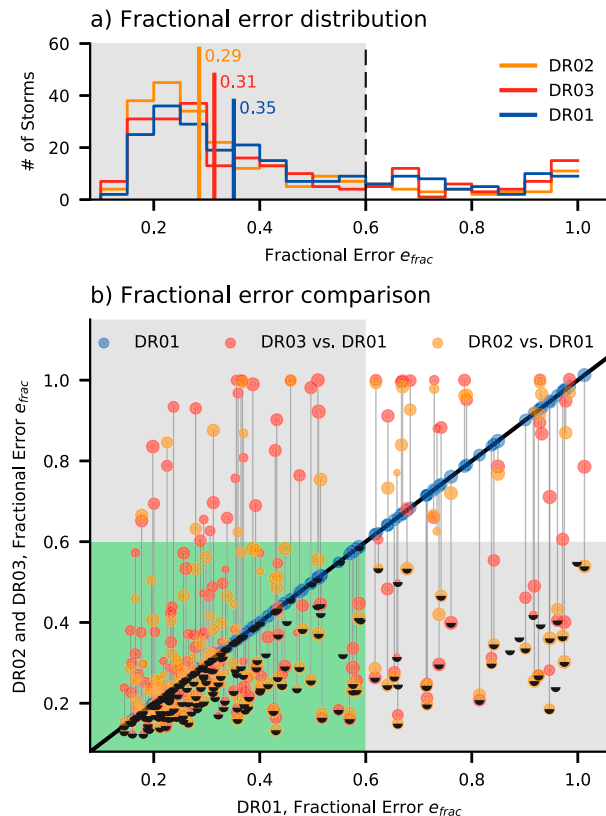
FIG. 10. Statistics of the fractional error. (a) Histograms of the fractional error for the 225 fitted events in the front stations DR01 (blue), DR02 (orange,) and DR03 (red). The colored vertical lines show the median fractional error per station. The vertical dashed line is the threshold chosen to identify good fits ($e_{frac} \leq 0.6$). (b) Fractional error of each event in DR01, DR02, and DR03, compared against DR01. Dot sizes indicate the event amplitude. Gray lines connect the same event in all three stations. For each case, the half-black circles indicate the station with the smallest fractional error. Gray areas correspond to $e_{frac} \leq 0.6$ in DR01, DR02, or DR03, respectively. Dots in the green area indicate $e_{frac} \leq 0.6$ in DR01 and DR02, or DR01 and DR03.

similar, then the fractional error is small, and we conclude that the observed event is indeed related to incident, dispersed swell events. In contrast, a large fractional error suggests a weaker signal-to-noise ratio, implying for example that additional processes have distorted the incident wave events, that more than one storm is contributing to forcing the swell, or that the observed features are not generated by dispersed swell waves.

d. Observed storm activity and wave spectra

Figure 11a shows the radial distance estimate, number of storms per degree latitude, radial uncertainty, and time uncertainty. Events with $e_{frac} \leq 0.6$ are black (Fig. 9c, green area), while the remaining events are gray. The vertical lines show one radial-distance standard deviation, inferred from the uncertainty estimate

(section 7b), while the standard deviation in time is not plotted because it is too small. The distributions of both radial distance and time uncertainty are shown in Figs. 11c and 11d.

The observed RIS events suggest that storm-generated packets of surface waves arrive at the ice shelf more than twice per week (2.3 events per week, all dots in Fig. 11a). In total, 208 of these events meet the criterion of a well identified/well fitted event ($e_{frac} < 0.6$, section 7). The majority (187 events, 1.8 week^{-1}) originate from the Southern Ocean, and we see no distinct seasonality in their occurrence rate. The rest (28 events) originate from the subtropics or the Northern Hemisphere in boreal winter, and are generated from tropical or extratropical cyclones in the North Pacific (Cathles et al. 2009; MacAyeal et al. 2006; Bromirski et al. 2010). The median estimated uncertainty is about 110 km in space and 2 h in time (Figs. 11c,d). These uncertainties are small enough to allow the wave events to be matched to specific high wind speed areas under storms that are often related to fronts or warm conveyor belts (Schemm and Wernli 2014), as we will explain in future work.

