

# Radiometer Calibration using Machine Learning

S. A. K. Leeney<sup>1,2\*</sup>, H. T. J. Bevins<sup>1,2</sup>, E. de Lera Acedo<sup>1,2</sup>,  
 W. J. Handley<sup>2,3</sup>, C. Kirkham<sup>1,2</sup>, R. S. Patel<sup>1,2</sup>, J. Zhu<sup>1,4, 19</sup>,  
 D. Molnar<sup>1,2</sup>, J. Cumner<sup>1,2</sup>, D. Anstey<sup>1,2</sup>, K. Artuc<sup>1,2</sup>,  
 G. Bernardi<sup>5</sup>, M. Bucher<sup>6,7</sup>, S. Carey<sup>1</sup>, J. Cavillot<sup>8</sup>, R. Chiello<sup>9</sup>,  
 W. Croukamp<sup>7</sup>, D. I. L. de Villiers<sup>7</sup>, J. A. Ely<sup>1</sup>, A. Fialkov<sup>2,3</sup>,  
 T. Gessey-Jones<sup>1,2</sup>, G. Kulkarni<sup>10</sup>, A. Magro<sup>11</sup>, P. D. Meerburg<sup>12</sup>,  
 S. Mittal<sup>1,2</sup>, J. H. N. Pattison<sup>1,2</sup>, S. Pegwal<sup>7,13</sup>, C. M. Pieterse<sup>7</sup>,  
 J. R. Pritchard<sup>14</sup>, E. Puchwein<sup>15</sup>, N. Razavi-Ghods<sup>1</sup>,  
 I. L. V. Roque<sup>1</sup>, A. Saxena<sup>12</sup>, K. H. Scheutwinkel<sup>1,2</sup>, P. Scott<sup>1</sup>,  
 E. Shen<sup>1,2</sup>, P. H. Sims<sup>16</sup>, M. Spinelli<sup>17,18</sup>

<sup>1</sup>Astrophysics Group, Cavendish Laboratory, University of Cambridge,  
 J. J. Thomson Avenue, Cambridge, CB3 0HE, UK.

<sup>2</sup>Kavli Institute for Cosmology in Cambridge, University of Cambridge,  
 Madingley Road, Cambridge, CB3 0HA, UK.

<sup>3</sup>Institute of Astronomy, University of Cambridge, Madingley Road,  
 Cambridge, CB3 0HA, UK.

<sup>4</sup>National Astronomical Observatory, Chinese Academy of Science,  
 Beijing, 100101, China.

<sup>5</sup>INAF-Istituto di Radio Astronomia, Via Gobetti 101, Bologna, 40129,  
 Italy.

<sup>6</sup>Laboratoire AstroParticule et Cosmologie, Université Paris-Cité, 10  
 Rue Alice Domon et Léonie Duquet, Paris, 75013, France.

<sup>7</sup>Department of Electrical and Electronic Engineering, Stellenbosch  
 University, Stellenbosch, 7602, South Africa.

<sup>8</sup>Antenna Group, Université catholique de Louvain, Louvain-la-Neuve,  
 1348, Belgium.

<sup>9</sup>Physics Department, University of Oxford, Parks Road, Oxford, OX1  
 3PU, UK.

<sup>10</sup>Department of Theoretical Physics, Tata Institute of Fundamental  
 Research, Homi Bhabha Road, Mumbai, 400005, India.

<sup>11</sup>Institute of Space Sciences and Astronomy, University of Malta,  
 Msida, Malta, MSD 2080, Malta.

<sup>12</sup>Faculty of Science and Engineering, University of Groningen,

Nijenborgh 4, Groningen, 9747 AG, Netherlands.

<sup>13</sup>South African Radio Astronomy Observatory, Black River Park, 2 Fir Street, Observatory, Cape Town, 7925, South Africa.

<sup>14</sup>Department of Physics, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK.

<sup>15</sup>Leibniz Institute for Astrophysics, An der Sternwarte 16, Potsdam, 14482, Germany.

