The Astrophysical Journal Supplement Series, 260:15 (19pp), 2022 May https://doi.org/10.3847/1538-4365/ac5259


# Time-domain Deep-learning Filtering of Structured Atmospheric Noise for Ground-based Millimeter Astronomy

Alejandra Rocha-Solache, Iván Rodríguez-Montoya, David Sánchez-Argüelles, and Itziar Aretxaga
Instituto Nacional de Astrofísica, Óptica y Electrónica, Apdo. Post. 51 y 216, 72000 Puebla, Pue., Mexico; irodriguez@inaoep.mx
*Received 2020 October 21; revised 2022 January 14; accepted 2022 February 2; published 2022 May 5*

## Abstract

The complex physics involved in atmospheric turbulence makes it very difficult for ground-based astronomy to build accurate scintillation models and develop efficient methodologies to remove this highly structured noise from valuable astronomical observations. We argue that a deep-learning approach can bring a significant advance to treat this problem because of deep neural networks' inherent ability to abstract nonlinear patterns over a broad scale range. We propose an architecture composed of long short-term memory cells and an incremental training strategy inspired by transfer and curriculum learning. We develop a scintillation model and employ an empirical method to generate a vast catalog of atmospheric-noise realizations and train the network with representative data. We face two complexity axes: the signal-to-noise ratio (S/N) and the degree of structure in the noise. Hence, we train our recurrent network to recognize simulated astrophysical pointlike sources embedded in three structured-noise levels, with a raw-data S/N ranging from 3 to 0.1. We find that a slow and repetitive increase in complexity is crucial during training to obtain a robust and stable learning rate that can transfer information through different data contexts. We probe our recurrent model with synthetic observational data, designing alongside a calibration methodology for flux measurements. Furthermore, we implement traditional matched filtering (MF) to compare its performance with our neural network, finding that our final trained network can successfully clean structured noise and significantly enhance the S/N compared to raw data and in a more robust way than traditional MF.

*Unified Astronomy Thesaurus concepts:* Millimeter astronomy (1061); Atmospheric scintillation (117); Diffuse radiation (383); Humidity (764); Time series analysis (1916); Red noise (1956); Neural networks (1933); Astronomy data modeling (1859); Astrostatistics techniques (1886)

## 1. Introduction

Millimeter waves offer a unique astronomical window to study very far and dusty regions of the universe. Coincidentally, millimeter waves suffer from absorption and reemission of the water vapor molecules composing Earth's atmosphere. At millimeter waves, there are $z \lesssim 6$ confirmed dusty star-forming galaxies, and it is possible to observe farther sources if dust is present (Blain et al. 2002; Casey et al. 2014; Zavala et al. 2018). Meanwhile, atmospheric turbulence distorts these far-traveling electromagnetic waves along their path through Earth's troposphere ($\lesssim$10 km), before reaching their final ground-based telescope destination. Nowadays, large-diameter telescopes and large-format continuum cameras are operational and on course to produce sky surveys with unprecedented extent and depth (Holland et al. 2013; Brien et al. 2018; Bryan et al. 2018; Montaña et al. 2019). However, the atmospheric turbulence still represents a critical challenge for any ground-based millimeter-astronomy campaign. The present situation raises the question of whether recent advances in deep learning can provide an attractive approach to deal with the highly structured atmospheric noise recorded in millimeter-wave continuum observations.

In the past two decades, observations in the far-infrared (IR) wavelengths have attracted increasingly more interest from different astronomy fields, probably due to two main particularities: Due to the abundance of dust (especially in rich star-forming regions) in the universe, UV and visible light are likely to be absorbed by dust and reemitted at infrared and millimeter wavelengths; nearly half the luminosity radiated from astrophysical objects in the universe is only accessible at millimeter-wave windows. Second, due to an observational effect known as the negative k-correction, the expansion of the universe does not dilute the thermal-dust energy flux observed at millimeter and submillimeter waves; consequently, very high-redshift galaxies can be accessible to millimeter astronomy with equal ease to lower-redshift sources. These two unique advantages make the millimeter-wave window an exciting opportunity to study the star formation history and the evolution of large-scale structures of the universe.

To continue the far millimeter-universe exploration, we need to count on large-area, deep surveys, compounding a representative sample of bright, dim, and high-redshift sources. Very-large-area surveys assembled by space-based missions like Spitzer, Planck, and Herschel (Aghanim et al. 2015; Ade et al. 2016; Martinache et al. 2018) compose a vast catalog of dusty star-forming galaxies that are being followed up by spectroscopic measurements; however, they lack the resolution to overcome source multiplicity, and, by design, they could only select the brightest sources, thereby, biasing population estimations. On the other hand, although ground-based telescopes have much better sensitivity and resolution, they are limited by comparatively lower mapping speeds than space-based telescopes. Therefore, it is imperative to count on large-aperture telescopes, which, coupled to multidetector instruments, can realize large-area, high-sensitivity sky maps in reduced scanning time.

The Large Millimeter Telescope (LMT), which is the largest millimeter-wave single-dish telescope, with an aperture diameter of 50m, and a field of view of 4″, contributes to this effort (Hughes et al. 2010). At 4600 m above sea level, on the summit of Sierra Negra Volcano in Mexico, the dry weather allows the LMT to perform high-quality astronomical observations, with winter 220 GHz opacities low as $\tau = 0.025$ at wavelengths of ∼1–3 mm. These excellent climate conditions are particularly necessary for infrared and millimeter astronomy to minimize water molecule absorption and anomalous emission effects. One of the LMT's key instruments is the Aztronomical Thermal Emission Camera (AzTEC) (Wilson et al. 2008), an array of 144 silicon nitride micromesh bolometers distributed into a hexagonal close-packed geometry, designed to perform continuum measurements in the 1.1 mm atmospheric window. With an average mapping speed of 20 arcmin$^2$ mJy$^{-2}$ hr$^{-1}$ (under favorable climate conditions), during the years that it was operational, AzTEC was capable of producing maps with ∼500 arcmin$^2$ angular size and 0.2 mJy per-beam depth, in ∼50 hr of total integration time.

A new generation of millimeter-wave continuum cameras is currently scheduled to be installed on the LMT, with significantly improved technologies, such as lumped-element kinetic inductance detectors (LEKIDs) (Austermann et al. 2018; Castillo-Dominguez et al. 2018). The first camera is the Mexico–UK Sub-millimeter Camera for Astronomy (MUSCAT), a large-format millimeter-wave camera made of 1600 horn-coupled LEKIDs cooled to an operating temperature of 100 mK (Castillo-Dominguez et al. 2018). This focal-plane array is designed to observe in the 1.1 mm bandwidth and cover the full 4′ of LMT's field of view. With a design mapping speed of 3 deg$^2$ mJy$^{-2}$ hr$^{-1}$, MUSCAT on the LMT would enable us to map in about 40 hr a 160 deg$^2$ region, with a 6″ resolution and 1 mJy depth. The second camera is TolTEC, a large-format millimeter-wave imaging polarimeter made of ∼7000 high-responsivity LEKIDs, arranged in three focal planes for the 1.1, 1.4, and 2.0 mm bandwidths, and cooled to a 150 mK operating temperature (DeNigris et al. 2020). With 1.9 deg$^2$ mJy$^{-2}$ hr$^{-1}$ as the design mapping speed for the 1.1 mm bandwidth, TolTEC would require about 40 hr to produce a 100 deg$^2$ multicolor map, with polarization information, 5″ resolution, and 1 mJy depth (at 1.1 mm).

These advanced instruments will produce valuable astronomical surveys whose raw data will yet be dominated by atmospheric emission. The first complication with the atmosphere is that it produces highly structured noise that, in some cases, is up to four orders of magnitude brighter than the astronomical emission of interest. The second complication, at least for infrared and millimeter waves, is a model scarcity in the literature—presumably due to the complex physics involved—that could be useful to simulate atmospheric noise in the time domain. The typical approach to clean atmospheric noise is to compute templates of the common-mode signals seen by all the detectors and then subtract them from the original time streams (Kovács 2008; Sayers et al. 2010). An equivalent approach is to perform principal component analysis (PCA) to remove the common-mode signals, the standard procedure in the AzTEC pipeline (Scott et al. 2008). However, this procedure could be problematic because, in some cases, it can remove astrophysical information too. Intrinsically, the problem demands an adequate treatment of randomly varying correlation lengths that cannot be easily handled by classical methods.

Aiming to solve this problem, some of us implemented independent component analysis in the map domain (Rodríguez-Montoya et al. 2018), a methodology that utilizes high-order statistical moments (negentropy) to exploit the richness of atmospheric-noise structure. However, it would be better to prevent (as much as possible) any contamination leakage to the map domain in the first place; thus, the cleaning step is more suitable at earlier stages in the domain of time. In this line of thinking, we see that the challenge of structured noise in time series does not pertain exclusively to millimeter astronomy; other fields have addressed similar issues. For example, the problem of noisy speech recognition has been addressed by various machine-learning strategies (see Zhang et al 2018 for a comprehensive review). Among a few examples in astrophysics, neural network applications have been useful in optical astronomy to correct atmospheric-turbulence-distorted images (Gómez et al. 2019) and in variable stars classification (Jamal & Bloom 2020) and have even succeeded in recovering gravitational waves from structured-noisy temporal series (George & Huerta 2018; Marulanda et al. 2020).

In this paper, we propose a deep-learning computational scheme to characterize the expected astronomical and atmospheric signals in the time domain. Specifically, we propose a structure made of long short-term memory units that can learn to identify astrophysical sources embedded in time series dominated by structured noise. To generate diverse sets of training data, we need to develop a physical model of turbulence to reproduce synthetic scintillation data. Moreover, we utilize real data taken with AzTEC to generate mock observations that are more representative of the kind of data to be produced by new-generation instrumentation. Given the physical and statistical complexity involved in this application and the inherently high computational demands, in this paper we aim to present proof of our concept by processing a single detector signal. Meanwhile, the number of detectors' scalability and map-making procedures are deferred to a subsequent paper.

The following section introduces the main aspects of ground-based millimeter-astronomy observations, which helps us to contextualize the statistical properties of the time streams used in our study. In Section 3, we review the deep-learning theory used to describe the building blocks of our proposed recurrent model. A more technical Section 4 is employed to detail our data preparation procedures, hyperparameter selection, and training strategies. We discuss most of our quantitative results in Section 5. Finally, Section 6 summarizes our conclusions.

## 2. Properties of Millimeter-wavelength Continuum Observations

In millimeter-continuum cameras, the detector's array and optical coupling are designed to match the angular size of the telescope beam. Hence, the full array does not entirely sample the sky brightness distribution at any given time. Consequently, a scanning pattern is executed by the telescope to perform a sequential integration recorded by all the detectors in temporal series. Each detector records a stream of time-ordered data that can be modeled as

$$d(t) = P(t) \cdot s(t) + A(t) + N(t), \tag{1}$$

where $s$ is the astronomical brightness distribution, $P$ the telescope-pointing matrix, $A$ the atmosphere emission, and $N$

the instrumental noise. In the AzTEC architecture, each bolometer couples to a negative temperature coefficient thermistor; hence, it records the sky brightness as a power decrease across the thermistor. The instrument electronics front-end measures the power signals across the thermistor, digitizes them at a 64 Hz sampling rate, and stores the time streams in electrical power units (nW). These values are converted later into radiation flux units,[1] using observations of calibration sources (planets and asteroids).

The data-reduction process is performed offline (after the observation time); it usually includes removing large spikes induced by cosmic rays, followed by an atmosphere-cleaning process. The pointing calibration data are then used to construct the pointing matrix and project the time streams into a map. Thus, every pixel in the final map comprises all the detector's temporal sampling average. This clean, reduced map is ready for astrophysical investigations. For an overview of data-reduction techniques used in millimeter astronomy, see Mairs et al. (2015), and for a complete description of the AzTEC pipeline, see Scott et al. (2008) and Rodríguez-Montoya et al. (2018).

In the following analysis, we consider only pointlike sources as our astrophysical target. A pointlike source is an astrophysical object with an angular extension significantly smaller than the telescope resolution; hence, in the map domain, they are imprinted as compact sources with the telescope's point-spread function morphology.

Considering typical observation parameters for AzTEC during the 2018 observation campaign (a scanning speed tuned to $50''\ \mathrm{s}^{-1}$ and an aperture diameter of 30 m), it leads to a pointlike source maximum temporal width of ~4.4 s, which we will use in Section 4.1 as the expected signal width to generate our synthetic data.

We can address two general types of millimeter observations: The most frequent are pointing calibrations; these are typically two-minute-long integrations on bright pointlike sources (a few janskys) used to calibrate the telescope position accuracy. The other type is science observations, consisting of integrations that can last from a few to hundreds of minutes. In most cases, science objects are much fainter than well-known calibration sources. However, clean and quick pointing calibrations are crucial to ensure the quality of scientific data and optimize the telescope's time. A successful deep-learning strategy can potentially help to improve both the pointing-calibration efficiency and science observations. Although intensive computational effort needs to be invested in the training stages, the employed time for inferences is much shorter.

### 2.1. Physical Modeling of Scintillation Noise in Millimeter Astronomy

To obtain theoretical predictions for the atmospheric noise typically recorded by continuum cameras, in the Appendix, we develop the physics of electromagnetic waves propagating through the turbulent atmosphere. We are particularly interested in the random variations of the log-amplitude $\chi \equiv \log(E_1/E_0)$, where $E_1$ represents a small fluctuation around the electric mean-field value $E_0$, which in turn relates to the observed intensity as $(I - I_0)/I_0 = |E_1|^2/|E_0|^2 = e^{2\chi}$ (Tatarskii 1961, 1971). We can see that $\chi$ is very useful in quantifying the intensity variations induced by the turbulent atmosphere, also known as scintillation.

In the wet atmosphere, millimeter waves' scintillation is caused by refraction and absorption turbulent fluctuations, which are caused by temperature, humidity, and pressure fluctuations. For infrared and millimeter wavelengths, refraction and absorption processes are respectively effectively captured by the real ($n_R$) and imaginary ($n_I$) parts of the refractive index. Consequently, the scintillation spectral density, $P_\chi(\omega)$, defined by

$$\overline{\langle \chi^2 \rangle} = \int d\omega\, P_\chi(\omega), \tag{2}$$

is given by $P_\chi(\omega) \approx P_R(\omega) + P_I(\omega)$, where $P_R$ is the contribution from $n_R$ and $P_I$ corresponds to $n_I$. Here $\omega$ is the Fourier temporal frequency. The cross-correlated spectral density $P_{\mathrm{IR}}$ appearing in Equation (A23) is typically much smaller than the previous two, and so, we ignore it. Considering a telescope with a circular-aperture radius $a_r$, observing at zenith angle $\vartheta$, the explicit expressions are

$$\begin{aligned} P_R(\omega) = {} & 2(2\pi k)^2 \sec\vartheta \int_{h_{\min}}^{h_{\max}} dz\, C_R^2(z) \\ & \times \int_{\omega/\upsilon(z)}^{\infty} d\kappa\, \frac{\kappa\, \Phi_0(\kappa)}{\sqrt{(\kappa\upsilon)^2 - \omega^2}} \\ & \times \sin^2\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right)\left(\frac{2J_1(\kappa a_r)}{\kappa a_r}\right)^2, \end{aligned} \tag{3}$$

$$\begin{aligned} P_I(\omega) = {} & 2(2\pi k)^2 \sec\vartheta \int_{h_{\min}}^{h_{\max}} dz\, C_I^2(z) \\ & \times \int_{\omega/\upsilon(z)}^{\infty} d\kappa\, \frac{\kappa\, \Phi_0(\kappa)}{\sqrt{(\kappa\upsilon)^2 - \omega^2}} \\ & \times \cos^2\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right)\left(\frac{2J_1(\kappa a_r)}{\kappa a_r}\right)^2, \end{aligned} \tag{4}$$

where $k = 2\pi/\lambda$ is the electromagnetic wavenumber, and $\kappa = 2\pi/l$ is the wavenumber associated with the atmosphere fluctuations. The functions $C_R$ and $C_I$ are the structure parameters related, respectively, to $n_R$ and $n_I$; $\Phi_0(\kappa)$ is the turbulence spectrum as in Equation (A12); $\upsilon$ is the transverse wind speed; and $J_1$ is the first-order spherical Bessel function. The integration variable is the altitude, and it goes from the observatory altitude above sea level $h_{\min}$ to the troposphere's end $h_{\max} \approx 10$ km.