9. Summary and conclusions

Ocean swell is commonly observed along coastlines and its origin is a long-standing question in oceanography (Munk 1947). Swell is known to be generated by strong winds associated with extratropical cyclones, but the exact positions of swell generation have not been well characterized. Areas of swell generation are hypothesized to correspond to areas of upper-ocean mixing and often of intense air–sea heat exchange. Here, instead of relying on in situ observations, we have developed a methodology to analyze the remote observation of ocean swell. This can be used to infer characteristics of the swell generation region.

To assess the locations of high surface winds over the Southern Ocean we developed a method that combines spectral analysis from time series of swell arrivals at a single point with a parametric model optimization. The parametric model used here represents a combination of the linearized dispersion relation and the JONSWAP spectrum. These models are based on different descriptions of the wave source: while the inversion of the linear dispersion relation assumes a source that is a delta function in space and time, the JONSWAP spectrum uses a wind speed over a given length or for a given time (Munk 1947; Hasselmann et al. 1973). Both concepts are approximations to reality, as is their combination (12). The resulting eight-parameter model provides a framework to describe the (point) origin and spectra of swell observed by a single point observation.
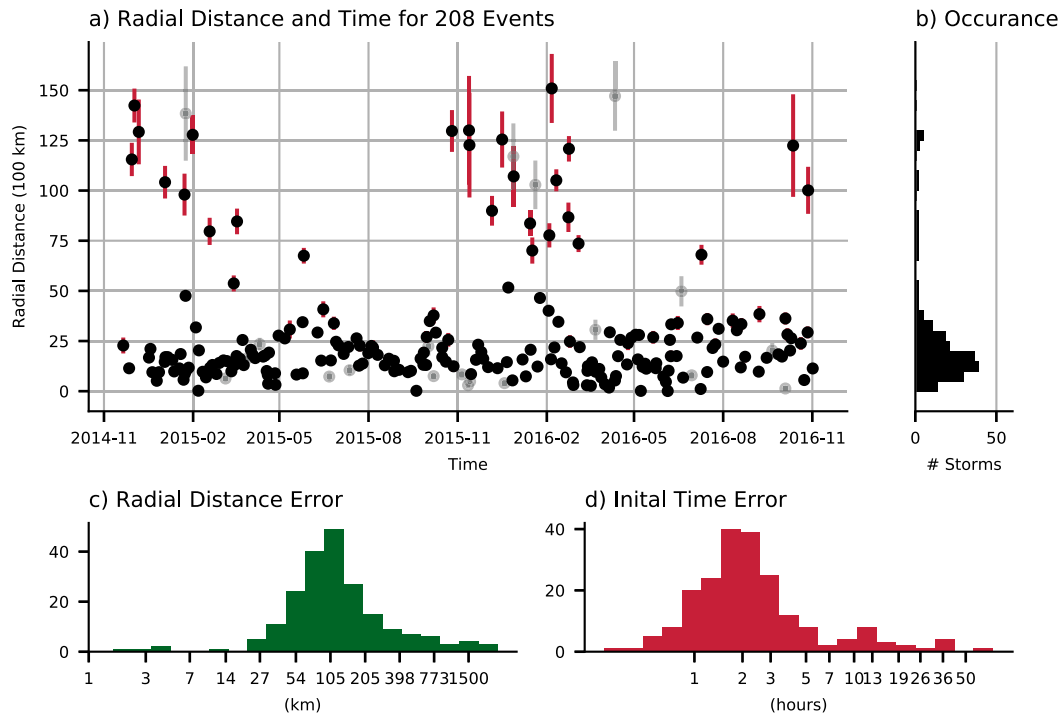
FIG. 11. (a) Estimated radial distances and initial times for 225 wave events. Black dots are events with a fractional error $e_{frac} \leq 0.6$ (208 events), while gray dots correspond to events with $e_{frac} > 0.6$ (17 events). Vertical red lines show the standard deviation of the time inferred from PTMCMC (section 7b). (b) Histogram of numbers of storms per radial distance. Histogram of (c) radial distance error and (d) initial time error.

