

# Spherical Harmonic Spectral Estimation on Arbitrary Grids

NICHOLAS R. CAVANAUGH

*Climate and Ecosystems Science Division, Lawrence Berkeley National Laboratory, Berkeley, California*

TRAVIS A. O'BRIEN

*Climate and Ecosystems Science Division, Lawrence Berkeley National Laboratory, Berkeley, and  
Department of Land, Air, and Water Resources, University of California, Davis, Davis, California*

WILLIAM D. COLLINS

*Climate and Ecosystems Science Division, Lawrence Berkeley National Laboratory, and Department  
of Earth and Planetary Science, University of California, Berkeley, Berkeley, California*

WILLIAM C. SKAMAROCK

*National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 19 July 2016, in final form 2 May 2017)

## ABSTRACT

This study explores the use of nonuniform fast spherical Fourier transforms on meteorological data that are arbitrarily distributed on the sphere. The applicability of this methodology in the atmospheric sciences is demonstrated by estimating spectral coefficients for nontrivial subsets of reanalysis data on a uniformly spaced latitude–longitude grid, a global cloud resolving model on an icosahedral mesh with 3-km horizontal grid spacing, and for temperature anomalies from arbitrarily distributed weather stations over the United States. A spectral correction technique is developed that can be used in conjunction with the inverse transform to yield data interpolated onto a uniformly spaced grid, with optional triangular truncation, at reduced computational cost compared to other variance conserving interpolation methods, such as kriging or natural spline interpolation. The spectral correction yields information that can be used to deduce gridded observational biases not directly available from other methods.


## 1. Introduction

Spectral coefficient and spectral density estimation are considered core techniques in geophysical data analysis. Spectra serve to concisely summarize variability as a function of spatial or temporal scale in both one-dimensional


and multidimensional data. In geophysical time series, for example, the strongest spectral peaks are often observed at frequencies corresponding to the daily and annual cycles. Spectra also capture aspects of variability intrinsic to the system being sampled, for example in the atmospheric sciences, the “weather,” “macroweather,” and “climate” regimes (Lovejoy and Schertzer 2013). Spectral coefficients are easily estimated for data with a uniform sampling frequency using the fast Fourier transform (FFT), a discrete computational algorithm that decreases computational complexity to  $\mathcal{O}(N \log N)$  versus brute force spectral estimation, which has a computational complexity of  $\mathcal{O}(N^2)$ .

Analogous to temporal spectra, horizontal (or wave-number) spectra can also be calculated using the FFT on data that are evenly geographically spaced to examine the signal’s power as a function of wavelength. This

---

 Denotes content that is immediately available upon publication as open access.

---

 Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/MWR-D-16-0259.s1>.

---

*Corresponding author:* Nicholas R. Cavanaugh, [nrcavanaugh@lbl.gov](mailto:nrcavanaugh@lbl.gov)

DOI: 10.1175/MWR-D-16-0259.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

technique can be extended to two or more dimensions by simply using higher-dimensional FFTs. For data on the sphere, the combination of the FFT across latitudinal bands and a (fast) Legendre transform across meridional bands leads to the (fast) spherical Fourier transform (FSFT), which has been shown to be very useful in the atmospheric sciences to yield estimates of power or energy as a function of spherical wavenumber (Baer 1972). In particular, spherical harmonic analysis is often used to examine atmospheric spectral scaling properties that relate to turbulence theory (Lovejoy and Schertzer 2013) and to understand the fidelity of climate model simulations at small scales (Baldwin and Wandishin 2002; Skamarock 2004; Hamilton et al. 2008; Skamarock et al. 2014). FSFT algorithms have also played a prominent role in Navier–Stokes solvers in global atmospheric models (Bourke 1974; Bourke et al. 1977; Wedi et al. 2013), and are used, for example, in the current implementations of the NOAA GFS (NCEP 2016) and ECMWF IFS (ECMWF 2016) models.

FFT-based algorithms have achieved such ubiquity in the atmospheric sciences due primarily to the availability of data on uniformly spaced grids. Weather and climate models most often output data in this format and observations are commonly interpolated onto grids using one of many available interpolation algorithms [e.g., optimal interpolation (OI) or kriging (Gandin 1963)]. OI is a simple yet effective probabilistic method of geospatial interpolation relying on the covariance statistics of arbitrarily distributed data sources to provide the least squares optimal estimate of the expected value at unobserved spatial points of interest. We use OI as a comparable method of interpolating unevenly spaced data in this study since OI generally outperforms other interpolation methods when interpolating common meteorological variables (Hofstra et al. 2008).

While spherical harmonic spectral estimation on nonuniformly spaced data using optimization techniques is possible, the sheer quantity and dimensionality of geophysical data makes the problem computationally intractable at scale. However, recent advances in computational mathematics have led to the introduction and distribution of nonuniform ( $N$ ) fast Fourier algorithms, namely the NFFT and the NFSFT (Potts et al. 1998; Keiner et al. 2009, as well as many others), which can yield transform estimates from data with an arbitrary sampling structure.

In this study, we illustrate the estimation of spherical harmonic spectral coefficients using the NFSFT algorithm, which utilizes both nonuniform fast Fourier and Legendre transforms. We first demonstrate the flexibility of the algorithm using uniformly spaced atmospheric data from reanalysis by estimating spectral coefficients

globally, hemispherically, and over global ocean and land taken separately. Second, we demonstrate the power and extensibility of the method by estimating spectral coefficients directly from a global cloud resolving model on an icosahedral mesh with 3-km horizontal grid spacing, followed by inverse transforming (iNFSFT) the data onto a uniformly spaced latitude–longitude grid. Finally, we demonstrate the use of the NFSFT in estimating spectral coefficients directly from weather observations placed at arbitrary geographic locations. The inverse transform, followed by a spectral correction (reweighting) that accounts for the original sampling distribution, yields a gridding algorithm that is more computationally efficient than kriging/OI. While we discuss applications of this algorithm within the context of the atmospheric sciences, we also note that these methods are broadly applicable in the geosciences and elsewhere. Throughout the study, we utilize power spectra strictly to summarize the results of spectral coefficient estimation.