<sup>16</sup>Department of Physics, Arizona State University, 781 South Terrace Rd, Tempe, 6004, US.

<sup>17</sup>Observatoire de la Côte d’Azur, Nice, France.

<sup>18</sup>Department of Physics and Astronomy, University of the Western Cape, Robert Sobukhwe Road, Bellville, 7535, South Africa.

<sup>19</sup>University of Chinese Academy of Sciences, Beijing, 100049, China.

\*Corresponding author(s). E-mail(s): [sakl2@cam.ac.uk](mailto:sakl2@cam.ac.uk);

### Abstract

Radiometers are crucial instruments in radio astronomy, forming the primary component of nearly all radio telescopes. They measure the intensity of electromagnetic radiation, converting this radiation into electrical signals. A radiometer’s primary components are an antenna and a Low Noise Amplifier (LNA), which is the core of the “receiver” chain. Instrumental effects introduced by the receiver are typically corrected or removed during calibration. However, impedance mismatches between the antenna and receiver can introduce unwanted signal reflections and distortions. Traditional calibration methods, such as Dicke switching, alternate the receiver input between the antenna and a well-characterised reference source to mitigate errors by comparison. Recent advances in Machine Learning (ML) offer promising alternatives. Neural networks, which are trained using known signal sources, provide a powerful means to model and calibrate complex systems where traditional analytical approaches struggle. These methods are especially relevant for detecting the faint sky-averaged 21-cm signal from atomic hydrogen at high redshifts. This is one of the main challenges in observational Cosmology today. Here, for the first time, we introduce and test a machine learning-based calibration framework capable of achieving the precision required for radiometric experiments aiming to detect the 21-cm line.

## 1 Introduction

Radiometers have been integral to the field of radio astronomy since its inception. They measure the intensity of incoming electromagnetic radiation within a specific

frequency band, producing a proportional electrical signal at their output. Typically, a radiometer consists of a radio antenna that captures the radiant energy and an amplifying device such as a Low Noise Amplifier (LNA) which, in combination with several other components described in Section 4.1, is commonly referred to as the “receiver”. The incident electromagnetic radiation induces a voltage across the antenna, which can be interpreted as Johnson-Nyquist noise representative of the sky temperature ( $T_{\text{sky}}$ ) [1]. For cosmological applications, particularly those aiming to detect faint signals like the sky-averaged 21-cm line of neutral hydrogen, this initial voltage signal is too weak to be directly interpreted. The signal must therefore be amplified by the receiver before it can be measured. This amplification modifies the signal non-linearly across frequencies, introducing a multiplicative gain factor,  $g$ , and an additive receiver noise term,  $T_{\text{rec}}$ .

It is typically very difficult to know these two values a priori, which motivates the need for external references or calibration sources to constrain them. The relationship between the measured output power and the input temperature from a source can be modelled as

$$\mathcal{P}_{\text{src}}^{\text{out}} = gM(T_{\text{sky}}^{\text{in}} + T_{\text{rec}}), \quad (1)$$

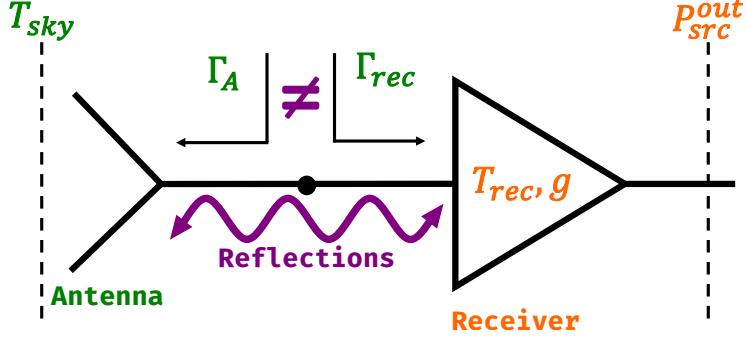
where  $M$  (defined in Section 4) is the mismatch factor. This equation highlights how the measured power depends on both the intrinsic properties of the source and the receiver’s internal noise and impedance mismatch.