The systematic comparison of the parametric model with the wave events allows us to quantitatively test models of generation and propagation. It provides a framework for learning about the underlying physical process, rather than letting the machine construct a model of the observations without physical constraints (often referred to as unsupervised machine learning; Editorial Board 2019).

A two-stage optimization procedure is used to fit the model to the data, by exploring the full eight-parameter space simultaneously. In our procedure, we first apply a gradient descent method and subsequently carry out Monte Carlo sampling to ensure 1) that the gradient descent minimum is "annealed" to a globally optimal position in parameter space, and 2) that there is a well-sampled uncertainty distribution for each parameter. The uncertainty in the fitting parameters is estimated using a simulated annealing approach (PTMCMC) that is based on the data and the model, with no prior assumptions about the functional form of the distribution of the uncertainty. The methodology exposes differences between events (Fig. A2) that are not obvious to the eye (Fig. 1) and also allows quantitative comparison of observations of the same events at three observation sites (section 8b), which can give insight into wave interactions with the ice shelf and into differences in wave

propagation to the three sites. We showed that the set of prior model parameters is sufficient to generate well behaved model fits, with fitted model parameters adjusted minimally. Larger parameter adjustments that introduce model costs (lasso regularization) are only introduced for a minimization of the total costs, due to a better fit of the model. Future work could investigate physical reasons for these extreme parameter by reinterpreting the model function.

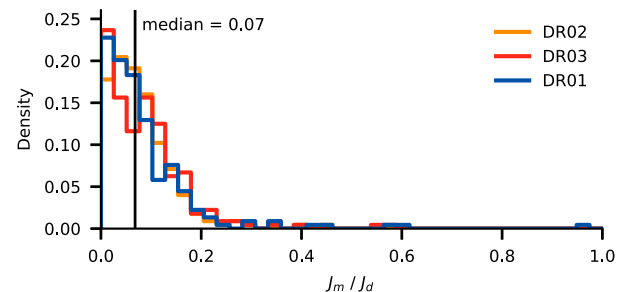The method developed here could in principle be used for any data documenting swell arrivals and unmarred