## 2. Methodology and data

### a. Fast, nonuniform spherical harmonic transforms

Any square integrable function  $f(\theta, \phi)$  distributed on the sphere can be decomposed into an infinite sum of Laplace spherical harmonic functions,  $Y_{lm}$ , such that

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \hat{f}_{lm} Y_{lm}(\theta, \phi), \quad (1)$$

where  $\theta$  is latitude,  $\phi$  is longitude, and  $l$  and  $m$  are the indices of the spectral coefficient matrix  $\hat{f}$ . A lower-resolution representation of  $f(\theta, \phi)$  can also be constructed by summing  $l \in [0, M]$ , where  $M$  corresponds to the desired total wavenumber (triangular) truncation. A spectrum,  $E_f(l)$ , can be constructed as a function of total wavenumber  $l$  such that

$$E_f(l) = \sum_{m=-l}^l c_{lm} |\hat{f}_{lm}|^2, \quad (2)$$

where  $c_{lm}$  is a normalizing constant. In the atmospheric sciences,  $f$  is often taken to be wind speed or potential temperature, making Eq. (2) proportional to kinetic or potential energy, respectively. Provided a finite data sample of  $f(\theta, \phi)$  at arbitrary locations on the sphere, a NFSFT algorithm can quickly and efficiently provide estimates of the spherical harmonic expansion coefficients  $\hat{f}_{lm}$ , from which a spectrum can be estimated.

For this study, we utilize the NFFT3C package (Keiner et al. 2009), which includes an implementation of the NFSFT algorithm described by Potts et al. (1998). The Potts et al. (1998) NFSFT algorithm leverages a

decomposition of the spherical harmonic transform into a Fourier transform in the zonal direction and a Legendre transform in the meridional direction. For the zonal component, the NFSFT utilizes an NFFT, which approximates a DFT to arbitrary precision by 1) spreading information from  $N$  nonuniformly spaced data points, using a convolution function with a known spectral transform (e.g., a Gaussian), onto a uniform grid with  $M$  points; 2) utilizing a standard FFT to transform the convolved data to spectral space; and 3) using the convolution theorem to approximately remove the spectral effects of the convolution (Dutt and Rokhlin 1993; Kunis and Potts 2003; Keiner et al. 2009). The NFFT approximation, for which errors are typically at the level of machine-precision, has a computational complexity of  $\mathcal{O}(M \log M + N \log \varepsilon)$ , where  $\varepsilon$  is the target accuracy of the approximation. Assuming  $M \approx N$ , this is faster than the  $\mathcal{O}(N^2)$  complexity required for a standard DFT. For the meridional component, the NFSFT utilizes a fast Legendre transform for data at arbitrary locations in which 1) Chebyshev transform coefficients are efficiently calculated by taking advantage of a recurrence property of Chebyshev polynomials that allows fast polynomial multiplication using fast, discrete cosine transformations; and 2) a change of basis from Chebyshev to Legendre polynomials (Potts et al. 1998). Altogether, the NFSFT algorithm requires  $\mathcal{O}(N^2 \log^2 N + N \log \varepsilon)$  calculations compared to the  $\mathcal{O}(N^3)$  calculations for a standard, discrete SFT.

### b. Interpolating nonuniform data

Data taken at arbitrary locations can be considered the product of indicator functions (centered at the data locations) and a continuous field  $f$ . When considering a sample of  $f$  taken at arbitrary locations on the sphere (identified here using dot notation, e.g.,  $\dot{f}$ ),

$$\dot{f}(\theta, \phi) = f(\theta, \phi) \cdot \dot{g}(\theta, \phi), \quad (3)$$

where  $\dot{g}(\theta, \phi) = 1(\theta, \phi)$  is an indicator function equal to 1 at the sample points, and zero everywhere else. By the convolution theorem, the SFT of  $\dot{f}$  is the convolution of  $f$  and  $\dot{g}$ :  $\hat{\dot{f}} = \hat{f}_{lm} * \hat{\dot{g}}_{lm}$ , where the asterisk denotes convolution. The inverse transform can be used to interpolate  $f$  to an alternate set of points [e.g., as demonstrated in Keiner et al. (2009)]; however, simple application of the inverse transform on  $\hat{\dot{f}}$  neglects the effect of  $\dot{g}$  on the spherical harmonic coefficients. The convolution theorem offers a simple means to correct for the nonuniform location of points  $\dot{g}$ , such that

$$f_{\text{alt}} = \frac{\tilde{f}_{\text{alt}}}{\tilde{g}_{\text{alt}}}, \quad (4)$$

where alt indicates  $\dot{f}$  represented on the alternate set of points and the tilde ( $\sim$ ) indicates the result of the inverse transform. When  $\dot{g}$  is densely sampled on a uniformly spaced grid,  $\tilde{g}$  is equal to 1 everywhere on that same uniformly spaced grid; however, for an arbitrary set of points,  $\tilde{g}_{\text{alt}}$  is a wavenumber- $M$ -dependent representation of  $\dot{g}$  at the alternate point locations that is related to the original, spatially varying sampling density. This factor provides the appropriate weights for  $\tilde{f}_{\text{alt}}$  to represent  $\dot{f}$  on the alternate spatial points. Henceforth, we will refer to  $f_{\text{alt}}$  as *corrected* and  $\tilde{f}_{\text{alt}}$  as *uncorrected*. In this text, we only consider uniformly spaced latitude–longitude grids as an alternate set of spatial points, but recognize that Eq. (4) holds for any arbitrary set of alternate points.

### c. Data

Reanalysis data used to demonstrate the horizontal spectral coefficient estimation on both densely sampled and geographically subsampled (sparse) uniformly spaced grids come from the Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) at the 200-hPa pressure level. CFSR has a uniform horizontal grid spacing of  $0.5^\circ$  and we estimate spectral coefficients using daily averages.

The Model for Prediction Across Scales (MPAS) cloud resolving model, which has 65 536 002 cells on an icosahedral mesh with 3-km horizontal grid spacing, is used to illustrate spectral analysis for nonuniformly spaced dense data on the sphere. We operate on 41 time slices of 200-hPa winds, each separated by 6 h, beginning on 0000 UTC 20 January and spanning to 0000 UTC 30 January 2009 to stay comparable with Skamarock et al. (2014). Further model details can be obtained from Skamarock et al. (2014).