The qualitative behavior of each spectrum is as follows. The scintillation spectrum due to the real part of the refractive index behaves as a $\omega^{-8/3}$ power-law for high frequencies, but the power depicts a plateau over lower frequencies (see Figure 12). Experimentally, $C_I^2$ has been found to be from $10^{-3}$ to $10^{-7}$ times smaller than $C_R^2$ (see the Appendix). One might have naively implied that $P_I$ ought to be negligible compared to $P_R$, but that is only true for higher frequencies. Indeed, the imaginary part dominates the lower-frequency power. It is essential to understand the physics involved in these two frequency domains, especially to plan technological strategies. For instance, a more sensitive camera with lower thermal noise or a larger telescope with higher aperture efficiency will significantly reduce the high-frequency scintillation noise. However, none of these two technological improvements

[1] 1 Jy = $10^{-26}$ W m$^{-2}$ Hz$^{-1}$.

would dramatically reduce the low-frequency scintillation noise caused by the millimeter waves' anomalous dispersion.

Early attempts to introduce a model for the atmospheric-noise temporal power spectrum (Church 1995; Lay 1997) included explicitly only the refractive-index real-part fluctuations (equivalent to $P_R$), missing the most dominant structure caused by water molecules, and encoded in the $P_I$ spectrum, corresponding to the imaginary part of the refractive-index fluctuations. A useful advantage of the IR scintillation spectrum in Equations (3) and (4) is that they are profile-generic expressions, i.e., different turbulence models $\Phi_0(\kappa)$, structure parameters $C_R^2(z)$, $C_I^2(z)$, or wind-velocity profiles $\upsilon(z)$ can be tested. Consequently, the IR scintillation spectrum can be ideally adapted by physically motivated parametric approaches (Errard et al. 2015) to study the atmospheric effects on astronomical observations.

### 2.2. Simulation of Atmospheric Temporal Series

If we know the noise spectral density beforehand, we can generate random realizations of structured atmospheric noise. Our model in Equations (3) and (4) allows us to generate simulations based on fundamental physics with the appeal of clearly interpreted parameters that are under control. However, in real observations, there can be many systematic effects introducing several structured-noise levels that are not always under control. Thus, we need a methodology to regard more representative observational conditions, such as the site's weather and instrumentation.

White Gaussian noise is the simplest example of unstructured noise, characterized by a constant (or flat) power spectrum, which implies stationary uncorrelated temporal series (Boyat & Joshi 2015). One consequence is that the white-noise standard deviation (std) remains approximately equal regardless of the temporal location and sample size. Atmospheric noise, on the other hand, is structured with a characteristic Brownian or $1/f$-like power spectrum (Milotti 2002). Structured noise is not necessarily stationary; it implies that the global std in a stream may significantly differ from local values within data patches (see e.g., Kirchgässner et al. 2012). The simplest way to classify the complexity of $1/f$-like noise is through the negative slope in log–log scale—the steeper the spectrum, the more complex the structure. Astronomical observations in millimeter waves contain both structured noise from the atmosphere and white noise from instrumentation.

In the presence of structured random noise, it is a common practice to use a whitening filter to enhance the signal detection (see, e.g., Cuoco et al. 2004; Allen et al. 2012; Abbott et al. 2016). Conversely, we implement an inverse-whitening transform to generate structured-noise time series, similar to the atmosphere emission at millimeter wavelengths. For the following analysis, we simulate sky observations employing a methodology described by Rodríguez-Montoya et al. (2018) using a spectral density, either from theoretical considerations or from observational calibrated data.

Let us assume that $\mathcal{P}$ denotes a target spectral density (derived theoretically or inferred from observations). Likewise, $\mathcal{N}_i$ denotes a white-noise flat power spectrum. Then, the $i$th simulated noisy realization $\mathcal{S}_i$ can be generated by

$$\mathcal{S}_i = \mathcal{F}^{-1}\{(\mathcal{P} + \mathcal{N}_i)^{1/2}\, \mathcal{G}_i\}, \tag{5}$$

where $\mathcal{F}^{-1}$ denotes the inverse-Fourier transform operator and $\mathcal{G}_i$ is a Fourier-transformed random-Gaussian sequence with the same length as the time stream. Effectively, $\mathcal{G}_i$ allows us to model small power shifts from the average power spectrum. Notice that $\mathcal{S}_i$ follows a Gaussian distribution, with correlation scales that depend on the random combination of $\mathcal{P}$, $\mathcal{N}_i$, and $\mathcal{G}_i$.[2]

Thus, the power-spectrum shape (either flat or $1/f$-like) determines the degree of structure in our atmospheric-noise simulations, and we can generate three levels of structured noise: (i) the simplest white Gaussian noise is generated with $\mathcal{P} = 0$, (ii) the IR scintillation noise with $\mathcal{P}$ given by Equations (3) and (4) with added white instrumental noise, and (iii) mock observational noise generated from an empirical spectral density. Rather than just picking a single one, we exploit these three incremental structure levels to train our neural network and improve overall learning performance. Figure 1 shows examples of the three types of simulated structured-noise inputs that we use throughout the neural network training.

## 3. Deep-learning and Recurrent Networks

### 3.1. Neural Networks and Recurrent Connections

Recurrent neural networks (RNNs) are a particularly attractive deep-learning architecture for applications involving ordered data. We can understand an RNN cell as a function that exploits recursiveness to find the current state of a dynamical variable. Keeping correlations among spaced events is indeed a powerful learning tool to handle long-standing structures. An RNN cell morphology is similar to the most basic cell but with a self-connected loop—namely, a hidden state—that feeds its own output back a number of times, turning the cell into a dynamical system that correlates the information between individual inputs (Cho et al. 2014). The simplest example is the Jordan RNN cell (Jordan 1997), which is composed of two units; the first one is a hidden state representing the memory of the cell,

$$h_t = f_h(U_h\, x_t + W_h\, y_{t-1} + b_h), \tag{6}$$

where $x_t$ is the current input state and $U_h$ is an input-to-hidden weight matrix. By the same token, $y_{t-1}$ is the previous output state, $W_h$ is an output-to-hidden weight matrix, and $b_h$ is a bias parameter. The function $f_h$ is a nonlinear transformation that produces the current hidden state $h_t$. The second component of the Jordan RNN cell is in itself the most basic unit used in deep learning (also called the dense cell), and it produces the output state in this case,

$$y_t = f_y(\, V_y\, h_t + b_y), \tag{7}$$

in which a hidden-to-output weight matrix $V_y$ multiplies the hidden state $h_t$, with the addition of a bias parameter $b_y$ and nonlinearly transformed by the function $f_y$. The output state $y_t$ can either be fed back to the cell, or delivered to other cells as their input. The nonlinear transformations $f_h$ and $f_y$ are called activation functions and help to assigns relevance to the cell's information output. The RNN cell abstracts features of the temporal sequence in its weight matrices, where $U_h$ is more

[2] Non-Gaussianities are expected in actual atmospheric data, and thus, the spectral density may not contain all relevant information. Nonetheless, our scintillation model could be combined with iterative techniques like the Johnson translator system (see, e.g., Wang et al. 2018, 2021) to simulate atmospheric series with a given skewness and kurtosis.

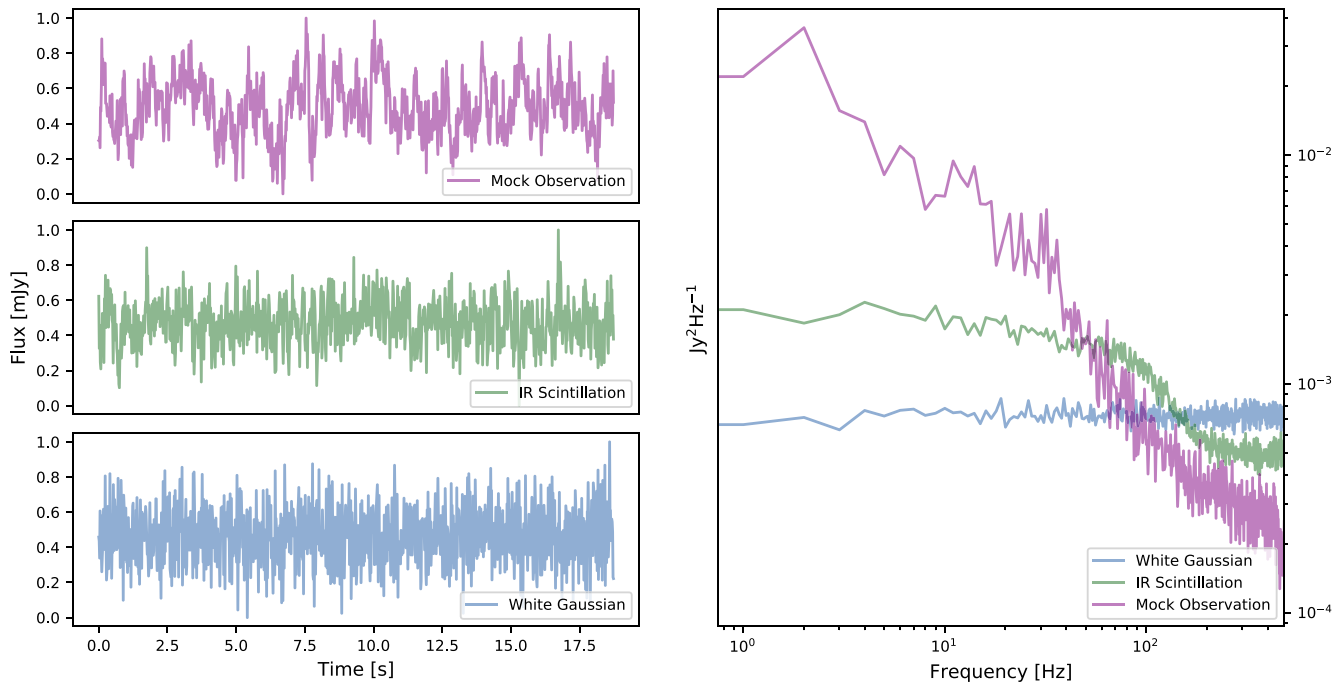


**Figure 1.** The left panel shows noise-realization examples of white Gaussian, IR scintillation, and mock observations in the time domain. The right panel shows a power spectra average of 200 noise realizations in the frequency domain. Mock observations show higher energy at low frequencies, with an average negative slope near 1 in log–log space. White noise resembles a flat spectrum with log–log zero slope. IR scintillation noise represents an intermediate complexity used in our training strategy, with a log–log slope of approximately $-1/2$. See Sections 4.1 and 4.3 for more details.

sensitive to new external data, and $W_t$ keeps a memory of the previous state. Because the only purpose of the hidden state $h_t$ is to abstract the sequence's temporal behavior, the hidden state $h_t$ may have arbitrary dimensions.

As a drawback of recurrent networks, the amount of information that can be abstracted into a simple RNN unit is typically short range. Also known as the vanishing gradient effect (Hochreiter et al. 2001), there is always the risk of inducing learning degradation over extended time intervals. Closely related is the problem of exploding gradients (Pascanu et al. 2013), when cumulative losses increase excessively and lead to unstable or impossible learning. A better practice is to use long short-term memory (LSTM) (Hochreiter & Schmidhuber 1997) to avoid the vanishing and exploding gradient effects.

The LSTM cell is a discriminating mechanism to either preserve or discard long-term information. The LSTM cell's ability to forget unnecessary information enables itself to keep temporal correlations for more extended periods. Intuitively, an LSTM unit internal structure is made up of two hidden mechanisms that act as pipelines for the information of distinct temporal scales. The first mechanism performs the operations of a single recurrent cell, compiling a representation of a few past data inputs. This hidden state is interpreted as a short-term memory unit, due to its limited ability to abstract small-extent behavior. The second mechanism encodes the sequence's global behavior, acting as a filter for the most significant long-term features. This new memory state allows the network to keep important information that would otherwise be overwritten by new data entries in a short-term memory unit.

In the vanilla LSTM layer (Hochreiter & Schmidhuber 1997), depicted on the right-hand side of Figure 2, the information flows through a system of logic gates whose parameters are trained to learn what information should be preserved or missed. The forget $f_t$ and input $i_t$ gates can be written as

$$f_t = \mathrm{Sig}(U_f\, x_t + W_f\, h_{t-1} + b_f),$$
$$i_t = \mathrm{Sig}(U_i\, x_t + W_i\, h_{t-1} + b_i). \tag{8}$$

Notice that each logic gate has a simple recurrent cell structure, with the only distinction of their performed task for the LSTM. The forget gate $f_t$ uses the sigmoid activation function to evaluate the data input $x_t$ and the previous hidden state $h_{t-1}$ (along with their respective weight matrices) and returns numerical labels between 0 and 1. A label close to zero means that a specific memory entry is not relevant for the global understanding of the sequence, and hence, it will be likely ignored in subsequent iterations. Conversely, a label close to 1 means that the state is relevant and is likely to be preserved. The second task is to evaluate how relevant the new input data are, which is done by the input gate $i_t$, also using the sigmoid function and the respective weight matrices. Both $f_t$ and $i_t$ are then combined to produce a new long-term memory called the cell state,

$$c_t = f_t \circ c_{t-1} + i_t \circ \tanh(U_c\, x_t + W_c\, h_{t-1}). \tag{9}$$

The Hadamard (or element-wise) product $f_t \circ c_{t-1}$ suppresses the irrelevant features of the past cell state $c_{t-1}$ and prevents them from persisting into the new cell state $c_t$. The tanh gate is an additional recurrent operation used to complement the input gate $i_t$ to create a list of candidates integrated into the long-term memory $c_t$. Finally, the output gate has the role of actually selecting from the feature candidates contained in the new cell state $c_t$ those that will be transferred to the new hidden state $h_t$. This operation is just a filtered version of $c_t$, with a simple