Mismatches between the impedances of the antenna and receiver result in partial reflections of both the power incoming from the sky and the noise power generated by the amplifier, shown in Figure 1. These reflections can lead to standing waves and other spectral distortions in the measured output power that complicate the interpretation of the signal. These are typically referred to as “systematics”.

Temporal instabilities in the receiver electronics amplify these challenges, introducing perturbations in the signal that are hard to calibrate. Although integrating time and frequency data can help mitigate random or radiometric noise, eliminating the systematic effects demands accurate calibration. Consequently, the accuracy of the radiometer calibration is crucial for converting the measured amplified signal [2] back to the true emission from the source captured by the antenna.

A particular application of radiometers in astronomy is in the measurement of relative (chromatic or spectral) distortions and of the absolute value of the temperature spectrum from a given celestial source. A classical example of this is the measurement of the Cosmic Microwave Background [3]. The aforementioned calibration challenges are particularly significant when aiming to detect the extremely faint 21-cm cosmological signal from atomic hydrogen at high redshift. This signal is expected to be a few tens or hundreds of milli-Kelvin in amplitude, which is 4–5 orders of magnitude dimmer than the foregrounds that obscure the signal [4]. Radiometers also have multiple uses beyond astronomy and science, being the integral part of many radio receiving systems in fields such as communications [5], medical instrumentation [6] and radar technology [5].

Receiver calibration methods can be subdivided into two primary categories. First, there are methods where the calibration “known” source is external to the antenna,



**Fig. 1:** A schematic overview explaining the core challenge in radiometer calibration. An impedance mismatch between the receiver and the source (usually a radio antenna) causes standing waves between the receiver and source. This adds power to the output signal and must be corrected for via calibration.  $\Gamma_A$  and  $\Gamma_{rec}$  represent the reflection coefficient of the antenna and the receiver and  $P_{src}^{out}$  is the measured power. Throughout the paper, green and orange represent “inputs” and “outputs”, which will be described in detail in Section 4.1.

such as a radio galaxy of known brightness. Second, there are methods where the calibration source is internal and typically connected to the input of the receiver, such as a load resistance. Although the internal approach is less optimal, it is often the only possibility in many applications. For example, when a large field of view prevents dedicated observation of a single external source.

One can also classify receiver calibration methods based on whether they aim to eliminate the effects of the receiver blindly or by fitting a model of the receiver using measured data. Traditional calibration methods used in radio astronomy include techniques such as “Dicke-switching” [7]. In Dicke-switching, the receiver input is continuously switched between the antenna and a reference noise source of constant noise power, which helps mitigate gain fluctuations and receiver noise. Another approach involves using a correlator to reduce the impact of uncorrelated random and systematic noise powers. A well-known alternative is the method known as noise or tone injection [8], where a strong calibration signal is added to the main sky signal without the need to switch the receiver from the antenna to a calibration source. For noise or tone injection to work effectively, the injected calibration signal must be of sufficiently high power relative to the ambient noise and the weak astronomical emissions you are trying to measure. If the astronomical signals are extremely faint, the calibration signal may need to be set at a level that is unrealistically high compared to the noise floor. Such high power levels can be impractical to implement and may even interfere with the measurement process. For example, by saturating the receiver or distorting the system’s response.

Recent advances in machine learning offer an attractive alternative for radiometer calibration. In this article, we introduce a technique for both absolute and spectral calibration that uses machine learning to predict the parameters of a receiver model. By connecting a radio receiver to a set of sources with measured physical temperatures

and reflection coefficients, and by measuring the corresponding output power spectral densities, we train a neural network to accurately infer the calibration parameters. This method shows significant promise for radio astronomy, providing advantages over traditional analytical techniques. We demonstrate its effectiveness using both measured and simulated data from REACH, a 21-cm sky-averaged cosmology experiment [9].