Weather station temperature anomalies are taken from the Hadley Centre's quality controlled subset of the Integrated Surface Database at 0000 UTC (HadISD; Dunn et al. 2012). We remove all data that have been flagged for quality control. We then calculate anomaly time series by removing the seasonal cycle, estimated from a centered 5-day running mean (with missing data removed) for stations that are at least 90% complete over 1981–2010.

## 3. Uniformly spaced grids

We first demonstrate spherical harmonic spectral coefficient estimation on a dense and uniformly sampled two-dimensional latitude–longitude grid: CFSR 200-hPa  $u$  and  $v$  winds during January 2005. We illustrate one time slice of these fields in Figs. 1a and 1b and illustrate spectra for the total winds in Fig. 1c. The spectral estimation procedure for data of this type is usually accomplished using discrete Fourier techniques. In the case of complete and uniformly spaced data, the NFSFT will

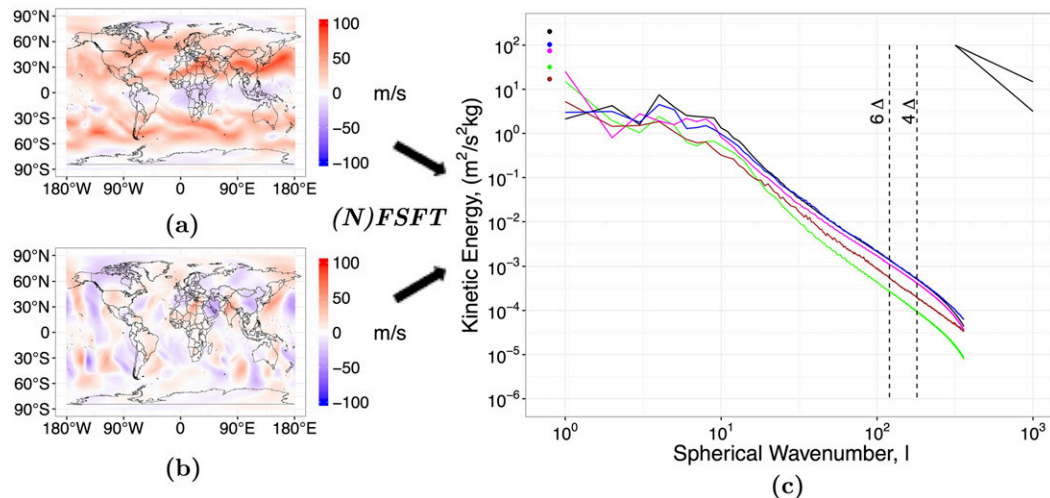


FIG. 1. Spectral estimation on a uniformly spaced latitude–longitude grid: CFSR 0.5°. (a)  $f_u(\theta, \phi)$ : 200-hPa  $u$  wind for 0000 UTC 1 Jan 2005. (b)  $f_v(\theta, \phi)$ : 200-hPa  $v$  wind for 0000 UTC 1 Jan 2005. (c)  $E_{f_u+v}(l)$ : 200-hPa KE spectra averaged over January 2005 for global (black), Northern Hemisphere (magenta), Southern Hemisphere (green), global land (brown), and global ocean (blue). Points indicate wavenumber-0 values. Vertical lines denote model effective resolution, as discussed in text. Diagonal lines show  $k^{-3}$  and  $k^{-5/3}$  scaling.

yield the same result as the standard FSFT, shown in Fig. 1c. Skamarock (2004) demonstrated that numerical model diffusion schemes yield unphysical spectral characteristics beyond the wavenumber associated with 4–6 times the grid scale (illustrated as vertical lines). These characteristics manifest themselves as a departure from the canonical  $k^{-3}$  or  $k^{-5/3}$  scaling regimes related to atmospheric turbulence (illustrated as diagonal lines in the top-right corner)—we revisit this assertion in section 5.

Next, we demonstrate the use of the NFSFT in estimating spectral coefficients derived from each hemisphere, taken separately. The NFSFT requires only the raw data input on the sampled half-sphere for the spectral estimation. In Fig. 1c, we show that the energy estimated for the Northern and Southern Hemispheres are similar both to each other and to the spectra for the entire sphere, with the Southern (summertime) Hemisphere showing a greater degree of  $k^{-5/3}$  scaling.

Finally, we illustrate in Fig. 1c spectra derived from a nontrivial subsampling of uniformly spaced data on the sphere, that is, for the global ocean and global land taken separately. The spectrum for global ocean behaves similarly to the spectrum for all data, initially decaying at near  $k^{-3}$  but flattening somewhat near wavenumber 50. The spectrum for global land, however, decays at close to  $k^{-2.4}$  over the entire inertial range. The authors do note, however, that these differences are subject to uncertainty estimates and require further investigation using higher-resolution data and optimal spectral estimation techniques.

While the FSFT algorithm is unable to estimate spectra for sparse matrices, we do acknowledge the methodological

developments of Harig et al. (2015, and references therein) and DelSole and Tippett (2015), which could be used to accomplish the same task. In particular, the problem of localized spherical spectral estimation has been studied in depth by Wiczorek and Simons (2005, 2007), who propose a family of spherical multitapers based on Slepian functions that can be used prior to spectral estimation using standard techniques. When performing spectral estimation on localized regions of the sphere, spherical harmonic basis functions are shown to be nonoptimal, since they are not strictly orthogonal on the region of interest. This nonoptimality contributes to spectral leakage in the estimated spectral coefficients. By contrast, the multitaper Slepian functions are orthogonal on both the whole sphere and the region of interest, and explicitly address the trade-off between the concentration of information in the spatial domain versus the concentration of information in the spectral domain through the number of taper functions, thereby mitigating or eliminating spectral leakage. Strictly speaking, the NFSFT methodology presented here maximizes concentration only in the spatial domain, and we therefore expect some level of nonoptimality (bias) in the estimated power spectra. We leave it up to the reader to determine the NFSFTs suitability in such situations.