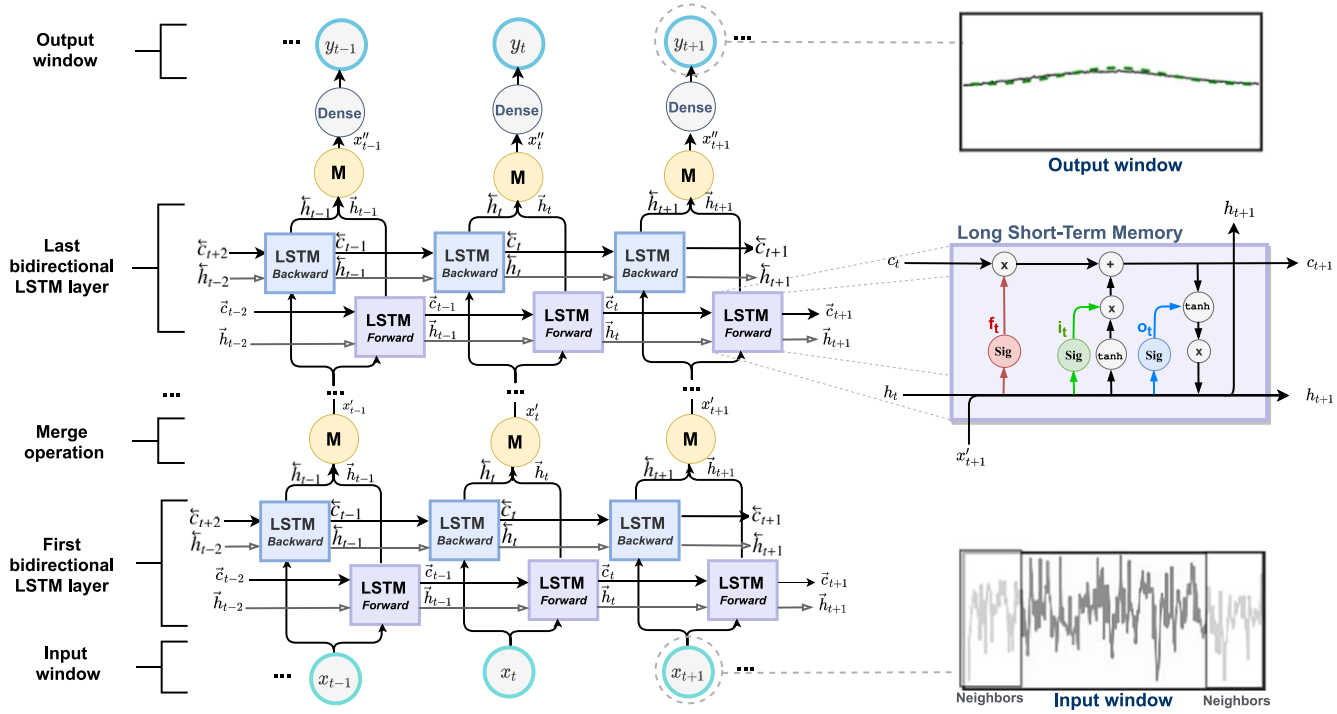


**Figure 2.** Visual representation of our deep recurrent model, mainly built up of LSTM units deployed in bidirectional layers. Left: At each training iteration $t$, a noisy input window $x_t$ is processed by the first LSTM forward and backward layers. The forward LSTM sublayer learns the temporal structure from left to right, while its twin backward sublayer learns in the opposite direction. An intermediate module M performs a merge of forward $\vec{h}_t$ and backward $\overleftarrow{h}_t$ hidden states. This operation yields a new state $x'$ with augmented information to feed the upper bidirectional LSTM layer. As a final processing step, a linear dense layer transforms the last layer output into an object $Y_t$ with the predefined output shape. On the right-hand side, the LSTM cell internal structure is exposed. The cell state $c_t$ or long-term memory and the hidden state $h_t$ or short-term memory both preserve an internal representation of distinctive temporal scales. The information shared between these two states is carried out by the forget $f_t$, input $i_t$, and output $o_t$ logic gates. See Section 3.1 for more details.

sigmoid gate $o_t$,

$$o_t = \mathrm{Sig}(U_o\, x_t + W_o\, h_{t-1} + b_o),$$
$$h_t = o_t \circ \tanh(c_t). \tag{10}$$

The output state $h_t$ can be either used as a hidden state fed back to the cell or delivered to the next LSTM layer as an input state.

The fact that each network logic gate can (be trained to) decide what information is more or less relevant (to describe the data of interest) is a remarkable advantage of the LSTM cell. In the family of recurrent architectures, the LSTM indeed represents the state of the art in modeling complex systems that require large scales of temporal understanding. For a comprehensive explanation of the LSTM cell internal structure and its variations, we refer the reader to Hochreiter & Schmidhuber (1997) and Greff et al. (2017).

### 3.2. LSTM Network for Time-domain Astronomical Cleaning

Even though the atmosphere evolution is chaotic, inside the universal inertial regime (see the Appendix), it is possible to find a timescale in which the fluctuations are coherent and well described by the spectral density in Equations (3). Consequently, the LSTM cell seems to be a natural choice for discerning between atmospheric and astrophysical behaviors. Below, we describe our recurrent architecture implementation to handle temporal astronomical data.

Given that the astronomical data-reduction process is performed offline (i.e., after the observation time), we can modify the LSTM structure to benefit not only from past events but also from future information. Thus, we choose a many-to-many input/output configuration: A sliding window replaces each time segment in the series with a fixed number of past and future neighbors. We feed every temporal window into the network with the aim to abstract the neighboring correlations.

Taking even more advantage of future temporal information, we implement a bidirectional approach (Schuster & Paliwal 1997; Graves & Schmidhuber 2005) that combines two identical recurrent units into a single layer: one processing the input series in a right-wise direction, and another processing the same time series left-wise.[3] Recalling the notation used in Equations (9) and (10), each bidirectional LSTM layer computes two independent cell states $\vec{c}_t$ and $\overleftarrow{c}_t$ that encode the long-term memory in the right-wise and left-wise directions, respectively. Likewise, each bidirectional LSTM layer produces two hidden states, $\vec{h}_t$ and $\overleftarrow{h}_t$. In our implementation (see the M modules in Figure 2), the merger of these two states creates the bidirectional LSTM layer output,

$$y_t = M(\vec{h}_t, \overleftarrow{h}_t). \tag{11}$$

The function $M$ is a concatenation operation that ties the states into a unique output $y_t$, which becomes in turn the input for the next bidirectional LSTM layer $y_t \rightarrow x'_t$.

According to Figure 2, the information flows through our architecture as follows. A time segment of a data stream $x_t$ is fed into the first bidirectional LSTM layer. The forward layer sees the sequence in the causal temporal sense, from left to

[3] During the preliminary stages of our model, we explored a recurrent encoder–decoder configuration (Cho et al. 2014). Yet, the implied high computational costs did not pair with a meaningful learning loss value improvement (for our purposes).

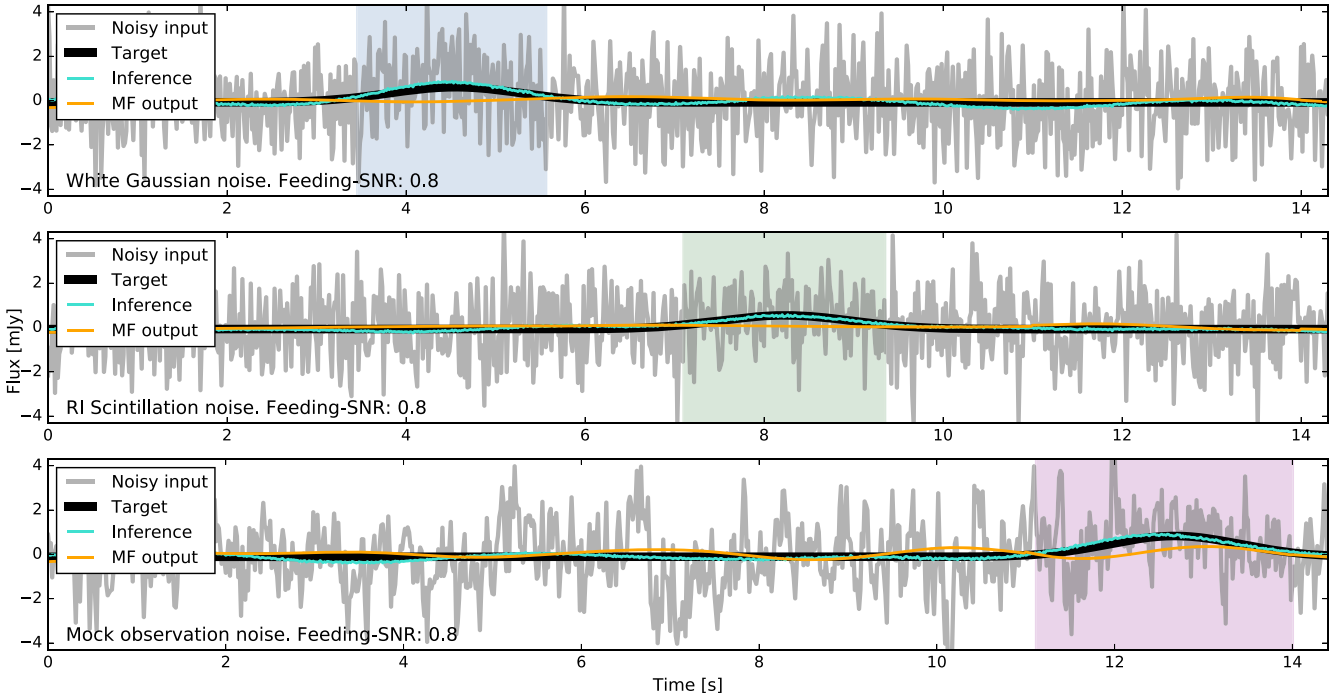


**Figure 3.** Examples of three levels of the structured noise and target signals (see Sections 2.2 and 4.1 for details on data preparation) filtered by our recurrent neural network. The gray lines are noisy inputs, the three of them with an initial signal-to-noise ratio of 0.8; black lines represent a simulated astronomical signal dubbed as the target; the light turquoise lines are the network inferences of the underlying astronomical signal, which are detected after the cleaning process with a signal-to-noise ratio above 3. For comparison, a matched-filtered (MF, Section 4.4) sequence is shown (orange line), failing to detect any source at this very low feeding S/N. (See Figure 6 for a zoom-in of the source detection windows.)

right. The backward layer processes the window simultaneously in the reverse direction. The cell states $\vec{c}_{t-1}$, $\overleftarrow{c_{t+1}}$, and the hidden states $\vec{h}_{t-1}$, $\overleftarrow{h_{t+1}}$ are initialized with the outputs from the preceding and following temporal windows. The forward and backward LSTM cells produce two hidden states $\vec{h}_t$ and $\overleftarrow{h_t}$ that are joined by a merge module M and then delivered as the input $x'_t$ to the next bidirectional layer. In this way, the information obtained from both temporal analyzes is packed into a new single matrix representation. The process continues up to the deepest bidirectional LSTM layer contained in the network. At the last processing stage, there is a simple dense layer (without any recurrent connection; see Equation (7)), which acts just as a parametric point-wise operation to map the LSTM output into the user-defined temporal size (we choose this hyperparameter value in Section 4.2).

## 4. Numerical Experiments

In the previous sections, we explained our expected signal's physical and statistical properties, connected to the deep-learning theoretical framework for our architecture's design. In this section, we describe the data preparation to train the network, the choices for hyperparameter values, and the training strategy that we adopt to produce the results discussed in the next section.

### *4.1. Data Preparation*

The effectiveness of deep-learning algorithms relies on the quality of training: the more massive and accurate the data sets, the higher are the expectations for improvement. These two conditions are obviously uncommon in many real-world applications and particularly for unexplored astronomical sources, for which the underlying brightness is a priori unknown and always mixed with foregrounds and instrumental noise. Notwithstanding these, it is possible to train the network with synthetic data created on the basis of physical models, intending to transfer the gained learning to more realistic applications. Provided that the mock and real data domains are closely related, it has been extensively documented that the knowledge can be transferred from the former to the latter domain (Pan & Yang 2009).

We generate temporal series with pointlike source random insertions, choosing their statistical properties according to instrumental and observation specifications, as described in Section 2. The network processes a single-bolometer recording as an independent input. As Figures 1 and 3 exemplify, we consider three types of structured noise with increasing complexity: (i) White Gaussian noise is the less structured noise under consideration. Despite its unrealistic simplicity, this noise distribution works as a handy starting point for our training strategy. (ii) IR scintillation noise samples represent a complexity midway point. We choose an intermediate value of $10^{-5}$ for the ratio $P_I/P_R$ in our simulations (also see the Appendix). (iii) Mock observations synthesized from real data are our closest approximations to the observed atmospheric foreground and expected instrumental features.

We perform a few preprocessing steps for the real AzTEC temporal series to make them numerically more treatable. We utilize the 2018 season real AzTEC observations with a 0.17 opacity, representing remarkably bad weather nights. We apply PCA to a 110 time-stream set and then subtract the first two principal components. This process removes typically the longest correlations linked mainly to bad weather conditions.[4]

[4] PCA can partially remove astronomical emission, too. Thus, a more accurate preprocessing methodology or a flux calibration (as in Downes et al. 2012) could be employed in the future with real-time observations.

Meanwhile, the remaining noise is still considerably more structured than the IR scintillation and white Gaussian simulations (see Figure 1). Finally, we pick a detector located at the edge of the array; because it is off source most of the time, the atmospheric flux dominates this detector's time stream, and we can conveniently use it for further simulation steps. Training the neural network with realistic simulations allows us to build any features from other systematic effects (e.g.,, the faint confusion background) into the LSTM parameters, helping the network to recognize the embedded targets.

Next, for the three types of noise, we fit and subtract a third-degree polynomial that allows us to work with centered data and filters out the larger scales. We proceed to insert pointlike source signals into the simulated noisy time streams. Due to long-range nonstationarity in some structured-noise samples,[5] we should avoid interpreting the global std as the relevant level of noise. The time-stream size is chosen from technical limitations such as computational memory or the telescope's time so there is nothing special about the global std. Instead, we choose the full pointlike source extent (4.4 s at $6\sigma$) to compute the local std as the reference noise. We adopt the following procedure: We randomly select a time segment that matches the pointlike source; then, we measure the noise std inside 4.4 s and adjust the pointlike source amplitude to meet the desired signal-to-noise ratio (S/N). We name this induced complexity feeding S/N, calculated as the maximum signal's amplitude ratio to the local noise level. This procedure allows us to control the complexity presented to the deep neural network at every training stage.

By construction, for any of the three types of noise, the source brightness relative to the local noise is the only relevant independent variable to classify the sources according to complexity. Thus, as a final preprocessing step, we scale the temporal series to have the same standard deviation. It means that our simulations could be interpreted either as bright calibration sources immersed in very noisy data or as dim sources in moderate noise.

We stress that our time-stream simulator feeds the network with unique random realizations of structured noise. This aspect is especially relevant because deep neural networks are vulnerable to overfitting when training is performed over long periods using a finite data set. In our case, however, we can generate virtually as many random realizations as the training procedure may require, while each noise pattern passes through the network only once.

### 4.2. Hyperparameters

Before designing the neural network architecture, we first need to choose the batch dimensions, i.e., the number of time steps feeding the LSTM. Then, the network uses each target stream and its corresponding inference to compute the relative loss function that serves as a reference to update the network weights. We can significantly reduce the memory resources by splitting the batch into a group of windows, which cannot be arbitrarily small. Although pointlike source signals are a few seconds long, the LSTM needs to abstract longer timescales to assemble the long-term memory structure patterns. Thus, the window size must be at least broader than the pointlike source length. Moreover, individual windows should conveniently share some intermediate time steps to help the network perform continuous inferences through cutting edges.

Balancing the underlying application requirements and computational costs, we choose a batch size of 1200 time steps (equivalent to 18.7 s), containing two 700 time-step-long windows, sharing 100 neighbors on both sides. It means that for each window, 700 time steps are fed into the network to infer the mid 500 time steps. At this point, the batches are ready to feed the network.

A deep neural network design requires selecting a set of hyperparameters representing the network capabilities and computational constraints: More extensive networks imply more massive processing time and memory. To avoid the exponential complexity involved in a blind search (i.e., exploring every combination inside the hyperparameter space grid), some systematic approaches like the random search, Bayesian models, and genetic algorithms, among others, can be used to produce efficient results (Luo 2016). Nonetheless, naive or expert-based explorations have not been entirely suboptimal in a wide variety of applications (Bergstra et al. 2011).