FIG. A1. Distribution of the ratio of model cost function J_m in (5) to data cost function J_d in (6) for all events in the three front stations. The median is shown by the black line.
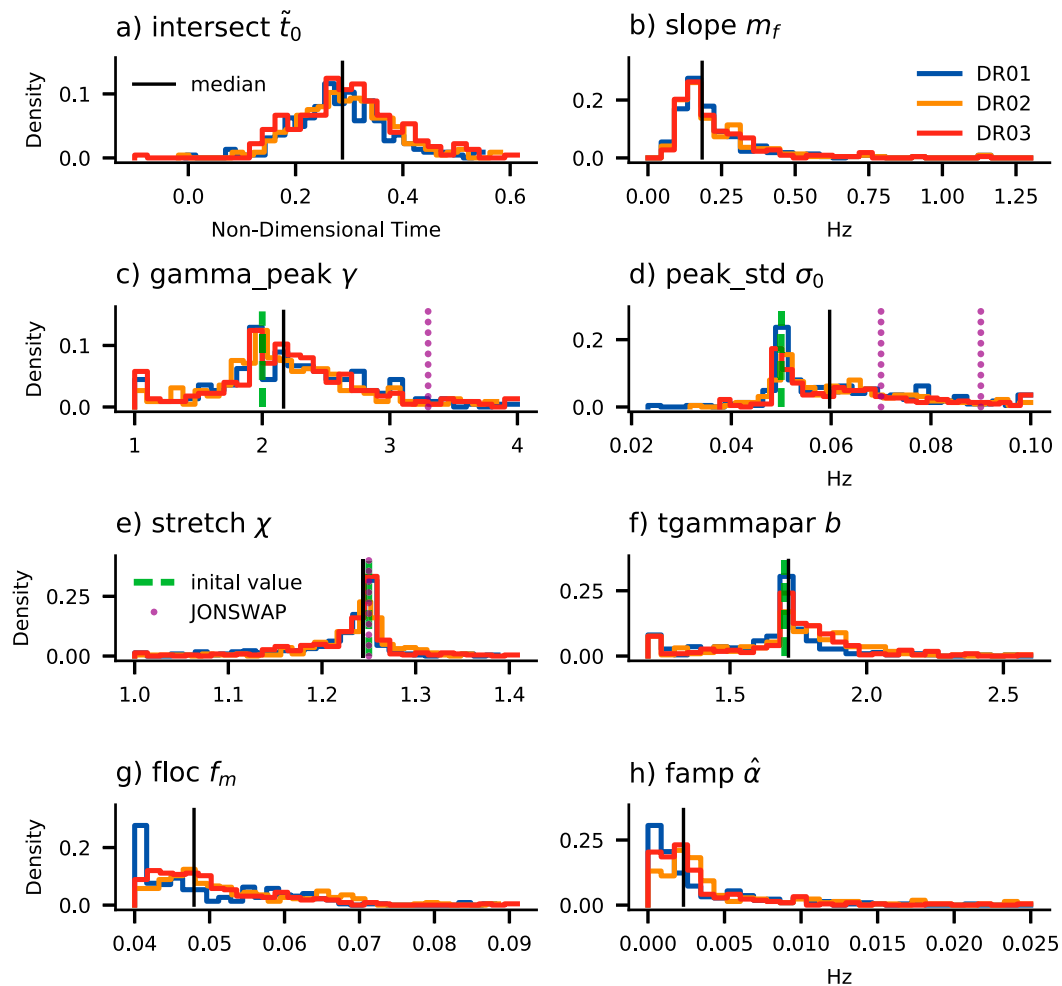
Fitting Parameter Statistics

FIG. A2. As in Fig. 8, but not normalized by prior uncertainty and initial guesses. Green dashed lines indicate the initial values of the model for cases with independent initial parameters, the black lines are the medians of all cases at all stations, and the purple lines are the standard JONSWAP values if applicable (section 5).

by locally generated noise. Here we apply the approach to three seismic stations on the RIS, which provide a dataset of well resolved swell arrivals close to one of most active cyclone-genesis regions of the globe (Hoskins and Hodges 2005). These point observations allow us to identify the swell origin in space and time.

For the two years of RIS data available, the optimization method results in a catalog of 208 (self) similar swell arrivals that can be detected in the ice shelf. In total 187 (90%) of these swell arrivals originate from the Southern Ocean and can be used to improve the understanding of the origin of ocean swell and its interaction with sea ice. The remaining 10% originate from the Northern Hemisphere midlatitudes during the boreal winter season. This is observational evidence of the distinctly different seasonality in both hemispheres,

and establishes the incidence of swell impacts on the Ross Ice Shelf throughout the year. The uncertainties of origin in location and time are correlated and found to be about 110 km or 2 h, respectively (Figs. 11c,d), which allows us to reduce the uncertainty of the wave origins to the scale of atmospheric surface fronts or other features within cyclones, rather than the scales of cyclones themselves (1000 km, 5 days).

A follow-up paper will apply this method to the RIS data to verify Southern Hemisphere storm position in atmospheric reanalysis data. Ice shelf seismic measurements, such as the RIS data, are particularly well suited for detecting swell arrivals, because the general lack of open water near the seismic stations means that local wind waves are absent in the seismic data, resulting in a relative noise-free set of swell arrivals. In addition, weaker

events that might complicate the interpretation of the swell arrivals appear to be largely attenuated when they propagate through sea ice.

Future work could also apply the methodology to a broader network of ice-shelf seismic stations or even to midlatitude wave buoys, bottom pressure sensors, or even land-based seismic stations, although further tuning would likely be needed to distinguish between locally and remotely generated waves. The methodology is suitable for data records of any duration, ranging from a few days to multiple years, provided that they provide sufficient temporal resolution. The RIS time series are relatively short, and a manual procedure was used to preselect candidate swell arrivals. If this method were applied to longer observation periods, the manual selection could be replaced with an automated search strategy from the catalog of available machine learning tools (Bergen et al. 2019). Finally, by combining observations at several stations with additional information about ocean currents or sea ice, the method could potentially be further extended to study storm intensity and/or sinuous wave travel paths.