#### 4. Nonuniformly spaced grids

In this section, we demonstrate spectral estimation directly from MPAS on an icosahedral mesh with 3-km horizontal grid spacing (Skamarock et al. 2014). MPAS



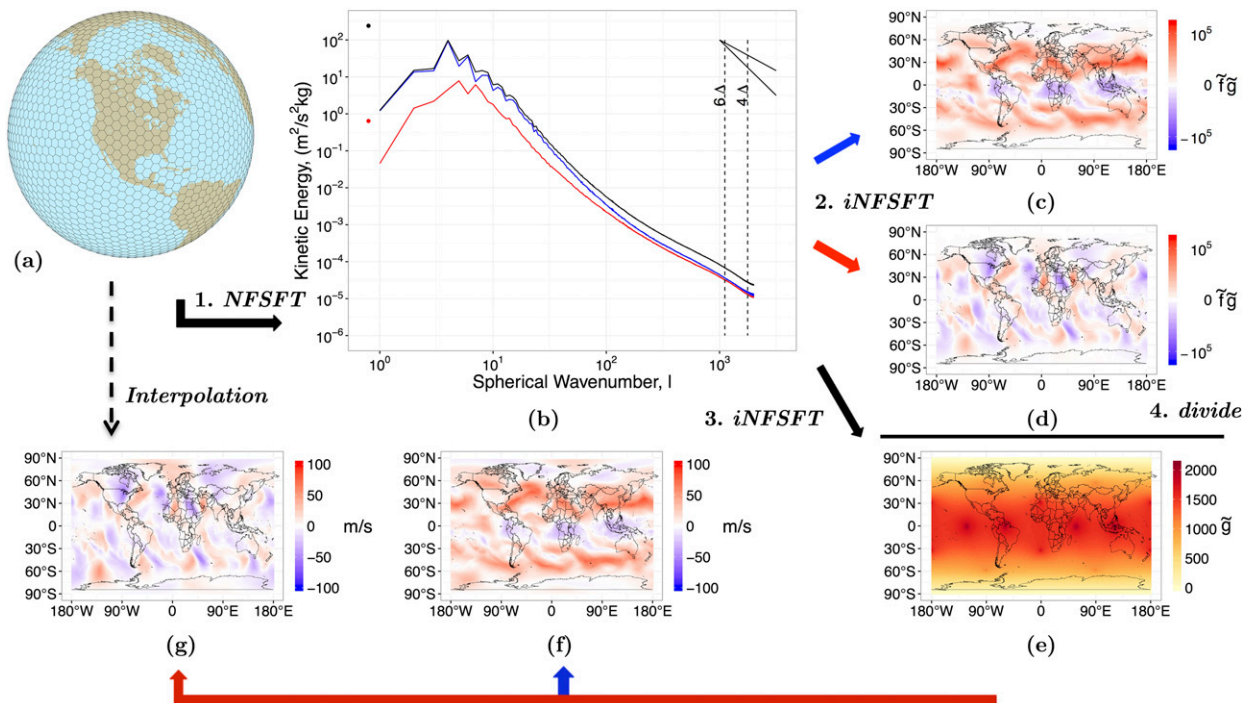


FIG. 2. Spectral estimation on an icosahedral mesh and interpolation to a uniformly spaced latitude–longitude grid: MPAS 3-km global cloud resolving model. See algorithm 1. (a) Schematic of MPAS icosahedral mesh. (b)  $E_f(l)$ : 200-hPa KE spectra estimated from the nonuniformly spaced data averaged over 41 (total) 6-h time slices for 200-hPa total wind (black), 200-hPa  $u$  wind (blue), and 200-hPa  $v$  wind (red). (c)  $\tilde{f}_{u,\text{alt}}(\theta, \phi)$ : uncorrected  $u$  winds transformed onto a  $1^\circ$  uniformly spaced latitude–longitude grid at T180 truncation. (d)  $\tilde{f}_{v,\text{alt}}(\theta, \phi)$ : as in (c), but for  $v$  winds. (e)  $\tilde{g}_{\text{alt}}(\theta, \phi)$ : spectral correction of the icosahedral mesh at T180 truncation. (f)  $f_{u,\text{alt}}(\theta, \phi)$ : corrected 200-hPa  $u$  winds on a  $1^\circ$  uniformly spaced latitude–longitude grid at T180 truncation. (g)  $f_{v,\text{alt}}(\theta, \phi)$ : as in (f), but for  $v$  winds. Points indicate wavenumber-0 values. Vertical and diagonal lines as in Fig. 1. The alternative cubic spline interpolation operation is indicated by the dashed arrow between (a) and (g).

performs numerical calculations on a global icosahedral mesh (Fig. 2a), which generally requires interpolation to a uniformly spaced grid for further analysis, including spectral estimation. This additional step can add both computational time and spectral bias if the method is not variance conserving. Here, the method is variance conserving and the rate limiting step for further analysis is the NFSFT estimation, which comes at a computational complexity of  $\mathcal{O}(N^2 \log^2 N)$  (Kunis and Potts 2003).

In Fig. 2b, we illustrate the kinetic energy spectra from MPAS averaged over 41 time slices for  $u$  and  $v$  winds at 200 hPa estimated via NFSFT. The kinetic energy for total winds (solid black) is directly comparable to Fig. 6 in Skamarock et al. (2014) (see online supplemental Fig. S1), demonstrating the accuracy of the NFSFT algorithm for global analysis. The largest differences between the spectra presented here and that of Fig. 6 in Skamarock et al. (2014), which are subtle, lie primarily in the largest and smallest wavenumbers. Skamarock et al.'s (2014) approach utilizes localized radial basis functions to reconstruct zonal and meridional velocities at the icosahedral cell centers, and then uses linear

interpolation to construct  $u$  and  $v$  winds on a uniformly spaced latitude–longitude grid. They then use the SPHEREPACK functions within NCL to estimate the spectral coefficients (W. C. Skamarock 2016, personal communication). We note that the spectral departure beyond  $4\Delta$  is stronger in Skamarock et al. (2014), suggesting that at least part of this spectral effect at small scales may be due to nonvariance conserving regridding algorithms. We also illustrate the individual  $u$ - and  $v$ -kinetic energy spectra in Fig. 2b.