Our approach consists of a semisystematic exploration to find suitable hyperparameter values. First, we set up an initial framework with a simple configuration: We employ the well-known Adam optimization algorithm (Kingma & Ba 2014), looking to minimize the mean square error (MSE) as the loss function. Then, we perform a two-step hyperparameter search corresponding to the size of the architecture. We explore the network's depth (number of layers) and width (number of cells per layer) parameters, monitoring the loss function to find the best performance. The depth dimension is explored from one to seven LSTM layers with increments of one. We sweep the width dimension in the range of 40 to 250 cells in each layer with increments of 30 cells. The depth–width search space is delimited into a $7 \times 9$ grid using a fixed learning rate equal to $10^{-3}$. Finally, we perform a fine-tuning of the learning framework by iterating over Adam's learning rate range between $10^{-3}$ and $10^{-9}$. As a result, we instantiate the model proposed in Section 3.2 with 190 bidirectional LSTM cells, four deep layers and a $10^{-5}$ value for Adam's learning rate.

### 4.3. Training

The structure degree reached by the atmospheric noise represents quite a tough starting point to be handled by any parametric architecture, especially for weak-brightness sources. Nevertheless, a training strategy inspired by curriculum (Zaremba & Sutskever 2014) and transfer learning (Weiss et al. 2016) is—as we see below—able to perform a gradual complexity exploration over several levels of structure and feeding S/N.

Transfer and curriculum learning are techniques to share knowledge between machine-learning models. On the one hand, curriculum learning aims for intuitive exploration, beginning with simple examples and gradually escalating complexity. On the other hand, transfer learning takes advantage of models trained in a data domain, using their resulting parameter values (or part of them) as initialization states to find better results in a related but new data domain.

To perform stable training with the highest attainable precision, we propose an incremental learning approach that combines transfer and curriculum learning to capture two complexity axes: the initial S/N and the type of noise structure.

[5] In a series coming from a non-flat power spectrum, we may have for instance that $\mathrm{std}(15s) \neq \mathrm{std}(2s)$.

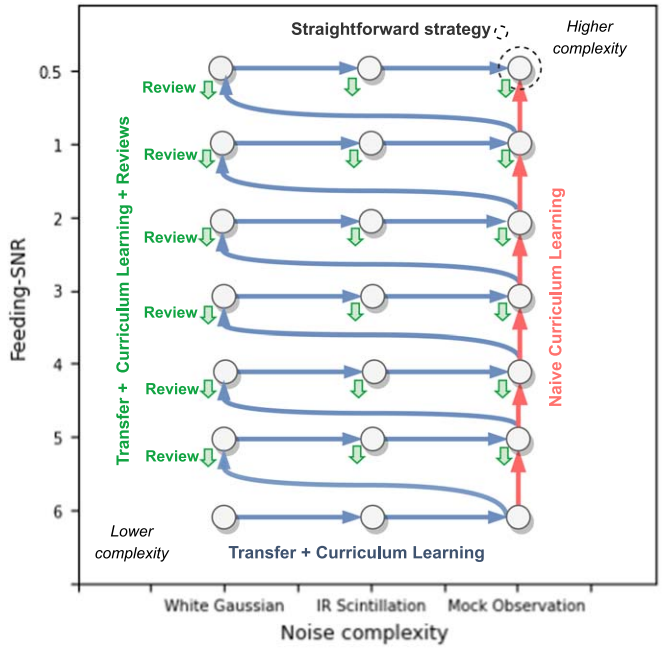


**Figure 4.** Pictogram of the paths taken by each training strategy tested. It should be read from the lower-left to the upper-right corner. The exploration space is built onto two axes: the feeding S/N on the vertical axis and the noise complexity level on the horizontal axis. The upper-right dotted circle represents the straightforward strategy. The red-solid line going upwards at the right corresponds to the naive curriculum learning strategy. The blue-solid zigzag line represents transfer + curriculum learning. Finally, the green downward-pointing arrows denote the transfer + curriculum + reviews strategy. (See Section 4.3 for details.)

Dubbing curriculum learning as the strategy's vertical axis, our application gradually decreases the feeding S/N from 3.5 to 0.1. This procedure helps to accelerate the optimization convergence by simulating the gradual inclusion of fainter astrophysical sources. Correspondingly, we dub transfer learning as the strategy's horizontal axis: At each feeding-S/N level, the network explores three structure degrees: white Gaussian, IR scintillation, and mock observations. This procedure pretends to mimic the conditions of incrementally worse weather.

We monitor the learning evolution along the training stages regarding the loss function (the target-inference MSE) at each training iteration, which is good practice for early identification of missed behaviors such as overfitting or deficient learning performance. Monitoring the loss curve at intermediate stages also allows us to register variations from different training strategies. For example, the simplest approach would be the straightforward strategy, in which the network is trained directly with the highest complexity, in this case, the noisier mock observations.

However, we probe three more learning strategies (explained below), each one exploring different routes of incremental complexity into a discretized two-dimensional space. Each training strategy path's design is illustrated in Figure 4. After a full network training, each strategy is tested on the same data set of 1250 series with mock-observation noise and a randomly varying feeding S/N between 0.1 and 3.5. In this test, the same data set is used to make a fair comparison of the learning strategies. In Figure 5, we record the loss-function value at each iteration for every tested training strategy. The thick lines represent the loss-function moving average inside a 25 iteration local window, and the colored bands represent the corresponding moving std.

For the straightforward strategy, we can see in Figure 5 that the loss function displays a sudden dropout at the beginning of training ($\lesssim$600 iterations), due to the initial network adaptation to the data domain. A few steps later ($\gtrsim$600 iterations), a plateau can be noticed, albeit with disadvantageous dispersion amounts. We call the second strategy naive curriculum learning (NCL), which consists of preliminary training over mock observations, with constant feeding-S/N decrements from 3.5 to 0.1 in steps of 0.1. At any given feeding-S/N bin, when no significant improvement in the last 400 iterations is noticed, the network proceeds to the next feeding-S/N bin. We notice that NCL considerably improves the loss mean value compared to the straightforward strategy, albeit the dispersion levels are still comparable. Besides, the NCL training checkpoint performs poorly with white Gaussian or IR scintillation noise inferences. This lack of robustness against structure degree variations suggests the main disadvantage of NCL.

Thus, we test a third strategy called transfer-curriculum learning (TCL), which alternates transfer learning iterations (using sequentially white Gaussian, IR scintillation, and mock observations) amid intermediate stages of curriculum learning. TCL reports a slightly lower average loss and less dispersion than NCL, but with the added benefit of performing better with the other two types of noise.

Notice that the NCL and TCL loss curves suffer considerable variability, some of them as large as those reported by the straightforward strategy. This dispersion is rooted in the network's adaptation to the latest stages while forgetting previous feeding-S/N levels, an effect also known as catastrophic forgetting (McCloskey & Cohen 1989; Ratcliff 1990). It means that a chronically poor understanding of past training samples might pile up as the relative complexity increases. This effect would appear after long periods of training and an overspecialization of the latest feeding-S/N bins.

Inspired by a human learning strategy known as spaced repetition,[6] we address the catastrophic forgetting problem by adding intermediate training stages to review the past complexities. We call this strategy transfer-curriculum learning with spaced reviews (TCL-SR). After each transfer and curriculum learning round (between two feeding-S/N bins), we schedule a review of the previous feeding-S/N range. By randomly choosing the feeding S/N, the network can "remember" high-S/N data distributions while continuing to learn lower-S/N complexities. During every reviewing stage, the iteration number is controlled on the fly by the loss-function convergence; after 400 iterations without significant MSE improvement, the training scheduler triggers the next TCL-SR stage to pursue the following complexity level. In comparison, the TCL-SR strategy yields better results in terms of narrower variability and lower mean values for the loss function. Although the invested training time in the TCL-SR strategy is more significant than the straightforward strategy, the benefits are appreciable, especially at late stages, when the low-S/N and high-structured-noise levels pose a more challenging task for network inferences.

[6] In cognitive psychology and pedagogy, spaced repetition is a learning technique based on the periodic review of concepts to reduce the probability of forgetting information after long studying periods.

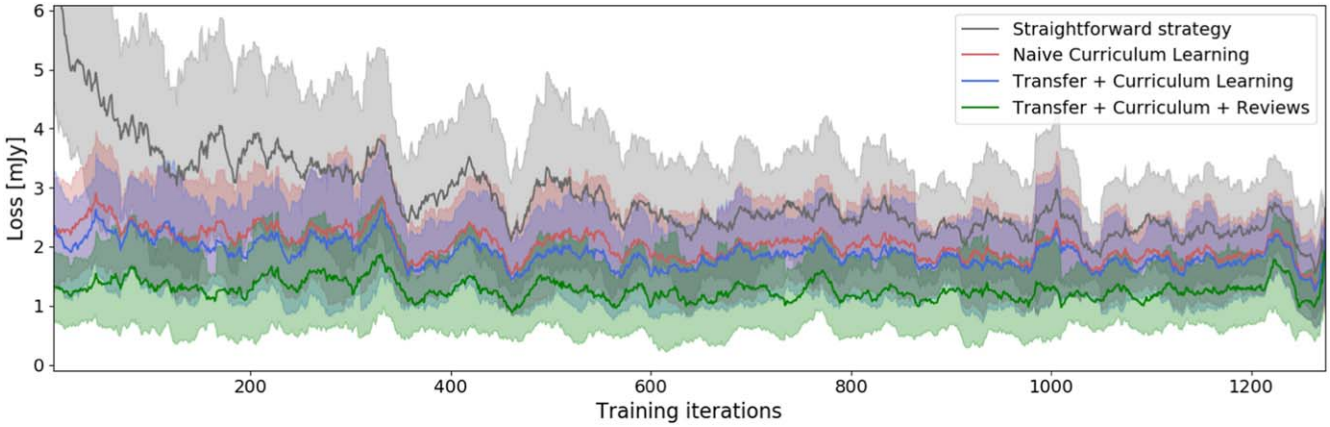


**Figure 5.** Learning curves along the first 1250 training iterations with mock-observation noise and random feeding S/Ns between 0.1 and 3.1. Previously, each model was thoroughly trained according to the learning strategies drawn in Figure 4. Each curve is measured as the average MSE evolution, and the colored regions represent the corresponding dispersion. Four training strategies are tested, of which the straightforward strategy is the least effective, while our proposed strategy of curriculum transfer learning with spaced reviews exhibits the best performance. (See Section 4.3 for further discussion.)

Thus, we implement the TCL-SR strategy to train our LSTM network; the training procedure continues until an improvement in the loss function is no longer noticeable. Then, the optimization module may be disconnected, and the free parameters get frozen.

### 4.4. Matched Filtering

We implement matched filtering (MF) to compare our trained network results to conventional signal-processing techniques. A similar approach has been used as a baseline to compare the performance of neural networks, particularly in gravitational-wave experiments (Cuoco et al. 2004; Allen et al. 2012; Abbott et al. 2016). The MF is defined in the frequency domain by Turin (1960),

$$\mathcal{M}(\omega) = \frac{\mathcal{H}^{\star}(\omega)}{\mathcal{P}_N(\omega)}, \tag{12}$$

where $\mathcal{H}$ is a template for the signal of interest (the star superscript denotes complex conjugation), and $\mathcal{P}_N(\omega)$ is the noise spectral density. The MF is applied to the raw signal $D(\omega)$ in Fourier space, and the filtered signal is computed from the inverse-Fourier transform,

$$\hat{s}(t) = \Re(\mathcal{F}^{-1}\{D(\omega)\star\mathcal{M}(\omega)\}). \tag{13}$$

For our purposes, the template $\mathcal{H}$ is prepared to match the pointlike-source shape used before, and $D$ denotes in this context the astrophysical signal embedded in the noisy data sets described in previous sections.

Although the MF can recover compact astronomical sources above a noise threshold, it is well known in general that the MF has a few limitations (Allen et al. 2012): It is not robust against non-white-noise artifacts, and as the signal is not precisely known, the filter output S/N is strongly dependent on the template. Thus, MF is expected to underperform our trained network, especially with highly structured noise, and yield more false positives from low-S/N raw data.

## 5. Results

We are now interested in performing a statistical characterization of the network's inferences through incremental complexity levels. Including white Gaussian and IR scintillation noise in this characterization is instructive to contextualize the results obtained with mock observations. Figure 3 shows three examples of network inferences with three types of noise; altogether, they exemplify the network's inference of a faint source (feeding S/N = 0.8) embedded in incremental levels of structured noise. Another set of time streams with higher feeding S/Ns is shown in the upper panel of Figure 6, focusing on the detection windows.

First, we need to define a criterion for window detection (in a single shot). We see that for any given realization of atmospheric noise $A_i$ and an astrophysical pointlike source $s_i$, the network inference $\hat{s}_i$ is a noisy representation of the true signal $s_i$. Notice that low-frequency modes could dominate the residual atmospheric noise. To estimate this residual noise, we feed the network with $A_i$ (which is equivalent to zero signal) and obtain an inference dubbed as $\hat{A}_i$, which is an approximation of the residual noise pattern contained in $\hat{s}_i$. Henceforth, we denote the level of residual noise as std($\hat{A}_i$) and look for peaks in $\hat{s}_i$ above 3 × std($\hat{A}_i$), labeling them as detection candidates. Because we know beforehand our simulated signal, we can classify each detection as a true positive or false positive. We also verify that the inference matches the target temporal location, allowing slight temporal displacements within a tolerance of ∼0.17 s, which is near one-fourth of the source width.

Using this criterion, we can count the number of sources successfully detected after a single inference over each structured-noise type. We generate a large number of new white Gaussian, IR scintillation, and mock-observation noisy data sets, each made of 150,000 realizations, created with feeding S/N = 0.1, 0.2, ..., 3.5. In Figure 7, we see the recovery rate as a function of feeding S/N; it is higher for the less structured noise, reaching 95% slightly above a feeding S/N of 0.6 for white Gaussian, 0.8 for IR scintillation, and 1 for mock observations. Further, we find false-positive counts below 10% at feeding S/N $\lesssim$ 0.3.