APPENDIX

Cost Function and Residual Distributions

The model and data cost functions are defined in section 4. Figure A1 shows the dominance of the data cost function as the distribution of their ratio.
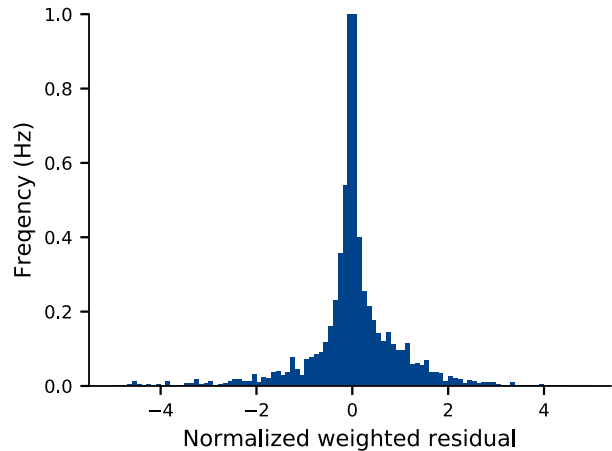

FIG. A3. Distribution of the residual between the data and model $\mathbf{w}(\mathbf{D} - \mathbf{M})$ for the example case in Fig. 4 (section 7).

The parametric model of dispersed swell arrivals is introduced in section 5, and the initial values and their prior estimated uncertainties are summarized in Table 1. The distribution of the final normalized distance for each parameter as shown in Fig. 8 is derived by using

$$H^i = \frac{p_0^i - p_{\text{final}}^i}{p_\sigma^i},$$

where p_0^i is the initial value, p_{final}^i the final value, and p_σ^i the prior weight of parameter i. The terms p_0^i and p_σ^i can vary from case to case (section 7a). Figure A2 shows the distributions for all parameters as in Fig. 8 but for p_{final}^i rather than for H^i.

The final model parameters are defined by the median of the Monte Carlo sampling that results in an eight-dimensional uncertainty distribution, shown as joint PDFs for an example case in Fig. S1 and the distribution of the residual between data and model is shown in Fig. A3.

REFERENCES

Ardhuin, F., P. Sutherland, M. Doble, and P. Wadhams, 2016: Ocean waves across the Arctic: Attenuation due to dissipation dominates over scattering for periods longer than 19 s. *Geophys. Res. Lett.*, **43**, 5775–5783, https://doi.org/10.1002/2016GL068204.

Barber, N. F., and F. Ursell, 1948: The generation and propagation of ocean waves and swell. I. Wave periods and velocities. *Philos. Trans. Roy. Soc. London*, **240**, 527–560, https://doi.org/10.1098/rsta.1948.0005.

Bergen, K. J., P. A. Johnson, M. V. de Hoop, and G. C. Beroza, 2019: Machine learning for data-driven discovery in solid Earth geoscience. *Science*, **363**, eaau0323, https://doi.org/10.1126/science.aau0323.

Bourassa, M. A., and Coauthors, 2013: High-latitude ocean and sea ice surface fluxes: Challenges for climate research. *Bull. Amer.*

Meteor. Soc., **94**, 403–423, https://doi.org/10.1175/BAMS-D-11-00244.1.

Bromirski, P. D., R. E. Flick, and N. Graham, 1999: Ocean wave height determined from inland seismometer data: Implications for investigating wave climate changes in the NE Pacific. *J. Geophys. Res.*, **104**, 20 753–20 766, https://doi.org/10.1029/1999JC900156.

——, O. V. Sergienko, and D. R. MacAyeal, 2010: Transoceanic infragravity waves impacting Antarctic ice shelves. *Geophys. Res. Lett.*, **37**, L02502, https://doi.org/10.1029/2009gl041488.