In Figs. 2c and 2d, we show the unweighted wind fields estimated by iNFSFT, transformed onto a  $1^\circ$  uniformly spaced latitude–longitude grid at T180 truncation, and in Fig. 2e, we show the weights (spectral correction) estimated on the same grid. Figure 2e clearly demonstrates a higher observation density near the equator, as would be expected from representing a 3-km icosahedral mesh on a latitude–longitude grid. Also apparent from Fig. 2e are the six grid transition points that have higher observation densities. One of these points is depicted in our icosahedral network schematic (Fig. 2a) just west of Baja California, Mexico. Note that the units for Figs. 2c–e are

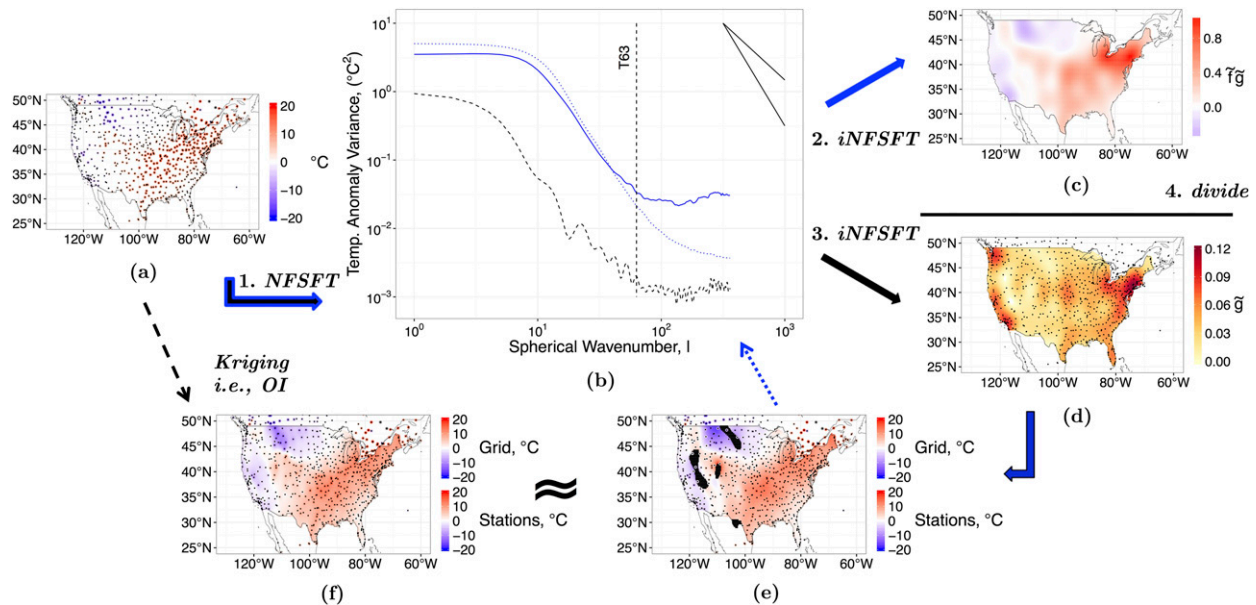


FIG. 3. Spectral estimation on a weather station network of arbitrarily located observations and interpolation to a uniformly spaced latitude–longitude grid: HadISD temperature anomalies over the United States. See algorithm 1. (a)  $\hat{f}(\theta, \phi)$ : station temperature anomalies for 0000 UTC 1 Jan 2005. (b)  $E(l)$ : 0000 UTC temperature variance spectra averaged over 2005 (solid blue), 0000 UTC network spectrum averaged over 2005 (dashed black), and corrected 0000 UTC temperature variance spectra averaged over 2005 (dotted blue). (c)  $\hat{f}_{\text{alt}}(\theta, \phi)$ : uncorrected temperature anomalies transformed onto a  $0.25^\circ$  uniformly spaced latitude–longitude grid at T63 truncation for 0000 UTC 1 Jan 2005. (d)  $\tilde{g}_{\text{alt}}(\theta, \phi)$ : spectral correction at T63 truncation. (e)  $\hat{f}_{\text{alt}}(\theta, \phi)$ : corrected temperature anomalies transformed onto a  $0.25^\circ$  uniformly spaced latitude–longitude grid at T63 truncation for 0000 UTC 1 Jan 2005. Black regions indicate negative or near-negative observation density corresponding to (d). (f) 0000 UTC temperature anomalies interpolated to a  $0.25^\circ$  uniformly spaced latitude–longitude grid using kriging/OI. In (a), (d), (e), and (f), black dots mark the locations of weather stations and in (a), (e), and (f) observed station temperatures are illustrated on the color scale as the gridded data. Vertical and diagonal lines are as in Fig. 1. The alternative kriging or OI operation is indicated by the dashed arrow between (a) and (f).

arbitrary and depend on the degree of triangular truncation  $M$ . By elementwise matrix division of the uncorrected wind fields (Figs. 2c and 2d) by the observation network correction (Fig. 2e), the corrected wind fields (Figs. 2f and 2g) can be recovered at the desired triangular truncation on a uniformly spaced latitude–longitude grid.

## 5. Arbitrary networks

We now direct our attention to estimating spectral coefficients directly from meteorological data with arbitrary network configurations. Here, we focus on a different scaling field—temperature anomalies. Figure 3a shows a high-quality subset of HadISD surface temperature anomaly observations at 0000 UTC 1 January 2005 between  $25^\circ$ – $50^\circ$ N and  $60^\circ$ – $130^\circ$ W. The NFSFT can directly estimate the temperature anomaly variance spectral coefficients from these observations, illustrated by the spectra with the solid blue line in Fig. 3b. Unlike the dense networks analyzed in sections 3 and 4, which yield energy spectra that decay monotonically as power laws, station data are clearly influenced by the network

configuration and tend to yield spectra that flatten at higher wavenumbers. In this case, we can see by analyzing the spectrum of the network (Fig. 3b, dashed black) that its structure is nontrivial and is very likely affecting the spectral behavior estimated from the temperature data. Taking cues from Skamarock (2004), we define the effective resolution of the meteorological network as the wavenumber in which the variance spectrum is no longer power-law scaling, but note that a more sophisticated methodology might be devisable. We move forward with our analysis at a triangular truncation of  $M = 63$ .