To assert a comparison to a classical method, we test our MF implementation (Section 4.4) over another set of 150,000 random realizations per type of noise and compute the source detection rate in Figure 7. As it might have been anticipated, we see that the MF technique cannot recover as many astronomical sources as our trained network, especially at low S/N. For example, with a feeding S/N equal to 1, the MF

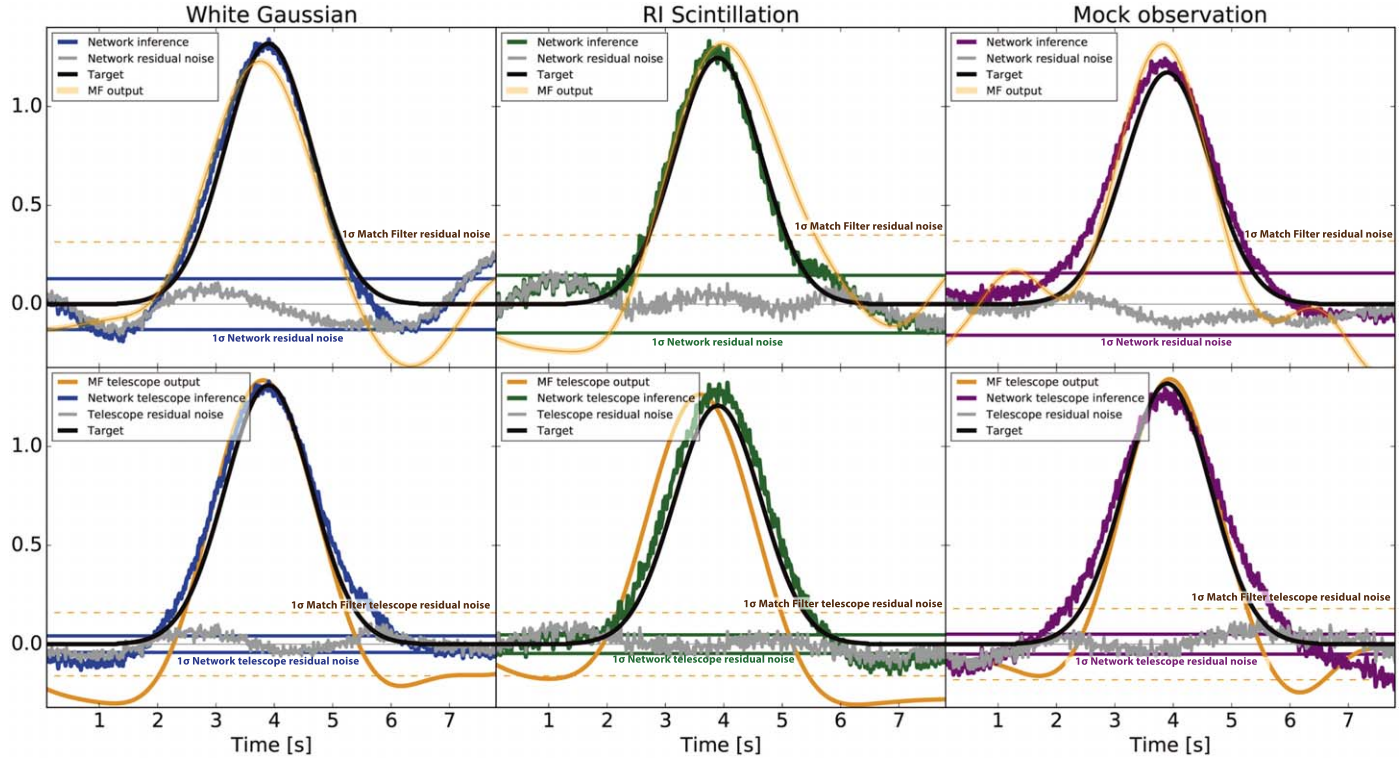


**Figure 6.** Upper panels: single-shot inferences of a ∼1.3 mJy source made by our network and MF. While in Figure 3, gray lines denote raw atmospheric noise, here, gray lines represent atmospheric-noise residuals (denoted by $\hat{A}_i$ in the text) after our network-based filtering. (MF noise residuals are not shown.) The horizontal colored solid lines represent the $1\sigma$ network-residual noise level in the complete batch, while the orange dashed lines are the size of the MF-residual noise. The network inferences (denoted as $\hat{s}_i$) are the colored noisy lines near the target ($s_i$), and the MF inferences are the orange solid lines. Lower panels: telescope measurements of the same sources after averaging 50 inferences of the same signal embedded in different noise realizations. The horizontal colored lines represent the corresponding $1\sigma$ residual noise levels in the complete batch. The telescope noise reduction from single-network inferences accounts for roughly 66% in these examples. This S/N is regarded as an approximation of the one that could be attained by a full-array integration in the map domain (see also Figure 11).

detection rates are ≃35% for white Gaussian noise, 34% for IR scintillation, and 4.4% for mock observations. These figures are in sharp contrast to our neural network results, generating a detection rate above 90% for the same feeding S/N (see Figure 7). Further, with MF, we find ∼20% false positives at feeding $\mathrm{S/N} \lesssim 1$.

Even when an astrophysical source is recovered in a single-shot inference (as Figure 7 shows), it does not necessarily imply an accurate measurement. We know that atmospheric-noise residuals harm single-shot temporal samplings, causing distorted flux measurements. However, given the random nature of atmospheric noise, residuals are expected to introduce random departures from the actual flux, which, on average, should cancel out at every pixel location in the map domain.

Henceforth, we use the term telescope measurement as the average of a large number of stacked inferences in the time domain. This term is reminiscent of what happens in the map domain, where the final measurement of an astronomical source is mainly the average of many (suitably cleaned) temporal samples. Consequently, although we do not address the map-making problem in our current analysis, we can, nonetheless, assert the expected map-domain detection quality.

Figure 6 (lower panel) shows the telescope measurements of a ∼1.3 mJy pointlike source after 50 stacked inferences. If the noise residuals were unstructured, the noise reduction after stacking 50 network inferences would have been $\sim 1/\sqrt{50}$, or ∼85%. However, we record an improvement of roughly 66%, indicating that the residual noise left by our network still has some degree of structure. We observe that the peak position is

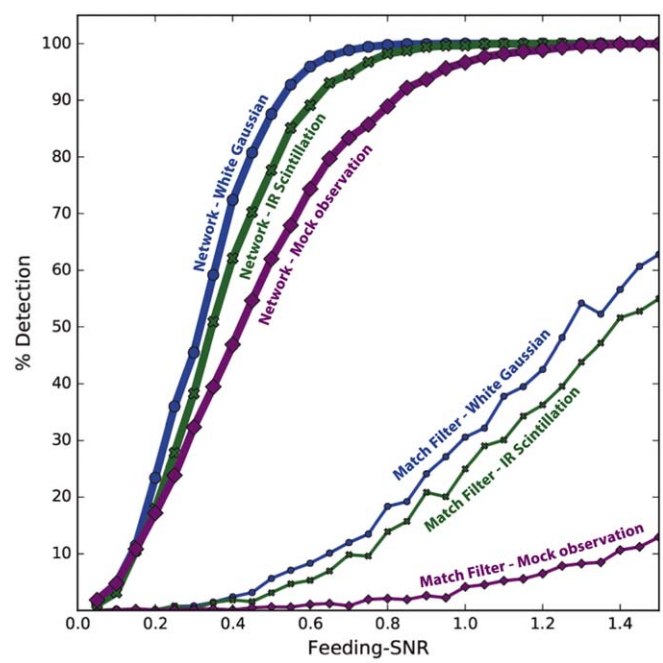


**Figure 7.** Source detection rate as a function of feeding S/N. The curves quantify the fraction of the total number of simulated sources that are successfully cleaned and detected above a $3\sigma$ noise level after a single-shot network or MF inference. Our neural network recovers a more significant fraction of sources than the MF. For example, the MF completeness for feeding $\mathrm{S/N} = 1$ is ≃35%, 34%, and 4.4% for white Gaussian, IR scintillation, and mock observations, respectively. In contrast, our neural network produces a source detection above 95% for the same feeding S/N.

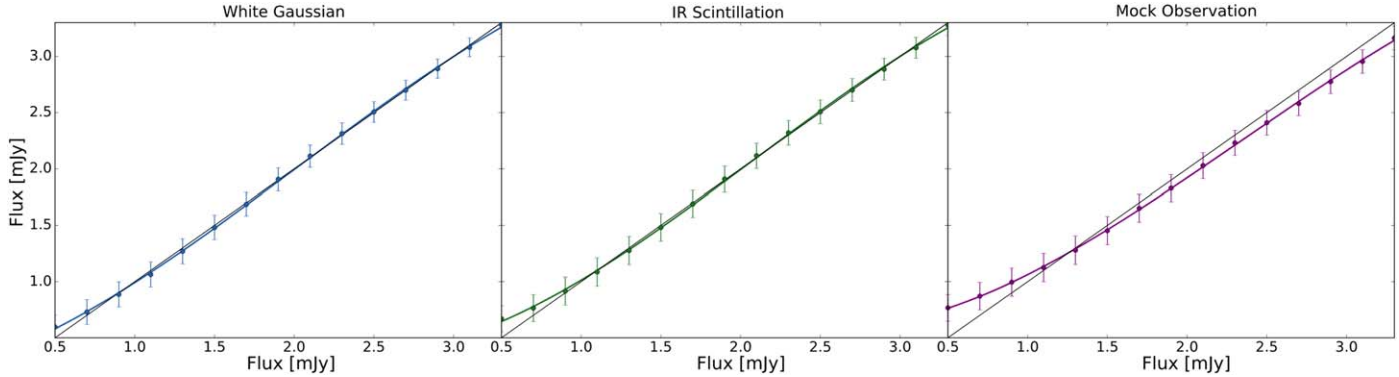


**Figure 8.** Flux calibration of network inferences. For each flux bin, 2000 network inferences are stacked, and an average signal is statistically inferred. The error bars represent the uncertainty in the amplitude of the pointlike source signal. The error bars joining curve represent a polynomial fit that can be used to calibrate posterior network inferences.

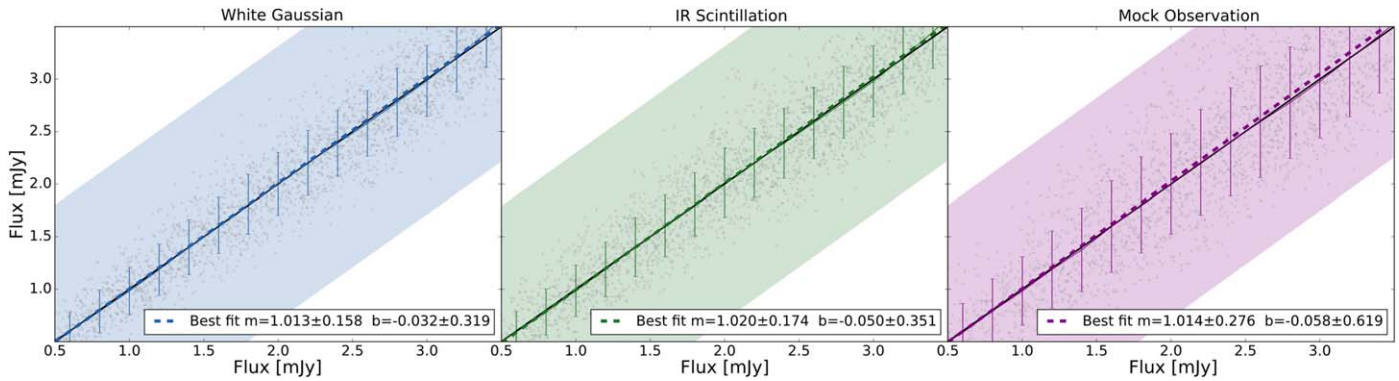


**Figure 9.** Comparison of network-inferred source fluxes vs. the real ones for the three types of simulated noise. Shaded regions represent $1\sigma$ dispersion of the feeding noise. Vertical error bars represent the inferred flux's $1\sigma$ dispersion (per flux bin of 0.2 mJy) due to atmospheric residuals. The cloud of dots is a subrepresentative sample of the 150,000 total inferences per type of noise.

accurately recovered, though the inferred pointlike source is broadened compared to the original target, clearly as an accumulated effect of small temporal shifts around the signal's peak. Atmospheric effects well known in astronomy as blurring or wandering are similar to the observed accumulated beam broadening (Tatarskii 1961, 1971; Wheelon 2001). For comparison, MF inferences in Figure 6 are significantly less accurate; even after being stacked to form telescope measurements, the average profile loses information at the skirts of the flux distribution. Moreover, notice that our network's $1\sigma$ residual noise is much smaller than the level of MF-residual noise in Figure 6.

Some sort of calibration is typically necessary for any cleaning methodology (see, e.g., Downes et al. 2012 and Rodríguez-Montoya et al. 2018). We can use telescope measurements not only to characterize the network's inferences but even more important to design a flux calibration procedure, which will be critical by the time when real observational data from any new-generation instrument will be available and processed by a scaled version of our current architecture. We split the flux range into 14 bins and generate 2000 noise realizations for each flux bin. We perform a Gaussian-curve fit to each averaged signal to measure the source flux and statistical error. From the 14 telescope measurements obtained, as shown in Figure 8, we see that they differ from the ideal one-to-one flux-correspondence line. Then, we perform a polynomial fit that we can use to calibrate subsequent network detection.

To verify the calibration effectiveness, we utilize, once again, our 150,000 noisy data realizations but correct the network flux measurements with the telescope measurements' best fit. The resulting cloud of calibrated flux points appears in Figure 9. Vertical error bars denote the $1\sigma$ dispersion amount. Shaded regions represent the typical $1\sigma$ feeding-error bar that was used to generate these simulations. For each type of noise, the dispersion-to-feeding-error ratio is $\lesssim 0.27$ for white Gaussian, $\lesssim 0.29$ for IR scintillation, and $\lesssim 0.44$ for mock observations. Performing a linear fit to the calibrated cloud of flux points, we find a good agreement with the ideal one-to-one correspondence line, indicating a good calibration. Subtracting the true flux and drawing the corresponding histograms on Figure 10, we can observe that the random distortion distributions tend to be symmetric, with a skewness-to-std ratio of $\lesssim 0.70$ for white Gaussian, $\lesssim 0.63$ for IR scintillation, and $\lesssim 0.70$ for mock observations. Thus, as we anticipated, we can assert that random flux distortions will indeed cancel out in the map domain, provided that they have been correctly calibrated.

From Figures 6, 8, and 9, we see that mock observations display comparatively lower recovery rates, larger flux bias, and more scattered dispersion with respect to those of white Gaussian and IR scintillation noise. We attribute this behavior to the structure degree contained in mock-observation noise and the underlying instrumental systematic features that are likely present in these more realistic simulations. Notwithstanding these, in preliminary setups, we also identified that the

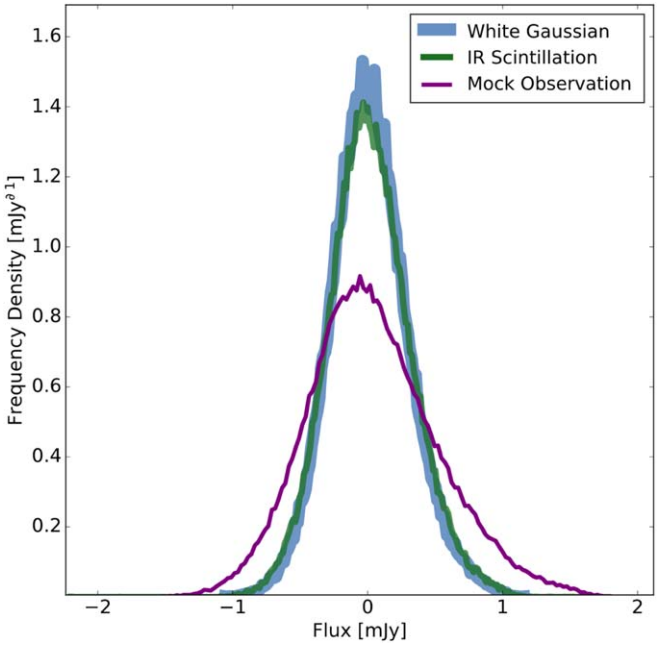


**Figure 10.** Distributions of random flux distortions on network inferences caused by atmospheric-noise residuals, using the three types of structured noise.

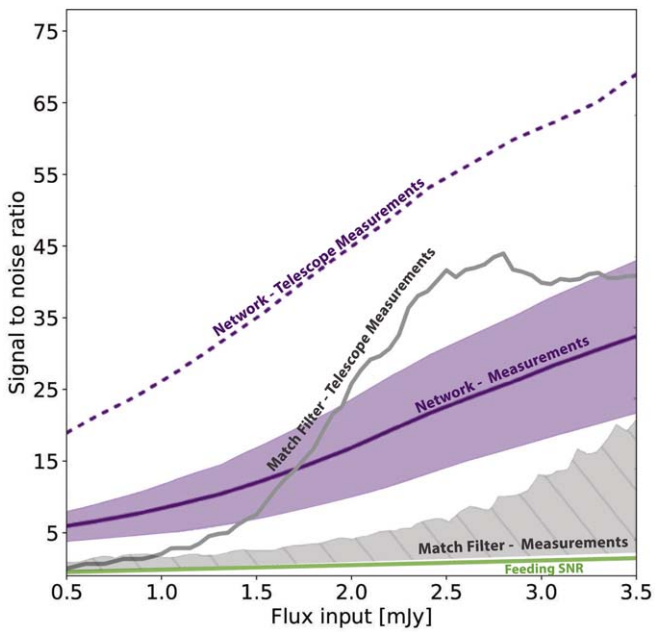


**Figure 11.** S/N enhancement from mock-observation noisy series. The ensemble average upon 150,000 independent realizations is shown to exhibit the network and MF detectability improvement. The curves represent the incrementally larger S/N per source flux in the case of raw data (feeding S/N), network and MF single-shot inferences, and telescope measurements.

training strategy could reduce both systematics. Indeed, to compensate for the higher complexity of low-feeding-S/N samples, our TCL-SR training strategy weighs more on the latest reviewing stages, having lower flux sources. This scheme increases the recovery of weak sources (Figure 7), reduces the random distortion dispersions (Figure 9), and explains the overestimations at low fluxes in Figure 8 (which we correct through flux calibration).