## 1.1 Machine learning for radiometer calibration

We introduce a machine learning-based calibration methodology to model the complex, non-linear behaviours inherent in radiometer systems. A high-level overview of this approach is shown in Figure 2, and its advantages compared to traditional methods are summarised in Table 1. Neural networks are particularly adept at capturing higher-order effects, making them suitable for calibrating instruments where analytical descriptions are insufficient. Whilst we acknowledge that fitting a derived analytic model of the system parameters would be preferable, in the absence of a complete model, machine learning offers a practical and effective alternative.

A similar calibration methodology was proposed in [10] and [11]. However, these methods were designed for systems observing point sources, which can be calibrated by exposing the antenna to previously characterised sources. This is often not possible, for example in sky-averaged experiments where antennas have wide beams and are non-steerable. Therefore, new methods are required.

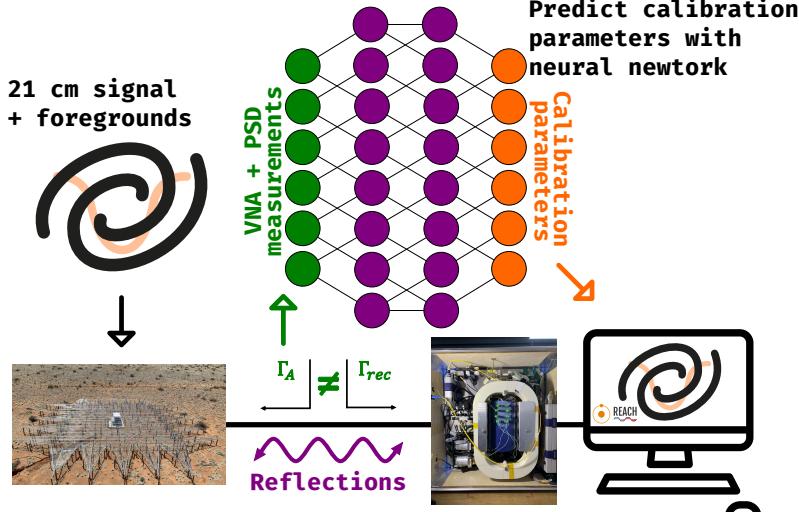
By training the neural network on a diverse set of internal reference sources that encompass the impedance properties of the antenna, we can accurately predict the calibration parameters.

We implement this method on the receiver used by the REACH 21-cm radio telescope. The neural network trains on calibration data gathered live in-field whilst the instrument physically switches between internal reference sources, which generates a comprehensive training set. Once the network has converged on the calibration parameters, it then runs on test data gathered as the instrument physically switches to the antenna (or any other previously unseen source, such as those used for demonstration purposes in this work) on which the neural network can make predictions. This strategy effectively learns the complex spectral properties of these receiver calibration parameters, such as the Noise Parameters, a description of which is given in Section 4.1.

## 1.2 Sky-averaged 21 cm Cosmology

Sky-averaged 21-cm cosmologists are attempting to build the world's most sensitive thermometer to measure the temperature of the early Universe. The redshifted 21-cm line of neutral hydrogen is a faint radio signal that carries the imprint of the first stars and galaxies formed during the Cosmic Dawn and Epoch of Reionisation [12]. By mapping these signals from redshift  $z \approx 6$  to 300 (corresponding to observation frequencies on Earth from 200 MHz down to  $\sim 5$  MHz), it is possible to probe the Universe's history from shortly after the Big Bang to the emergence of complex structures, enabling inference on fundamental physics and cosmology [20, 21].