——, A. Diez, P. Gerstoft, R. A. Stephen, T. Bolmer, D. A. Wiens, R. C. Aster, and A. Nyblade, 2015: Ross ice shelf vibrations. *Geophys. Res. Lett.*, **42**, 7589–7597, https://doi.org/10.1002/2015GL065284.

——, and Coauthors, 2017: Tsunami and infragravity waves impacting Antarctic ice shelves. *J. Geophys. Res. Oceans*, **122**, 5786–5801, https://doi.org/10.1002/2017JC012913.

Cathles, L. M., E. A. Okal, and D. R. MacAyeal, 2009: Seismic observations of sea swell on the floating Ross Ice Shelf, Antarctica. *J. Geophys. Res.*, **114**, F02015, https://doi.org/10.1029/2007JF000934.

Collard, F., F. Ardhuin, and B. Chapron, 2009: Monitoring and analysis of ocean swell fields from space: New methods for routine observations. *J. Geophys. Res.*, **114**, C07023, https://doi.org/10.1029/2008jc005215.

Collins, C. O., W. E. Rogers, A. Marchenko, and A. V. Babanin, 2015: In situ measurements of an energetic wave event in the Arctic marginal ice zone. *Geophys. Res. Lett.*, **42**, 1863–1870, https://doi.org/10.1002/2015GL063063.

Delpey, M. T., F. Ardhuin, F. Collard, and B. Chapron, 2010: Space-time structure of long ocean swell fields. *J. Geophys. Res.*, **115**, C12037, https://doi.org/10.1029/2009jc005885.

Earl, D. J., and M. W. Deem, 2005: Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, **7**, 3910–3916, https://doi.org/10.1039/b509983h.

Editorial Board, 2019: Use machines to tame big data. *Nat. Geosci.*, **12**, 1, https://doi.org/10.1038/s41561-018-0290-6.

Elfouhaily, T., B. Chapron, K. Katsaros, and D. Vandemark, 1997: A unified directional spectrum for long and short wind-driven waves. *J. Geophys. Res.*, **102**, 15 781–15 796, https://doi.org/10.1029/97JC00467.

Foreman-Mackey, D., D. W. Hogg, D. Lang, and J. Goodman, 2013: Emcee: The MCMC Hammer. *Publ. Astron. Soc. Pac.*, **125**, 306–312, https://doi.org/10.1086/670067.

Fox, C., and V. A. Squire, 1994: On the oblique reflexion and transmission of ocean waves at shore fast sea ice. *Philos. Trans. Roy. Soc. London*, **347A**, 185–218, https://doi.org/10.1098/rsta.1994.0044.

Fretwell, P., and Coauthors, 2013: Bedmap2: Improved ice bed, surface and thickness datasets for Antarctica. *Cryosphere*, **7**, 375–393, https://doi.org/10.5194/tc-7-375-2013.

Gallet, B., and W. R. Young, 2014: Refraction of swell by surface currents. *J. Mar. Res.*, **72**, 105–126, https://doi.org/10.1357/002224014813758959.

Goodman, J., and J. Weare, 2010: Ensemble samplers with affine invariance. *Comm. App. Math. Comp. Sci.*, **5**, 65–80, https://doi.org/10.2140/camcos.2010.5.65.

Gruber, N., and Coauthors, 2019: The oceanic sink for anthropogenic CO_2 from 1994 to 2007. *Science*, **363**, 1193–1199, https://doi.org/10.1126/science.aau5153.

Haran, T., and J. Bohlander, 2014: MODIS Mosaic of Antarctica 2008-2009 (MOA2009) Image Map, Version 1. National Snow and Ice Data Center, accessed 10 May 2018, https://doi.org/10.7265/N5KP8037.

——, ——, T. Scambos, T. Painter, and M. Fahnestock, 2005: MODIS Mosaic of Antarctica 2003-2004 (MOA2004) Image Map, Version 1. National Snow and Ice Data Center, accessed 10 May 2018, https://doi.org/10.7265/N5ZK5DM5.