On the other hand, notice the significant noise-level drop in the telescope measurements in Figure 6. This improvement is more clearly represented in Figure 11, where we show the S/N obtained from mock observations (only) as a function of flux. The single measurements' S/N is computed as the expectation value of the 150,000 total inferences (previously used for Figure 7). The telescope measurements' S/N is computed directly from their corresponding error bars (see Figure 8). Both MF and our neural network inferences are represented in Figure 11. For example, we see that a pointlike source of 1 mJy, recorded in mock-observation raw data with feeding S/N = 1.0, is expected to be measured by our neural network with an S/N between ∼5–13. Likewise, after enough samples, the same source is expected to be measured by the telescope with an S/N of ∼27. In contrast, a single MF inference of a 1 mJy source is expected with an S/N below ∼2.5, while an MF-telescope measurement at the same flux would only slightly improve the S/N. MF-telescope measurements show a maximum S/N at fluxes above ∼2.5 mJy. This plateau is explained by the information loss at the skirts of the MF inference profile (as shown in Figure 6), which is intensified for brighter sources. Overall, our trained network reports a significant S/N enhancement of astrophysical sources with respect to raw signals and is superior to our tested MF technique.

As a side note, it is worth mentioning that our network obtained well-behaved inferences on temporal series generated from a spectral density absent in the training data sets. Given the variability of climate conditions along different observation nights, this capability to infer from diverse noise structures is an advantageous transfer learning capability to be further explored.

We should finally specify that our model was built in TENSORFLOW (Abadi et al. 2016) and trained in a tower-data configuration. The acceleration hardware consists of four Volta-100. The total training procedure required nearly 144 hr of GPU time. The final frozen model takes up 520 Mb in memory.

## 6. Conclusions

Structured noise in the time domain poses a significant challenge for communications and astronomy, especially under low-S/N conditions. Here, we have shown that a deep-learning approach with a representative training data set could significantly advance the treatment of structured noise. This contribution is of great interest for ground-based infrared and millimeter-wave astronomy in the advent of a new technological generation of continuum cameras and large telescopes because the turbulent atmosphere of Earth introduces highly structured noise that can be removed only through advanced data-processing techniques and physical modeling.

We propose a bidirectional RNN model with the ability to abstract long- and short-term nonlinear features from sequential data. Our architecture consists of bidirectional LSTM cells that assemble a many-to-many input/output configuration. We utilize sliding windows of atmospheric noise with an embedded astrophysical signal to feed and train the network to retrieve a clean version of the signal. This configuration allows us to capture some of the larger atmospheric fluctuations while optimizing memory space and processing capabilities.

Because neural networks need to learn from large amounts of representative data, we develop and present here a complete model for the IR scintillation spectral density, taking into account the Kolmogorov–Obukhov (KO) theory, the dry-air turbulent fluctuations (real part of the refractive index), and the anomalous dispersion caused by water-molecule emission lines (imaginary

part of the refractive index). Furthermore, we generate synthetic observations to train the network with a realistic type of noise, containing intrinsic structures induced possibly by instrumental effects. We have tried four training strategies and adopt our TCL-SR strategy, a useful combination of curriculum and transfer learning with intermediate reviewing blocks.

We test our trained network with several complexity levels, reporting more than 95% of source recovery (see Figure 7) at feeding $S/N \gtrsim 0.6$, 0.8, and 1, in the cases of white Gaussian, IR scintillation, and mock observational noise, respectively. We characterize random flux distortions caused by atmospheric-noise residuals, helping us design an efficient TCL-SR training strategy and calibrate inferences against flux bias. After an effective calibration, random flux distortions cancel out on average, leading to unbiased statistical measurements. It means that atmospheric-noise residuals shall be diluted in the map domain, provided a large enough number of temporal samples are stacked at every pixel. Furthermore, our network reports a significant $S/N$ enhancement of astronomical signals (see Figure 11). We notice that the network (single-shot) measurements improve the $S/N$ from $\sim$5 to 44 times, on average, compared to raw data. Meanwhile, the telescope (statistical) measurements improve the $S/N \sim$19–71 times compared to raw data.

By comparison, our tests with the simple MF technique deliver a poor detection rate on single inferences and look nonrobust to handle different nonwhite Gaussian noise complexities (more sophisticated versions like the adaptive MF may yield better results, though; Robey et al. 1992). Thus, we assert that our neural network presents the potential of a much more efficient approach to converge to accurate astrophysical measurements than more traditional filtering methods.

We conclude that our deep-learning methodology reports a satisfactory cleaning performance over structured-noise temporal series and considerably high astronomical-signal enhancement. Additionally, our training strategy shows some evidence of effectiveness over diverse degrees of simulated noise structure (using spectral densities not contained in the training data set).

It is worth stressing that although we have focused on pointlike sources, they represent a wide variety of applications for extragalactic astronomy, ranging from telescope-pointing calibrations to the observation of far and faint submillimeter galaxies. Regarding new-generation instruments, both atmospheric and pointlike source simulations are fairly representative realizations of the expected data. Nevertheless, more specialized simulations capturing each instrument's particular features may be generated from their observational data.

Our present study may be extended in many ways. For instance, growing the LSTM's long-term memory could make it possible to abstract longer correlations to handle higher-structure degrees or recover fainter sources. Mechanisms such as Attention Models (Vaswani et al. 2017) or Compressive Transformers (Rae et al. 2019) could provide interesting improvements as well. Multifeature networks may be another route for further studies, as it can be useful to train the network with environmental data taken at the observation site. For example, it would be useful to perform independent measurements of the temperature and specific-humidity structure parameters along the troposphere, which are well known to correlate with the recorded structured noise. In this direction, our physical IR scintillation model may be beneficial for further investigations attempting to disentangle the scintillation noise from a signal propagating through Earth's atmosphere. In turn, deep-learning research applied to atmospheric sciences could provide a parametric characterization of environmental variables from rough data, understanding which is of great interest not only for astronomy but also for a wide variety of fields where turbulence is involved.

The authors thankfully acknowledge the computer resources provided by the Laboratorio Nacional de Supercómputo del Sureste de México, CONACYT network of national laboratories. This project was possible owing to partial support from CONACYT research grants 237004, 490769, F.C. 2016/1848, and FORDECYT-297324. Finally, we would like to thank an anonymous referee for a critical review that significantly improved our paper.

## Appendix
## Propagation of Millimeter Waves through the Turbulent Atmosphere

In this appendix, we review atmospheric-turbulence physics responsible for imprinting highly structured noise on the recorded signals. Earth's atmosphere is a very complex medium; even during calm nights ideal for astronomical observations, turbulent fluctuations of wind velocity, air temperature, pressure, and specific humidity dominate the atmosphere dynamics. Although these thermodynamic variables are studied as climate phenomena, their small perturbations are entirely chaotic and make any deterministic description impossible.

Fortunately, Kolmogorov's theory (Pope 2000; Nieuwstadt et al. 2016; Vallis 2017) provides a successful description of turbulent phenomena by distinguishing three physical scales: (i) the outer scale $L_0$, characterized by large fluctuations (eddies) that are responsible for injecting mechanical energy into the system through progressive decays to smaller and smaller eddies. This energy cascade proceeds without (significant) loss down to (ii) the inner scale $l_0$, where, due to the air viscosity and molecular diffusion, all the energy is finally dissipated into heat. (iii) The universal inertial regime is an intermediate range of longitudes $l_0 \ll l \ll L_0$, where dissipation is negligible and turbulence dynamics is locally homogeneous and isotropic. For Earth's atmosphere, typically, $L_0 \sim 5$–10 km and $l_0 \sim 1$ mm. Further, at millimeter wavelengths, light interacts mainly along its path through the troposphere, which has a height of $H \sim 10$ km. For a receiver pointing to a zenith angle $\theta$, the Fresnel length (Wheelon 2001, 2003) (which is comparable to the size of eddies causing light diffraction) is at most $\sqrt{\lambda H \sec\theta}$. Because, for typical observation angles, it is true that $l_0 \ll \sqrt{\lambda H \sec\theta} \ll L_0$, Kolmogorov's universal inertial regime provides a proper prescription of the atmospheric turbulence affecting millimeter astronomy.

### *A.1. Wave Equations and the Turbulence Description*

The theory of wave propagation through turbulent media was developed in Tatarskii's (1961) seminal treatise (Tatarskii 1961), followed by his own review (Tatarskii 1971). Subsequent extensions have been summarized in more recent reviews (Beland 1993; Wheelon 2001, 2003; Andrews & Phillips 2005). However, the fluctuations in the refractive index's imaginary part—which account for absorption and reemission (anomalous dispersion) of light—are often overlooked. Given that the former is a dominant effect at infrared and millimeter wavelengths, below, we shall review the equations governing the propagation of infrared light through a turbulent atmosphere.

An electromagnetic wave propagates through the atmosphere obeying the wave equation

$$\nabla^2 E(\boldsymbol{r}) + k^2 n^2(\boldsymbol{r}) E(\boldsymbol{r}) = 0, \tag{A1}$$

where $E$ is a component of the electric field, $k = |\boldsymbol{k}|$ is the wavenumber amplitude, and $n(\boldsymbol{r})$ is the atmosphere's refractive index. The random nature of $n(\boldsymbol{r})$ prevents us from obtaining any exact solution to the wave equation. Nevertheless, approximate descriptions are still possible, like the Rytov approximation, which uses a transformation $E(\boldsymbol{r}) = E_0(\boldsymbol{r})\, e^{\psi(\boldsymbol{r})}$ in order to perform a series expansion in terms of the surrogate function $\psi = \psi_0 + \psi_1 + \psi_2 + ...$, where the field $E_0$ obeys the vacuum wave equation.

We can describe the random behavior of the refractive index in terms of small variations,

$$n(\boldsymbol{r}) = 1 + n_1(\boldsymbol{r}), \tag{A2}$$

where we have defined $\langle n(\boldsymbol{r})\rangle = 1$. In nature, $n_1(\boldsymbol{r}) = n_R(\boldsymbol{r}) - i n_I(\boldsymbol{r})$ is a complex passive scalar field whose real part is responsible for variations of the propagating wave due to refraction, and the imaginary part is responsible for absorption. For visible light $n_I$ is usually negligible; meanwhile, at infrared and millimeter wavelengths, the atmospheric molecules cause resonant absorption and reemission of light; thus, anomalous dispersion is nonnegligible. Considering up to linear terms ($n^2 \approx 1 + 2n_1$), we can see that the refractive index conveniently appears as the source in a first-order wave equation,

$$\nabla^2 \psi_1(\boldsymbol{r}) + 2\nabla\psi_0(\boldsymbol{r}) \cdot \nabla\psi_1(\boldsymbol{r}) = -2k^2 n_1(\boldsymbol{r}). \tag{A3}$$

With an additional transformation $\psi_1(\boldsymbol{r}) = Q(\boldsymbol{r}) e^{-\psi_0(\boldsymbol{r})}$, one recovers the familiar Helmholtz equation,

$$\nabla^2 Q + k^2 Q = -2k^2\, n_1(\boldsymbol{r}) e^{\psi_0(\boldsymbol{r})}, \tag{A4}$$

whose solution is well known in terms of the Green function $G(\boldsymbol{R}, \boldsymbol{r}) = e^{i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{r})}/(4\pi|\boldsymbol{R} - \boldsymbol{r}|)$,

$$\psi_1(\boldsymbol{r}_o) = -2k^2 \int d^3r\; G(\boldsymbol{R}, \boldsymbol{r}) \frac{E_0(\boldsymbol{r})}{E_0(\boldsymbol{r}_o)} n_1(\boldsymbol{r}), \tag{A5}$$

where $\boldsymbol{r}_o$ and $\boldsymbol{r}$ represent the observer and scattering positions, respectively. The first-order Rytov solution is readily found from $E_1(\boldsymbol{r}_o) = E_0(\boldsymbol{r}_o) e^{\psi_1(\boldsymbol{r}_o)}$. Implicitly, the former equations contain the expressions for the amplitude $A$ and phase $\varphi$ of the random-field fluctuation,

$$E_1(\boldsymbol{r}_o) = A(\boldsymbol{r}_o) \exp\left[i(\phi_0 + \varphi(\boldsymbol{r}_o)\right], \tag{A6}$$

where $\phi_0$ is the mean-field value of the phase fluctuations. Yet, instead of the amplitude itself, it is commonly more useful to work with the log-amplitude $\chi \equiv \log(A/E_0)$. Consequently, the astronomical-signal fluctuations propagating through the atmosphere are described by the complex scattering integral,

$$\chi(\boldsymbol{r}_o) + i\varphi(\boldsymbol{r}_o) = -2k^2 \int d^3r\; G(\boldsymbol{r}_o, \boldsymbol{r}) \frac{E_0(\boldsymbol{r})}{E_0(\boldsymbol{r}_o)} n_1(\boldsymbol{r}). \tag{A7}$$

Because the observed light intensity is the field square amplitude, $I/I_0 = |E_1|^2/|E_0|^2 = e^{2\chi}$, random variations of $\chi$ are a proxy to quantify the noise induced by the turbulent atmosphere, an effect also known as scintillation. Notice that as $\langle n_1\rangle = 0$, we also have $\langle\chi\rangle = 0$ by definition. However, the second-order statistical moments are nonnegligible; indeed, the scintillation noise variance can be computed from

$$\sigma_I^2 = \frac{\langle [I - \langle I\rangle]^2\rangle}{\langle I\rangle^2} \approx 4\langle\chi^2\rangle, \tag{A8}$$

where

$$\langle\chi^2\rangle = 4k^4 \int d^3r \int d^3r' \Bigg\langle \mathrm{Re}, ,\{, G, ,(, \boldsymbol{r}_o,, ,\boldsymbol{r},), , \frac{E_0(\boldsymbol{r})}{E_0(\boldsymbol{r}_o)}, n_1, ,(\boldsymbol{r}) \times \mathrm{Re}, ,\{, G, ,(, \boldsymbol{r}_o,, ,\boldsymbol{r}',), , \frac{E_0(\boldsymbol{r}')}{E_0(\boldsymbol{r}_o)}, n_1, ,(, \boldsymbol{r}',),\} \Bigg\rangle. \tag{A9}$$