As in section 4, we can iNFSFT both  $\hat{f}$  and  $\tilde{g}$  to obtain the uncorrected temperature field (Fig. 3c) on a uniformly spaced latitude–longitude grid (in this case,  $0.25^\circ$  horizontal grid spacing at T63 truncation), as well as the observation weights (Fig. 3d). From Fig. 3d it is clear that observation density varies spatially, resulting in a nontrivial spectral correction. Comparison of Figs. 3c and 3d show strong spatial similarities in the patterns of  $\tilde{g}_{\text{alt}}$  and  $\hat{f}_{\text{alt}}$ , which indicates that the network density is modulating the interpolated field. We can correct for this modulation by dividing  $\hat{f}_{\text{alt}}$  by  $\tilde{g}_{\text{alt}}$ . At this point, we choose to remove the geographic locations that have

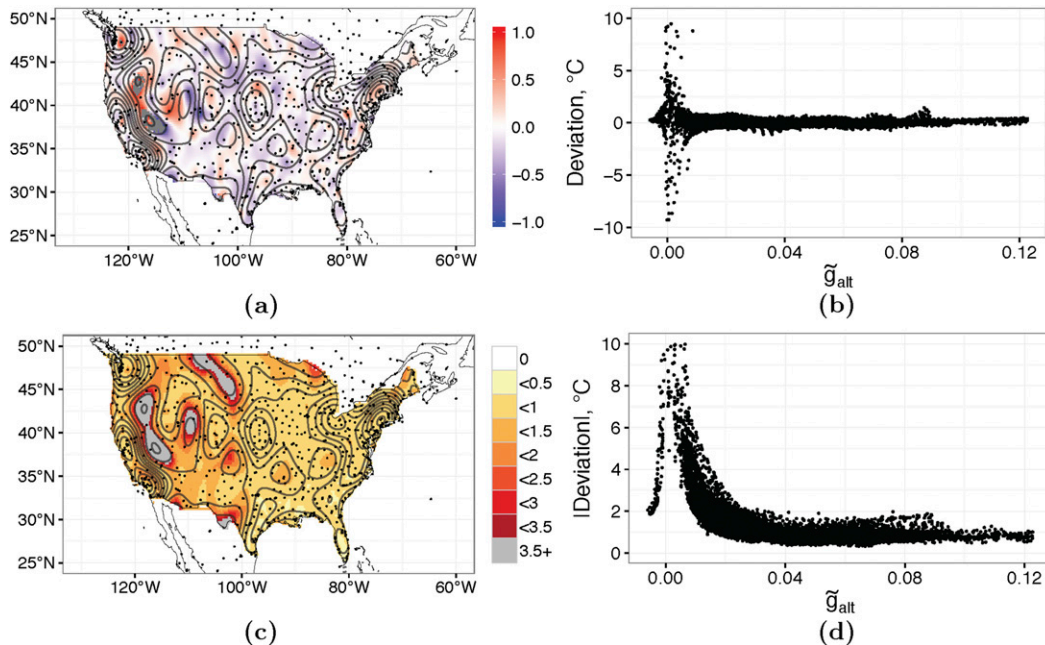


FIG. 4. Performance of spherical harmonic interpolation vs kriging. (a) Average deviation (color shaded) between T63 spherical harmonic interpolation vs kriging over 2005 estimated on a  $0.25^\circ$  uniformly spaced latitude–longitude grid. Average sampling density (contours) at T63 truncation over 2005, roughly corresponding to Fig. 3d. Bin width is 0.01. (b) Average deviation plotted against average sampling density. (c) As in (a), but for absolute deviation (color shaded) with contours as in (a). (d) As in (b), but for absolute deviation.

negative or near-negative weights, which lead to degeneracies in the corrected temperature field. The degeneracy arises when attempting to divide by  $\tilde{g}_{alt}$  when  $\tilde{g}_{alt} \rightarrow 0$ , leading  $\tilde{f}_{alt} \rightarrow \pm\infty$  in circumstances where station density is very low. We then elementwise divide the uncorrected temperature field by the network correction, yielding a corrected temperature field that has been interpolated by spherical harmonics onto a uniformly spaced  $0.25^\circ$  latitude–longitude grid over the United States. Finally (and optionally), we can estimate a corrected temperature variance spectrum from the corrected uniformly spaced temperature field by transforming only the areas with positive network densities—we illustrate this corrected spectrum in Fig. 3b (dotted blue).

Traditionally, kriging or OI, being the best least squares predictor, is used to transform data observed on arbitrary networks onto uniformly spaced latitude–longitude grids. Kriging has a computational complexity of  $\mathcal{O}(N^3)$ . Simple kriging assumes that the expectation of the field is known (in our case for temperature anomalies, zero) and relies on covariance information to yield a probabilistic prediction at unobserved locations. In this sense, simple kriging is quite similar mathematically to principal component analysis, where the covariance matrix is factored to yield orthogonal basis functions, termed empirical orthogonal functions (EOFs). An infinite

(sufficient) weighted sum of these EOFs yields a complete representation of the original data field. Similarly, a Laplace spherical harmonic expansion also has the completeness and orthogonality properties, signifying its equivalence with EOF-based approaches for spatial reconstruction on a global domain (Shen et al. 1994). Moreover, the Laplace spherical harmonics are the OI functions for homogeneous isotropic random fields on the sphere (Shen et al. 1994), and thus the back-transformed observations truncated at their effective resolution can be treated similarly to OI data if the appropriate conditions are met. The NFSFT algorithm, however, achieves this result with greater computational efficiency when using a large number of samples when compared to OI algorithms:  $[\mathcal{O}(N^2 \log^2 N)]$  vs  $\mathcal{O}(N^3)$ .