Hasselmann, K., and Coauthors, 1973: Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). Ergänzungsheft 8-12, Deutches Hydrographisches Institut, 93 pp., http://resolver.tudelft.nl/uuid:f204e188-13b9-49d8-a6dc-4fb7c20562fc.

——, W. Sell, D. B. Ross, and P. Müller, 1976: A parametric wave prediction model. *J. Phys. Oceanogr.*, **6**, 200–228, https://doi.org/10.1175/1520-0485(1976)006<0200:APWPM>2.0.CO;2.

Hoskins, B. J., and K. I. Hodges, 2005: A new perspective on Southern Hemisphere storm tracks. *J. Climate*, **18**, 4108–4129, https://doi.org/10.1175/JCLI3570.1.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi, 1983: Optimization by simulated annealing. *Science*, **220**, 671–680, https://doi.org/10.1126/science.220.4598.671.

Kohout, A. L., M. J. M. Williams, S. M. Dean, and M. H. Meylan, 2014: Storm-induced sea-ice breakup and the implications for ice extent. *Nature*, **509**, 604–607, https://doi.org/10.1038/nature13262.

Lipovsky, B. P., 2018: Ice shelf rift propagation and the mechanics of wave-induced fracture. *J. Geophys. Res. Oceans*, **123**, 4014–4033, https://doi.org/10.1029/2017JC013664.

MacAyeal, D. R., and Coauthors, 2006: Transoceanic wave propagation links iceberg calving margins of Antarctica with storms in tropics and Northern Hemisphere. *Geophys. Res. Lett.*, **33**, L17502, https://doi.org/10.1029/2006gl027235.

——, E. A. Okal, R. C. Aster, and J. N. Bassis, 2009: Seismic observations of glaciogenic ocean waves (micro-tsunamis) on icebergs and ice shelves. *J. Glaciol.*, **55**, 193–206, https://doi.org/10.3189/002214309788608679.

Marone, C., 2018: Training machines in Earthly ways. *Nat. Geosci.*, **11**, 301–302, https://doi.org/10.1038/s41561-018-0117-5.

Marshall, D. P., D. R. Munday, L. C. Allison, R. J. Hay, and H. L. Johnson, 2016: Gill's model of the Antarctic Circumpolar Current, revisited: The role of latitudinal variations in wind stress. *Ocean Modell.*, **97**, 37–51, https://doi.org/10.1016/j.ocemod.2015.11.010.

Massel, S. R., 1996: *Ocean Surface Waves: Their Physics and Prediction*. 2nd ed. Advanced Series on Ocean Engineering, Vol. 11, World Scientific, 491 pp.

Massom, R. A., T. A. Scambos, L. G. Bennetts, P. Reid, V. A. Squire, and S. E. Stammerjohn, 2018: Antarctic ice shelf disintegration triggered by sea ice loss and ocean swell. *Nature*, **558**, 383–389, https://doi.org/10.1038/s41586-018-0212-1.

Miles, J. W., 1957: On the generation of surface waves by shear flows. *J. Fluid Mech.*, **3**, 185–204, https://doi.org/10.1017/S0022112057000567.

——, 1960: On the generation of surface waves by turbulent shear flows. *J. Fluid Mech.*, **7**, 469–478, https://doi.org/10.1017/S0022112060000220.

Munday, D. R., and X. Zhai, 2017: The impact of atmospheric storminess on the sensitivity of Southern Ocean circulation to wind stress changes. *Ocean Modell.*, **115**, 14–26, https://doi.org/10.1016/j.ocemod.2017.05.005.

Munk, W. H., 1947: Tracking storms by forerunners of swell. *J. Meteor.*, **4**, 45–57, https://doi.org/10.1175/1520-0469(1947)004<0045:TSBFOS>2.0.CO;2.

——, and F. E. Snodgrass, 1957: Measurements of southern swell at Guadalupe Island. *Deep-Sea Res.*, **4**, 272–286, https://doi.org/10.1016/0146-6313(56)90061-2.