To compute this double integral we need some prior knowledge of the complex refractive-index autocovariance function $\langle n_1(\boldsymbol{r}) n_1(\boldsymbol{r}')\rangle$. For cell sizes $l_0 \ll |\boldsymbol{r} - \boldsymbol{r}'| \ll L_0$, well inside the universal inertial regime, we know that turbulence is homogeneous, thus, the autocovariance can only depend on the relative position $\boldsymbol{r} - \boldsymbol{r}'$ but not on the individual vectors $\boldsymbol{r}$ and $\boldsymbol{r}'$. Furthermore, each component of the autocovariance is associated with its Fourier power spectrum by

$$\begin{aligned}
\langle n_R(\boldsymbol{r})\, n_R(\boldsymbol{r}')\rangle &= \int d^3\kappa\; e^{i\boldsymbol{\kappa}\cdot(\boldsymbol{r}-\boldsymbol{r}')}\, \Phi_R(\boldsymbol{\kappa}),\\
\langle n_I(\boldsymbol{r})\, n_I(\boldsymbol{r}')\rangle &= \int d^3\kappa\; e^{i\boldsymbol{\kappa}\cdot(\boldsymbol{r}-\boldsymbol{r}')}\, \Phi_I(\boldsymbol{\kappa}),\\
\langle n_R(\boldsymbol{r})\, n_I(\boldsymbol{r}')\rangle &= \int d^3\kappa\; e^{i\boldsymbol{\kappa}\cdot(\boldsymbol{r}-\boldsymbol{r}')}\, \Phi_{\mathrm{IR}}(\boldsymbol{\kappa}),
\end{aligned} \tag{A10}$$

where $\boldsymbol{\kappa}$ is the eddy wavevector. Moreover, provided the turbulence is isotropic inside the universal inertial regime, we know that $\Phi_n$ should not depend on any specific direction but just on the magnitude of the wavevector $\kappa = 2\pi/l$. According to a simple dimensional analysis, Kolmogorov and Obukhov showed that in this approximation $\Phi_n$ is given by the KO spectrum, $\Phi_n(\kappa) = 0.033\; C_n^2\; \kappa^{-11/3}$ (Kolmogorov 1991; Obhukov 1941), where the structure constant $C_n^2$ must be determined experimentally. For cell sizes barely comparable to the outer scale, $|\boldsymbol{r} - \boldsymbol{r}'| \lesssim L_0$, the KO spectrum is still valid locally, but $C_n^2$ needs to be corrected as a function of position. For Earth's atmosphere, the structure parameter depends mainly on the height above the observer $C_n^2(\boldsymbol{r}) \to C_n^2(z)$. More generally, the turbulence spectrum can be characterized by

$$\Phi_n(\boldsymbol{\kappa}; z) = C_n^2(z)\, \Phi_0(\kappa), \tag{A11}$$

where from here on the subscript $n$ denotes $R$, $I$, or $IR$. $\Phi_0(\kappa)$ could be written as the original KO spectrum $0.033\, \kappa^{-11/3}$, but because both the outer and dissipation scales suppress power at the largest and smallest scales, respectively, several empirical models have been developed in order to account for deviations from the inertial regime; e.g., Tatarskii's (1961) spectrum (Tatarskii 1961), von Kármán's (1948) spectrum (Von Karman 1948), and yet a more versatile version (that contains the former two) was proposed by Pope (2000),

$$\begin{aligned}
\Phi_0(\kappa) &= 0.033\; \kappa^{-11/3} f_L(\kappa L_0) f_l(\kappa l_0),\\
f_L(\kappa L_0) &= \left(\frac{\kappa L_0}{(\kappa L_0)^2 + c_L^2}\right)^{5/3+p_0},\\
f_l(\kappa l_0) &= e^{-\beta_0\, \kappa l_0},
\end{aligned} \tag{A12}$$

with $L_0 = 5$–$10$ km, $c_L = 2.6$, and $p_0 = 4$ as some typical values. Here we will not consider dissipation scale effects, so we choose $\beta_0 = 0$.

A careful analysis shows that $C_R^2$ depends mainly on the fluctuations of temperature and humidity, while $C_I^2$ depends mainly on the humidity fluctuations (Hill et al. 1980). Yet, it is not possible to determine any of the structure parameters from first principles, and they have to be measured directly (see e.g., (Osborn et al. 2017; Qian et al. 2018; Avila et al. 2019; Avila 2021). For many years, these kinds of experiments (in different locations and altitudes) have generated empirical profiles for $C_n^2(z)$; one of the most popular is the so-called Hufnagel-Valley 5/7 (HV5/7) model (Beland 1993),

$$C_n^2(z) = 1.7 \times 10^{-14} e^{-z/100} + 2.7 \times 10^{-16} e^{-z/1500} + 8.2 \times 10^{-26} \upsilon_{\rm rms}^2 \left(\frac{z}{1000}\right) e^{-z/1000}, \tag{A13}$$

where $\upsilon_{\rm rms}$ is the transverse rms wind speed. Ideally, one needs special measurements of the structure parameter but these measurements are not always available. Nonetheless, we can make use of an empirical profile like the HV5/7 model to make some approximations. By substituting Equations (A10), (A11), (A12), and (A13) into Equation (A9), we can estimate the second-order statistical moments of the astronomical signal passing through the turbulent atmosphere.

### A.2. Scintillation Variance

For simplicity, we may initially think of a point receiver and take the origin at the receiver position $\boldsymbol{r}_o = 0$. Consider the incident astronomical signal as a plane wave propagating from the zenith $E_0(\boldsymbol{r}) = \mathcal{E}_0\, e^{-ikz}$. We can also use the approximation of a small scattering angle, which is valid under conditions of weak scattering, in cylindrical coordinates, $\sqrt{\rho^2 + z^2} \approx z + \rho^2/2z$. The argument of the scattering integral is, in this case,

$$G(0, \boldsymbol{r}) \frac{E_0(\boldsymbol{r})}{E_0(0)} n_1(\boldsymbol{r}) = \frac{1}{4\pi z}\left[ n_R(\boldsymbol{r}) \cos\left(\frac{k\rho^2}{2z}\right) + n_I(\boldsymbol{r}) \sin\left(\frac{k\rho^2}{2z}\right)\right] + \frac{i}{4\pi z}\left[ n_R(\boldsymbol{r}) \sin\left(\frac{k\rho^2}{2z}\right) + n_I(\boldsymbol{r}) \cos\left(\frac{k\rho^2}{2z}\right)\right]. \tag{A14}$$

This can be substituted in Equation (A9) with the aid of Equation (A10) to compute the fluctuation autocovariance, which is the sum of the real, imaginary, and the imaginary-real contributions,

$$\langle \chi^2 \rangle = \langle \chi_R^2 \rangle + \langle \chi_I^2 \rangle + \langle \chi_{\rm IR}^2 \rangle. \tag{A15}$$

Let us define the weighting functions

$$A_n(\boldsymbol{\kappa}) = \int d^3 r \frac{C_n(z)}{z} \cos\left(\frac{k\rho^2}{2z}\right) e^{i\boldsymbol{\kappa}\cdot\boldsymbol{r}},$$
$$B_n(\boldsymbol{\kappa}) = \int d^3 r \frac{C_n(z)}{z} \sin\left(\frac{k\rho^2}{2z}\right) e^{i\boldsymbol{\kappa}\cdot\boldsymbol{r}}, \tag{A16}$$

in order to express each component as

$$\langle \chi_R^2 \rangle = \frac{k^4}{(2\pi)^2} \int d^3\kappa\, \Phi_0(\kappa)\, A_R(\boldsymbol{\kappa}) A_R(-\boldsymbol{\kappa}),$$
$$\langle \chi_I^2 \rangle = \frac{k^4}{(2\pi)^2} \int d^3\kappa\, \Phi_0(\kappa)\, B_I(\boldsymbol{\kappa}) B_I(-\boldsymbol{\kappa}), \tag{A17}$$

$$\langle \chi_{\rm IR}^2 \rangle = \frac{k^4}{(2\pi)^2} \int d^3\kappa\, \Phi_0(\kappa) [A_R(\boldsymbol{\kappa}) B_I(-\boldsymbol{\kappa}) + B_R(\boldsymbol{\kappa}) A_I(-\boldsymbol{\kappa})],$$

where $\Phi_0(\kappa)$ is given in Equation (A12). The integrals over the angular coordinates are easily done under the assumption of local isotropy,

$$\langle \chi_R^2 \rangle = (2\pi k)^2 \sec\vartheta \int_{h_{\rm min}}^{h_{\rm max}} dz\, C_R^2(z) \times \int_0^\infty d\kappa\, \kappa\, \Phi_0(\kappa) \sin^2\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right),$$
$$\langle \chi_I^2 \rangle = (2\pi k)^2 \sec\vartheta \int_{h_{\rm min}}^{h_{\rm max}} dz\, C_I^2(z) \times \int_0^\infty d\kappa\, \kappa\, \Phi_0(\kappa) \cos^2\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right),$$
$$\langle \chi_{\rm IR}^2 \rangle = 2(2\pi k)^2 \sec\vartheta \int_{h_{\rm min}}^{h_{\rm max}} dz\, C_{\rm IR}^2(z) \times \int_0^\infty d\kappa\, \kappa\, \Phi_0(\kappa) \sin\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right) \times \cos\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right). \tag{A18}$$

where we have made the line-of-sight change of variable $z \to s = z \sec\vartheta$, with $\vartheta$ the zenith angle, $h_{\rm min}$ is the altitude at which the observatory sits on, and $h_{\rm max}$ is the maximum relevant height, in our case, the end of the troposphere. Several investigations (Lee & Harp 1969; Hill et al. 1980; Lüdi et al. 2005; Solignac et al. 2012; Yuan et al. 2015) have shown that $C_R^2$ is typically $10^3$–$10^7$ times larger than $C_I^2$. This might naively lead to the wrong assumption that the imaginary-part contributions would be negligible with respect to the purely real part, but as we will see, this is not the case for infrared and millimeter wavelengths.

### A.3. Covariance and Aperture Averaging

Consider a pair of point receivers separated by a (constant) baseline vector $\boldsymbol{\rho}_b$. The spatial covariance between two point receivers is very similar to the previous computation in Equations (A17),

$$\langle \chi_I(\boldsymbol{r}_o) \chi_I(\boldsymbol{r}_o + \boldsymbol{\rho}_b) \rangle = \frac{k^4}{(2\pi)^2} \int d^3\kappa\, \Phi_0(\kappa)\, e^{i\boldsymbol{\kappa}\cdot\boldsymbol{\rho}_b}\, B_I(\boldsymbol{\kappa}) B_I(-\boldsymbol{\kappa}).$$

The expressions for the $R$ and IR components are closely similar. Now, the integration over the angular coordinates

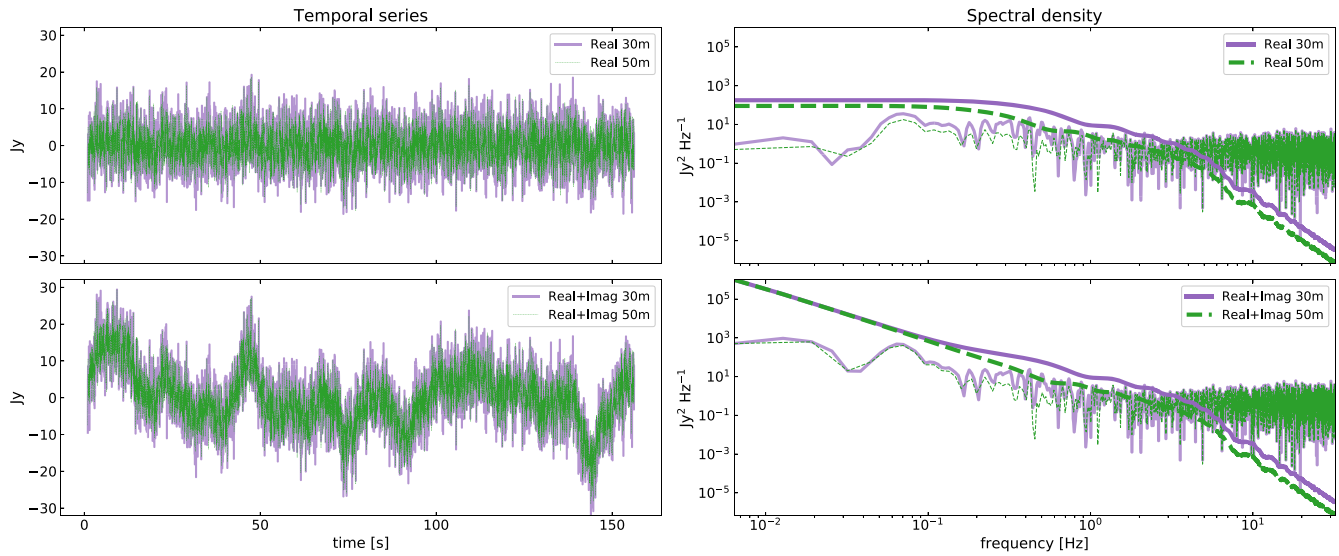


**Figure 12.** Theoretical realizations of the scintillation noise (left) and its corresponding spectral density (right), caused by the atmospheric turbulence perturbing the electromagnetic waves propagating through Earth's troposphere. The infrared and millimeter waves are affected by the dry air—or real refractive index—turbulent fluctuations (top panel), but more dominantly, they are affected by the water vapor—or imaginary refractive index—turbulent fluctuations (bottom). A circular telescope aperture with 30 and 50 m of diameter is considered to compute each noise simulation. Theoretical scintillation power spectra are represented by thick lines (corresponding to $\mathcal{P}$ in Equation (5)). Random-noise realizations are created according to Equation (5) and represented by thin lines in the spectral densities and temporal series.

yields

$$\begin{aligned}
\langle \chi_R(\boldsymbol{r}_o)\chi_R(\boldsymbol{r}_o+\boldsymbol{\rho}_b)\rangle &= (2\pi k)^2 \sec\vartheta \int_{h_{\min}}^{h_{\max}} dz\, C_R^2(z) \\
&\times \int_0^\infty d\kappa\, \kappa\, \Phi_0(\kappa)\, J_0(\kappa\rho_b) \sin^2\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right), \\
\langle \chi_I(\boldsymbol{r}_o)\chi_I(\boldsymbol{r}_o+\boldsymbol{\rho}_b)\rangle &= (2\pi k)^2 \sec\vartheta \int_{h_{\min}}^{h_{\max}} dz\, C_I^2(z) \\
&\times \int_0^\infty d\kappa\, \kappa\, \Phi_0(\kappa)\, J_0(\kappa\rho_b) \cos^2\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right), \\
\langle \chi_R(\boldsymbol{r}_o)\chi_I(\boldsymbol{r}_o+\boldsymbol{\rho}_b)\rangle &= 2(2\pi k)^2 \sec\vartheta \int_{h_{\min}}^{h_{\max}} dz\, C_{\mathrm{IR}}^2(z) \\
&\times \int_0^\infty d\kappa\, \kappa\, \Phi_0(\kappa)\, J_0(\kappa\rho_b) \sin\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right) \\
&\times \cos\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right),
\end{aligned} \tag{A19}$$

where $J_0$ is the zeroth-order spherical Bessel function. Using Equation (A19), we want to correct the scintillation variance in Equation (A18) by considering the electromagnetic wave collecting average of a large telescope. To do so, we can assume for simplicity a circular dish of radius $a_r$ (more realistic aperture corrections can be considered; Osborn 2015; Osborn et al. 2015). Then, we need to sum over all point-receiver spatial covariances located at pair positions $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ inside the circular aperture, such that their (now variable) baseline distance $\rho_b = |\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2| = [\rho_1^2 + \rho_2^2 - 2\rho_1\rho_2\cos(\phi_1 - \phi_2)]^{1/2}$ is $\rho_b \leqslant a_r$. Then, the aperture-average scintillation variance, defined by

$$\begin{aligned}
\overline{\langle \chi^2 \rangle} &= \frac{1}{\pi^2 a_r^2} \int_0^{a_r} d\rho_1\, \rho_1 \int_0^{a_r} d\rho_2\, \rho_2 \\
&\times \int_0^{2\pi} d\phi_1 \int_0^{2\pi} d\phi_2 \langle \chi(\rho_1, \phi_1)\chi(\rho_2, \phi_2)\rangle,
\end{aligned}$$

can be computed for each component,

$$\begin{aligned}
\overline{\langle \chi_R^2 \rangle} &= (2\pi k)^2 \sec\vartheta \int_{h_{\min}}^{h_{\max}} dz\, C_R^2(z) \\
&\times \int_0^\infty d\kappa\, \kappa\, \Phi_0(\kappa) \sin^2\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right)\left(\frac{2J_1(\kappa a_r)}{\kappa a_r}\right)^2, \\
\overline{\langle \chi_I^2 \rangle} &= (2\pi k)^2 \sec\vartheta \int_{h_{\min}}^{h_{\max}} dz\, C_I^2(z) \\
&\times \int_0^\infty d\kappa\, \kappa\, \Phi_0(\kappa) \cos^2\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right)\left(\frac{2J_1(\kappa a_r)}{\kappa a_r}\right)^2, \\
\overline{\langle \chi_{\mathrm{IR}}^2 \rangle} &= 2(2\pi k)^2 \sec\vartheta \int_{h_{\min}}^{h_{\max}} dz\, C_{\mathrm{IR}}^2(z) \\
&\times \int_0^\infty d\kappa\, \kappa\, \Phi_0(\kappa) \sin\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right) \\
&\times \cos\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right)\left(\frac{2J_1(\kappa a_r)}{\kappa a_r}\right)^2,
\end{aligned} \tag{A20}$$

where $J_1$ is the first-order spherical Bessel function. The overall effect of a large dish is to reduce the scintillation noise. But notice that the $\sin^2(x)$ function has significantly less overlap with $J_1^2(x)/x^2$ than $\cos^2(x)$. Consequently, the noise reduction is quite notable for the refractive-index real part; however, the scintillation noise due to the imaginary part remains almost unchanged irrespective of dish size (see Figure 12).