In Fig. 3f, we show the original data (Fig. 3a) optimally interpolated onto the  $0.25^\circ$  grid. We indicate the alternative OI operation in Fig. 3 by a dashed arrow between Figs. 3a and 3f. We argue that the OI fields are comparable to fields interpolated using spherical harmonics on variance conservation principals; however, we note that additional methodological research may be required to hone the technique. In Fig. 4a, we show the average deviation between OI 0000 UTC HadISD temperature anomaly fields with fields that have been interpolated via spherical harmonics over 2005. While these deviations do show some spatial structure (perhaps further

indicating a need for additional methodological research), they are in general small in amplitude and spatial scale compared to the climatological anomalies themselves, and are likely related in part to our T63 truncation choice. In Fig. 4c, we show the average absolute deviations between the OI dataset and our spherically interpolated dataset. In most regions, the average absolute deviation is less than  $1^{\circ}\text{C}$ , particularly in regions of high station density. It would be interesting for a future study to examine a full uncertainty analysis of the kriging methodology related to the validity of the kriged estimate.

In Figs. 4b and 4d we show average deviation and absolute deviation as a function of the spectral sampling density (the contours in Figs. 4a and 4c). These figures indicate that the difference between the spectral representation and the kriged representation increases as the spectral station density (correction) approaches 0. This is to be expected, since  $\hat{f}_{\text{alt}} \rightarrow \pm\infty$  as  $\hat{g}_{\text{alt}} \rightarrow 0$  as a consequence of spectral transform degeneracy in instances of (near) zero sampling density. In these same circumstances, that is the case where the interpolation is probabilistically uninformed, kriging yields an estimate of  $f$  that converges on 0. The degenerative zones, shown in Fig. 3e in black and corresponding to the regions of high deviation and absolute deviation in Figs. 4b and 4d, should be considered regions in which both the kriging and spherical harmonic interpolation methodologies are unlikely to be valid.

## 6. Discussion and conclusions

In this study, we demonstrate the use of the NFSFT for spherical harmonic spectral estimation by applying the algorithm directly to arbitrarily sampled atmospheric data. We illustrate the utility of this methodology by estimating spectral coefficients on subsamples of reanalysis data on a uniformly spaced grid, MPAS model data on an icosahedral mesh, and weather station data directly.

Spectral analysis is a common tool in geophysics involving a coordinate transformation from data in real space (i.e., latitude and longitude) to data in a complex spectral space. Numerical atmospheric models are often truncated at a specific wavenumber that corresponds to the smallest resolved spatial scale; we analogously show that weather observations can be treated in the same manner using the NFSFT. In this sense, the NFSFT algorithm allows observations to be de-noised and spectrally truncated at their effective resolution, perhaps providing better quality observations for ingestion by the assimilation schemes of global reanalyses. The NFSFT back transform has the same desirable properties as the

forward transform (i.e., the flexibility of arbitrary spatial configurations); as a result, arbitrarily sampled observations can be back transformed onto uniformly spaced grids and spectrally corrected for ease of use. With respect to the steps enumerated in Figs. 2 and 3, the algorithm is as follows:

### Algorithm 1 Spectral Regrid

```

input:  $\hat{f}(\theta, \phi)$ 
output:  $f_{\text{alt}}(\theta, \phi)$ 
begin
  1: analysis:
     $\hat{f}(\theta, \phi) \rightarrow \hat{f}_{lm}, \hat{g}(\theta, \phi) \rightarrow \hat{g}_{lm}$  via NFSFT
  2: signal synthesis:
     $\hat{f}_{lm} \rightarrow \hat{f}_{\text{alt}}(\theta, \phi)$  via iNFSFT
  3: network synthesis:
     $\hat{g}_{lm} \rightarrow \hat{g}_{\text{alt}}(\theta, \phi)$  via iNFSFT
  4: correction/deconvolution:
     $\hat{f}_{\text{alt}}(\theta, \phi)/\hat{g}_{\text{alt}}(\theta, \phi) = f_{\text{alt}}(\theta, \phi)$ 
end

```

Spectra are often used to help understand the behavior of the turbulent regimes that describe large-scale atmospheric dynamics (Lovejoy and Schertzer 2013). Nonuniform spectral methods allow for a more systematic examination of these turbulent regimes because the input data are not required to be densely populated or uniformly sampled. Beyond the simple geographic subsampling shown here, data can also be selected based on dynamical considerations (e.g., convective vs non-convective, low pressure vs high pressure), potentially yielding new fundamental insights about the nature of atmospheric dynamics and turbulence.

Finally, the NFSFT algorithm can be readily applied to climate model simulations computed on non-uniformly spaced or variable-mesh grids without having to first interpolate data onto a uniformly spaced grid. This methodology may help alleviate the practical burden of dealing with data computed on more sophisticated grids, which is becoming increasingly common, and may come at a reduced computational cost and bias compared to interpolation onto uniformly spaced grids.

Wieczorek and Simons (2005, 2007) have shown that spherical harmonic basis functions are nonoptimal for power spectral estimation on localized regions because they are nonorthogonal within localized domains. They argue that optimal spectral estimation cannot be performed without first multiplying the data by a specially designed multitaper before spectral estimation. The observation density-related spectral correction  $\hat{g}$  operates very much like a taper function, and can be optimized to various levels of smoothness (e.g., through spectral truncation) or to be orthogonal on the region of interest; however, a full mathematical treatment of these



procedures is beyond the scope of this paper. We, therefore, leave it up to the reader to decide whether the NFSFT methodology is suitable for their purposes.

The NFSFT algorithm is also just one in a growing suite of nonuniform fast spectral algorithms (Keiner et al. 2009). Many of the concepts documented in this study also follow for temporal spectral analysis, where the NFFT can be used to provide information about systems that have been nonuniformly sampled in time (e.g., biological systems). By weaving these components together, complete (space–time) spectral representations can be efficiently estimated for any arbitrarily sampled multidimensional system, with the atmosphere being just one of many examples.

**Acknowledgments.** This research was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy Regional and Global Climate Modeling Program (RGCM) and used resources of the National Energy Research Scientific Computing Center (NERSC), which is also supported by the Office of Science of the U.S. Department of Energy under Contract DE-AC02-05CH11231. The MPAS simulations were performed on the Yellowstone supercomputer at the National Center for Atmospheric Research. The authors also thank Editor Hillary Weller and two anonymous reviewers for their meaningful contribution to the study.