——, G. R. Miller, F. E. Snodgrass, and N. F. Barber, 1963: Directional recording of swell from distant storms. *Philos. Trans. Roy. Soc. London*, **255**, 505–584, https://doi.org/10.1098/rsta.1963.0011.

Newville, M., T. Stensitzki, D. B. Allen, and A. Ingargiola, 2014: LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python. Zenodo, https://doi.org/10.5281/zenodo.11813.

O'Reilly, W. C., C. B. Olfe, J. Thomas, R. J. Seymour, and R. T. Guza, 2016: The California coastal wave monitoring and prediction system. *Coastal Eng.*, **116**, 118–132, https://doi.org/10.1016/j.coastaleng.2016.06.005.

Phillips, O. M., 1957: On the generation of waves by turbulent wind. *J. Fluid Mech.*, **2**, 417–445, https://doi.org/10.1017/S0022112057000233.

——, 1985: Spectral and statistical properties of the equilibrium range in wind-generated gravity waves. *J. Fluid Mech.*, **156**, 505–531, https://doi.org/10.1017/S0022112085002221.

Pierson, W. J., and L. Moskowitz, 1964: A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. *J. Geophys. Res.*, **69**, 5181–5190, https://doi.org/10.1029/JZ069i024p05181.

Rapizo, H., A. V. Babanin, E. Schulz, M. A. Hemer, and T. H. Durrant, 2015: Observation of wind-waves from a moored buoy in the Southern Ocean. *Ocean Dyn.*, **65**, 1275–1288, https://doi.org/10.1007/s10236-015-0873-3.

Rintoul, S. R., 2018: The global influence of localized dynamics in the Southern Ocean. *Nature*, **558**, 209–218, https://doi.org/10.1038/s41586-018-0182-3.

Robinson, W., and T. G. Haskell, 1990: Calving of Erebus Glacier tongue. *Nature*, **346**, 615–616, https://doi.org/10.1038/346615b0.

Schemm, S., and H. Wernli, 2014: The linkage between the warm and the cold conveyor belts in an idealized extratropical cyclone. *J. Atmos. Sci.*, **71**, 1443–1459, https://doi.org/10.1175/JAS-D-13-0177.1.

Snodgrass, F. E., G. W. Groves, K. F. Hasselmann, G. R. Miller, W. H. Munk, and W. H. Powers, 1966: Propagation of ocean swell across the Pacific. *Philos. Trans. Roy. Soc. London*, **259A**, 431–497, https://doi.org/10.1098/rsta.1966.0022.

Squire, V., 2007: Of ocean waves and sea-ice revisited. *Cold Reg. Sci. Technol.*, **49**, 110–133, https://doi.org/10.1016/j.coldregions.2007.04.007.

——, W. H. Robinson, M. Meylan, and T. G. Haskell, 1994: Observations of flexural waves on the Erebus Ice Tongue, McMurdo Sound, Antarctica, and nearby sea ice. *J. Glaciol.*, **40**, 377–385, https://doi.org/10.1017/S0022143000007462.

Swart, N. C., S. T. Gille, J. C. Fyfe, and N. P. Gillett, 2018: Recent Southern Ocean warming and freshening driven by greenhouse gas emissions and ozone depletion. *Nat. Geosci.*, **11**, 836–841, https://doi.org/10.1038/s41561-018-0226-1.

Vaughan, G. L., L. G. Bennetts, and V. A. Squire, 2009: The decay of flexural-gravity waves in long sea ice transects. *Proc. Roy. Soc. London*, **465A**, 2785–2812, https://doi.org/10.1098/rspa.2009.0187.

Wiens, D., P. D. Bromirski, R. C. Aster, R. A. Stephen, P. Gerstoft, and A. Nyblade, 2014: Collaborative research: Dynamic response of the ross ice shelf to wave-induced vibrations and collaborative research: Mantle structure and dynamics of the ross sea from a passive seismic deployment on the ross ice shelf. International Federation of Digital Seismograph Networks, accessed April 2017, https://doi.org/10.7914/SN/XH_2014.