### *A.4. Temporal Power Spectrum*

For the next step, we need to invoke Taylor's frozen turbulence hypothesis (Taylor 1938), which postulates that for short enough periods of time and for slowly changing wind velocities, the turbulent atmosphere can be seen as a frozen screen moving in a horizontal plane along the line of sight. In other words, Taylor's hypothesis states an equivalence between spatial and temporal covariances, allowing us to make the replacement $\rho_b \to \tau\upsilon$ in Equation (A19), in terms of the wind

velocity transverse component $\upsilon(z)$. Thus, the temporal covariance can be written as

$$\overline{\langle \chi_R(t)\chi_R(t+\tau)\rangle} = (2\pi k)^2 \sec\vartheta \int_{h_{\min}}^{h_{\max}} dz\, C_R^2(z) \times \int_0^\infty d\kappa\, \kappa\, \Phi_n^{(0)}(\kappa)\, J_0[\kappa\tau\upsilon(z)] \times \sin^2\left(\frac{\kappa^2 z \sec\vartheta}{2k}\right) \times \left(\frac{2J_1(\kappa a_r)}{\kappa a_r}\right)^2, \tag{A21}$$

with similar expressions for the $I$ and IR components and where we already include the aperture-averaging factor. The frequency power density, defined by $\overline{\langle\chi^2\rangle} = \int d\omega\, P_\chi(\omega)$, can be computed from the Fourier transformation of the temporal covariance,

$$P_\chi(\omega) = \int_{-\infty}^{\infty} d\tau\, e^{i\omega\tau}\, \overline{\langle\chi(t)\chi(t+\tau)\rangle}. \tag{A22}$$

The full power spectrum is the sum of the complex refractive-index contributions, $P_\chi = P_R(\omega) + P_I(\omega) + P_{\rm IR}(\omega)$. The temporal integration can be done analytically under the condition $\omega < \kappa\upsilon < \infty$, leading us to

$$\begin{aligned}
P_R(\omega) &= 2(2\pi k)^2 \sec\vartheta \int_{h_{\min}}^{h_{\max}} dz\, C_R^2(z) \times \int_{\omega/\upsilon(z)}^{\infty} d\kappa\, \frac{\kappa\,\Phi_0(\kappa)}{\sqrt{(\kappa\upsilon)^2-\omega^2}} \times \sin^2\left(\frac{\kappa^2 z\sec\vartheta}{2k}\right)\left(\frac{2J_1(\kappa a_r)}{\kappa a_r}\right)^2,\\
P_I(\omega) &= 2(2\pi k)^2 \sec\vartheta \int_{h_{\min}}^{h_{\max}} dz\, C_I^2(z) \times \int_{\omega/\upsilon(z)}^{\infty} d\kappa\, \frac{\kappa\,\Phi_0(\kappa)}{\sqrt{(\kappa\upsilon)^2-\omega^2}} \times \cos^2\left(\frac{\kappa^2 z\sec\vartheta}{2k}\right)\left(\frac{2J_1(\kappa a_r)}{\kappa a_r}\right)^2,\\
P_{\rm IR}(\omega) &= 4(2\pi k)^2 \sec\vartheta \int_{h_{\min}}^{h_{\max}} dz\, C_{\rm IR}^2(z) \times \int_{\omega/\upsilon(z)}^{\infty} d\kappa\, \frac{\kappa\,\Phi_0(\kappa)}{\sqrt{(\kappa\upsilon)^2-\omega^2}} \sin\left(\frac{\kappa^2 z\sec\vartheta}{2k}\right) \times \cos\left(\frac{\kappa^2 z\sec\vartheta}{2k}\right)\left(\frac{2J_1(\kappa a_r)}{\kappa a_r}\right)^2.
\end{aligned} \tag{A23}$$

These integrals can be easily evaluated numerically once some profiles for the structure parameters and wind speed have been specified. Bufton's wind profile (Andrews & Phillips 2005) is a simple empirical model suggesting a Gaussian shape for $\upsilon(z)$, with a stream peaking barely below the troposphere's height. Notice that the structure parameter integration over $z$ contributes effectively as an overall numerical prefactor; thus, we can use Equation (A13) as the altitude profile for the three structure parameters, and modulate parametrically their relative power.

From the theoretical scintillation spectral density in Equations (A23), we can generate many simulations of atmospheric temporal series using Equation (5). Figure 12 exemplifies these random realizations, with 30 and 50 m telescope-aperture diameters, and observations at the zenith. The upper-left panel shows two dry-air scintillation (real refractive index) simulated time streams. The corresponding theoretical spectral densities are represented by the smooth curves on the upper-right panel, which are convolved with a Gaussian-random series to simulate an observation. The lower-left panel shows two wet-air scintillation (real+imaginary refractive index) simulations with $C_I^2 = 10^{-5}C_R^2$. The lower-right panel shows the corresponding theoretical and random spectral densities. From the theoretical spectral density, we observe a high-frequency power reduction by the 50 m aperture compared to the 30 m one; it means that a large telescope can notably reduce the dry-air scintillation noise. On the other hand, the low-frequency power remains equal for both 30 and 50 m; it means that a larger dish cannot defeat the wet-air scintillation noise.

In summary, Equation (A23) represents the general scintillation spectral density at infrared and millimeter waves, derived from turbulence physics and telescope-aperture-averaging factors. The large-scale highly structured noise prevailing in the observational temporal series is mainly caused by the amplitude of the imaginary-part structure parameter $C_I(z)$. This derivation is presented for an astronomical application for the first time and should be the preferred model to simulate atmospheric noise at infrared and millimeter waves.

## ORCID iDs

Iván Rodríguez-Montoya https://orcid.org/0000-0001-7164-7037
David Sánchez-Argüelles https://orcid.org/0000-0002-7344-9920
Itziar Aretxaga https://orcid.org/0000-0002-6590-3994

## References

Abadi, M., Barham, P., Chen, J., et al. 2016, in 12th USENIX Symp. on Operating Systems Design and Implementation (OSDI 16), 265
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, PhRvL, 116, 061102
Ade, P., Aghanim, N., Arnaud, M., et al. 2016, A&A, 596, A100
Aghanim, N., Altieri, B., Arnaud, M., et al. 2015, A&A, 582, A30
Allen, B., Anderson, W. G., Brady, P. R., Brown, D. A., & Creighton, J. D. E. 2012, PhRvD, 85, 122006
Andrews, L. C., & Phillips, R. L. 2005, Laser Beam Propagation through Random Media (Bellingham, WA: SPIE Optical Engineering Press), 152
Austermann, J., Beall, J., Bryan, S., et al. 2018, JLTP, 193, 120
Avila, R. 2021, MNRAS, 507, L11
Avila, R., Valdés-Hernández, O., Sánchez, L., et al. 2019, MNRAS, 490, 1397
Beland, R. R. 1993, The Infrared and Electro-Optical Systems Handbook, Vol. 2 (Bellingham, WA: SPIE Optical Engineering Press), 157, https://apps.dtic.mil/dtic/tr/fulltext/u2/a364019.pdf#page=170
Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. 2011, in 25th Annual Conf. on Neural Information Processing Systems (NIPS 2011) (San Diego, CA: Neural Information Processing Systems Foundation), https://proceedings.neurips.cc/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf
Blain, A. W., Smail, I., Ivison, R., Kneib, J.-P., & Frayer, D. T. 2002, PhR, 369, 111
Boyat, A. K., & Joshi, B. K. 2015, arXiv:1505.03489
Brien, T. L. R., Ade, P. A. R., Barry, P. S., et al. 2018, Proc. SPIE, 10708, 107080M
Bryan, S., Austermann, J., Ferrusca, D., et al. 2018, Proc. SPIE, 10708, 107080J
Casey, C., Narayanan, D., & Cooray, A. 2014, PhR, 541, 45
Castillo-Dominguez, E., Ade, P., Barry, P., et al. 2018, JLTP, 193, 1010
Cho, K., Van Merriënboer, B., Gulcehre, C., et al. 2014, arXiv:1406.1078
Church, S. 1995, MNRAS, 272, 551
Cuoco, E., Cella, G., & Guidi, G. M. 2004, CQGra, 21, S801
DeNigris, N., Wilson, G., Eiben, M., et al. 2020, JLTP, 199, 789
Downes, T., Welch, D., Scott, K., et al. 2012, MNRAS, 423, 529

Errard, J., Ade, P., Akiba, Y., et al. 2015, ApJ, 809, 63
George, D., & Huerta, E. 2018, PhRvD, 97, 044039
Gómez, S. L. S., González-Gutiérrez, C., Alonso, E. D., et al. 2019, PASP, 131, 108012
Graves, A., & Schmidhuber, J. 2005, NN, 18, 602
Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. 2017, IEEE Trans. Neural Netw. Learn. Syst., 28, 2222
Hill, R., Clifford, S. F., & Lawrence, R. S. 1980, JOSA, 70, 1192
Hochreiter, S., Bengio, Y., Frasconi, P., Schmidhuber, J., et al. 2001, Int. J. Uncertain., Fuzziness Knowl.-Based Syst., 6, 107
Hochreiter, S., & Schmidhuber, J. 1997, Neural Comput., 9, 1735
Holland, W., Bintley, D., Chapin, E., et al. 2013, MNRAS, 430, 2513
Hughes, D. H., Jáuregui Correa, J.-C., Schloerb, F. P., et al. 2010, Proc. SPIE, 7733, 773312
Jamal, S., & Bloom, J. S. 2020, ApJS, 250, 30
Jordan, M. I. 1997, Neural-Network Models of Cognition (Amsterdam: North-Holland), 471
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Kirchgässner, G., Wolters, J., & Hassler, U. 2012, Introduction to Modern Time Series Analysis (Berlin: Springer)
Kolmogorov, A. N. 1991, RSPSA, 434, 9
Kovács, A. 2008, Proc. SPIE, 7020, 70201S
Lay, O. 1997, A&AS, 122, 535
Lee, R. W., & Harp, J. 1969, Proc. IEEE, 57, 375
Lüdi, A., Beyrich, F., & Mätzler, C. 2005, BoLMe, 117, 525
Luo, G. 2016, Netw. Model. Anal. Health Inform. Bioinform., 5, 18
Mairs, S., Johnstone, D., Kirk, H., et al. 2015, MNRAS, 454, 2557
Martinache, C., Rettura, A., Dole, H., et al. 2018, A&A, 620, A198
Marulanda, J. P., Santa, C., & Romano, A. E. 2020, PhLB, 810, 135790
McCloskey, M., & Cohen, N. J. 1989, Psychology of Learning and Motivation (New York: Academic), 109
Milotti, E. 2002, arXiv:physics/0204033
Montaña, A., Chávez Dagostino, M., Aretxaga, I., et al. 2019, MmSAI, 90, 632
Nieuwstadt, F. T., Westerweel, J., & Boersma, B. J. 2016, Turbulence: Introduction to Theory and Applications of Turbulent Flows (Berlin: Springer)
Obhukov, A. 1941, Izv. Akad. Nauk SSSR, Ser. Geofiz, 5, 453
Osborn, J. 2015, MNRAS, 446, 1305
Osborn, J., Butterley, T., Townson, M., et al. 2017, MNRAS, 464, 3998
Osborn, J., Föhring, D., Dhillon, V., & Wilson, R. 2015, MNRAS, 452, 1707
Pan, S. J., & Yang, Q. 2009, IEEE Trans. Knowl. Data Eng., 22, 1345
Pascanu, R., Mikolov, T., & Bengio, Y. 2013, in Proc. of Machine Learning Research 28, Proc. of the 30th Int. Conf. on Machine Learning, ed. S. Dasgupta & D. McAllester (MLResearchPress), 1310, https://proceedings.mlr.press/v28/pascanu13.html
Pope, S. B. 2000, Turbulent Flows (Cambridge: Cambridge Univ. Press)
Qian, X., Yao, Y., Wang, H., et al. 2018, PASP, 130, 125002
Rae, J. W., Potapenko, A., Jayakumar, S. M., & Lillicrap, T. P. 2019, arXiv:1911.05507
Ratcliff, R. 1990, Psychol. Rev., 97, 285
Robey, F., Fuhrmann, D., Kelly, E., & Nitzberg, R. 1992, IEEE Trans. Aerosp. Electron. Syst., 28, 208
Rodríguez-Montoya, I., Sánchez-Argüelles, D., Aretxaga, I., et al. 2018, ApJS, 235, 12
Sayers, J., Golwala, S. R., Ade, P. A. R., et al. 2010, ApJ, 708, 1674
Schuster, M., & Paliwal, K. K. 1997, ITSP, 45, 2673
Scott, K., Austermann, J., Perera, T. A., et al. 2008, MNRAS, 385, 2225
Solignac, P. A., Brut, A., Selves, J.-L., Béteille, J.-P., & Gastellu-Etchegorry, J.-P. 2012, BoLMe, 143, 261
Tatarskii, V. I. 1961, Wave Propagation in a Turbulent Medium (New York: McGraw-Hill)
Tatarskii, V. I. 1971, The Effects of the Turbulent Atmosphere on Wave Propagation (Jerusalem: Israel Program for Scientific Translations)
Taylor, G. I. 1938, RSPSA, 164, 476
Turin, G. 1960, IRE Trans. Inform. Theory, 6, 311
Vallis, G. K. 2017, Atmospheric and Oceanic Fluid Dynamics (Cambridge: Cambridge Univ. Press)
Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, Advances in Neural Information Processing Systems 30, ed. I. Guyon et al., 5998, https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
Von Karman, T. 1948, PNAS, 34, 530
Wang, Y., Azam, A., Wilson, M. C., Neville, A., & Morina, A. 2021, J. Eng. Tribol., 235, 2640
Wang, Y., Liu, Y., Zhang, G., & Wang, Y. 2018, J. Tribol., 140, 021403
Weiss, K., Khoshgoftaar, T. M., & Wang, D. 2016, J. Big Data, 3, 9
Wheelon, A. D. 2001, Electromagnetic Scintillation, Vol. 1 (Cambridge: Cambridge Univ. Press)
Wheelon, A. D. 2003, Electromagnetic Scintillation, Vol. 2 (Cambridge: Cambridge Univ. Press)
Wilson, G., Austermann, J., Perera, T., et al. 2008, MNRAS, 386, 807
Yuan, R., Luo, T., Sun, J., et al. 2015, Atmos. Chem. Phys., 15, 2521
Zaremba, W., & Sutskever, I. 2014, arXiv:1410.4615
Zavala, J. A., Montaña, A., Hughes, D. H., et al. 2018, NatAs, 2, 56
Zhang, Z., Geiger, J., Pohjalainen, J., et al. 2018, ACM Trans. Intell. Syst. Technol., 9, 49