## REFERENCES

- Baer, F., 1972: An alternate scale representation of atmospheric energy spectra. *J. Atmos. Sci.*, **29**, 649–663, doi:[10.1175/1520-0469\(1972\)029<0649:AASROA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1972)029<0649:AASROA>2.0.CO;2).
- Baldwin, M. E., and M. S. Wandishin, 2002: Determining the resolved spatial scales of Eta Model precipitation forecasts. *19th Conf. on Weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., 3.2. [Available online at [https://ams.confex.com/ams/SLS\\_WAF\\_NWP/techprogram/paper\\_47735.htm](https://ams.confex.com/ams/SLS_WAF_NWP/techprogram/paper_47735.htm).]
- Bourke, W., 1974: A multi-level spectral model. I. Formulation and hemispheric integrations. *Mon. Wea. Rev.*, **102**, 687–701, doi:[10.1175/1520-0493\(1974\)102<0687:AMLSMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1974)102<0687:AMLSMI>2.0.CO;2).
- , B. McAvaney, K. Puri, and R. Thurling, 1977: Global modelling of atmospheric flow by spectral methods. *General Circulation Models of the Atmosphere*, J. Chang, Ed., Academic Press, 267–324.
- DelSole, T., and M. K. Tippett, 2015: Laplacian eigenfunctions for climate analysis. *J. Climate*, **28**, 7420–7436, doi:[10.1175/JCLI-D-15-0049.1](https://doi.org/10.1175/JCLI-D-15-0049.1).
- Dunn, R. J. H., K. M. Willett, P. W. Thorne, E. V. Woolley, I. Durre, A. Dai, D. E. Parker, and R. S. Vose, 2012: HadISD: A quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. *Climate Past*, **8**, 1649–1679, doi:[10.5194/cp-8-1649-2012](https://doi.org/10.5194/cp-8-1649-2012).
- Dutt, A., and V. Rokhlin, 1993: Fast Fourier transforms for non-equispaced data. *SIAM J. Sci. Comput.*, **14**, 1368–1393, doi:[10.1137/0914081](https://doi.org/10.1137/0914081).
- ECMWF, 2016: ECMWF IFS documentation—Cy41r2, operational implementation 8 March 2016. Part III: Dynamics and numerical procedures. ECMWF, 31 pp. [Available online at <https://www.ecmwf.int/sites/default/files/elibrary/2016/16647-part-iii-dynamics-and-numerical-procedures.pdf>.]
- Gandin, L. S., 1963: *Objective Analysis of Geophysical Fields*. Israeli Program for Scientific Translations, 242 pp.
- Hamilton, K., Y. O. Takahashi, and W. Ohfuchi, 2008: Mesoscale spectrum of atmospheric motions investigated in a very fine resolution global general circulation model. *J. Geophys. Res.*, **113**, D18110, doi:[10.1029/2008JD009785](https://doi.org/10.1029/2008JD009785).
- Harig, C., K. W. Lewis, A. Plattner, and F. J. Simons, 2015: A suite of software analyzes data on the sphere. *Eos, Trans. Amer. Geophys. Union*, **96**, 1–10, doi:[10.1029/2015EO025851](https://doi.org/10.1029/2015EO025851).
- Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei, 2008: Comparison of six methods for the interpolation of daily, European climate data. *J. Geophys. Res.*, **113**, D21110, doi:[10.1029/2008JD010100](https://doi.org/10.1029/2008JD010100).
- Keiner, J., S. Kunis, and D. Potts, 2009: Using NFFT 3—A software library for various nonequispaced fast Fourier transforms. *ACM Trans. Math. Software*, **36**, 1–30, doi:[10.1145/1555386.1555388](https://doi.org/10.1145/1555386.1555388).
- Kunis, S., and D. Potts, 2003: Fast spherical Fourier algorithms. *J. Comput. Appl. Math.*, **161**, 75–98, doi:[10.1016/S0377-0427\(03\)00546-6](https://doi.org/10.1016/S0377-0427(03)00546-6).
- Lovejoy, S., and D. Schertzer, 2013: *The Weather and Climate*. Cambridge University Press, 475 pp.
- NCEP, 2016: The Global Forecast System (GFS)—Global Spectral Model (GSM). Tech. Rep., Global Climate and Weather Modeling Branch. [Available online at <http://www.emc.ncep.noaa.gov/GFS/doc.php>.]
- Potts, D., G. Steidl, and M. Tasche, 1998: Fast and stable algorithms for discrete spherical Fourier transforms. *Linear Algebra Appl.*, **275–276**, 433–450, doi:[10.1016/S0024-3795\(97\)10013-1](https://doi.org/10.1016/S0024-3795(97)10013-1).
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi:[10.1175/2010BAMS3001.1](https://doi.org/10.1175/2010BAMS3001.1).
- Shen, S. S. P., G. R. North, and K. Y. Kim, 1994: Spectral approach to optimal estimation of the global average temperature. *J. Climate*, **7**, 1999–2007, doi:[10.1175/1520-0442\(1994\)007<1999:SATOE0>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<1999:SATOE0>2.0.CO;2).
- Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032, doi:[10.1175/MWR2830.1](https://doi.org/10.1175/MWR2830.1).
- , S.-H. Park, J. B. Klemp, and C. Snyder, 2014: Atmospheric kinetic energy spectra from global high-resolution non-hydrostatic simulations. *J. Atmos. Sci.*, **71**, 4369–4381, doi:[10.1175/JAS-D-14-0114.1](https://doi.org/10.1175/JAS-D-14-0114.1).
- Wedi, N., M. Hamrud, and G. Mozdzynski, 2013: A fast spherical harmonics transform for global NWP and climate models. *Mon. Wea. Rev.*, **141**, 3450–3461, doi:[10.1175/MWR-D-13-00016.1](https://doi.org/10.1175/MWR-D-13-00016.1).
- Wieczorek, M. A., and F. J. Simons, 2005: Localized spectral analysis on the sphere. *Geophys. J. Int.*, **162**, 655–675, doi:[10.1111/j.1365-246X.2005.02687.x](https://doi.org/10.1111/j.1365-246X.2005.02687.x).
- , and —, 2007: Minimum-variance multitaper spectral estimation on the sphere. *J. Fourier Anal. Appl.*, **13**, 665–692, doi:[10.1007/s00041-006-6904-1](https://doi.org/10.1007/s00041-006-6904-1).