In 2018, the EDGES collaboration announced the detection of an absorption profile centred at 78 MHz [22], which was nearly twice as deep as the theoretical models



**Fig. 2:** High-level overview of the machine learning-based calibration framework, applied to a 21-cm cosmology experiment. The 21-cm signal obscured by foregrounds [12] is received by the radio telescope as  $T_{\text{sky}}$ . The signal then passes through an amplifier and other components in the receiver chain. The effect of the receiver chain on the signal is “calibrated” by the neural network, which recovers the antenna signal where the 21-cm signal can be extracted. A detailed description of the calibration parameters is provided in Section 4.1.

predicted [20, 21]. This finding was particularly interesting as it suggested potential inaccuracies in the current physical models of the early universe. However, this detection has faced scepticism within the community [23–25] and was challenged by observations from the SARAS3 telescope [26–29]. One of the foremost challenges in observational cosmology today is to verify this measurement or to identify an alternative, true signal.

Despite significant efforts since the potential SARAS3 non-detection in 2022, there have been no new detection. Neither the SARAS3 non-detection nor the EDGES result has been repeated, even though EDGES, SARAS, and now REACH [9] are all commissioning new advanced systems and observing in various locations around the globe. The challenge in observational, sky-averaged 21-cm cosmology lies in the signal’s extreme faintness, which is predicted to be five orders of magnitude dimmer than the bright radio emission from our own and other galaxies [12]. Therefore, to achieve a confident detection, the radio antenna and receiver must be modelled and calibrated to an unprecedented level of precision and accuracy.

The precision of sky-averaged experiments critically depends on precise radiometer calibration [30]. This sensitivity requirement pushes, and in some cases breaks, the limits of current radiometer calibration technology and methodologies. Sky-averaged radiometer calibration is particularly challenging because the antennas used measure the average sky signal. Consequently, they observe the entire sky at all times and are

**Table 1:** Key advantages of the machine learning-based calibration framework for radiometry, compared against traditional methods. Each advantage is discussed in detail within the indicated sections of this document and supporting references.

Advantage	Explanation and Relevant Sections
No impedance mismatch assumptions	Machine learning directly accounts for and learns from measured reflections without requiring ideal matching (see Supplementary Information Section 5 for a discussion of mismatch assumptions in classical noise wave approaches [13, 14]). This avoids systematic offsets that can arise if these assumptions are violated.
Three-dimensional calibration in frequency, temperature and time domains	The model can predict noise parameter <i>surfaces</i> , adding in the temporal axis and thus can naturally handle daily drifts and environmental changes (demonstrated in Supplementary Information B, Section 6). It is thus well-suited for next-generation space-based 21-cm missions [15–17], where hardware and environmental conditions can shift significantly, and analytical calibration assumptions are often insufficient. The method operates on individual frequency channels rather than relying on frequency-polynomial fits employed in some other commonly used methods [18, 19], preserving finer spectral features. See an example of this in Figure 3.
Frequency-by-frequency calibration	The method also recovers the true system gain $g$ (see Section 2), highlighted in Figure 3, which is lost in traditional Dicke-switching-based approaches (derived in Supplementary Information A, Section 5).
Direct gain recovery	The method also recovers the true system gain $g$ (see Section 2), highlighted in Figure 3, which is lost in traditional Dicke-switching-based approaches (derived in Supplementary Information A, Section 5).
Provides absolute calibration	The machine learning framework infers the full set of receiver Noise Parameters (Section 4.1) and maps the measured power spectra onto absolute source temperatures. This enables robust offset recovery for ultra-precise applications whilst reducing dependencies on pre-calibrated reference sources. See Figure 4.

non-steerable. Resultantly, they cannot specifically target well-characterised external sources for calibration, as is standard procedure for other cosmological and astrophysical probes. Therefore, the calibration must be performed using internal reference sources connected to the input of the receiver.

A key challenge lies in the separation of the antenna signal from instrumental effects caused by impedance mismatches between the antenna and internal components [31]. These mismatches lead to reflections that cause standing waves between the antenna and the LNA [32]. This results in an interfering signal with complex spectral properties that obscure the true antenna signal as it passes through the receiver. Traditional calibration methods, such as using impedance-matched references and/or employing Dicke-switching [7], rely on assumptions (detailed in Supplementary Information A, Section 5) that may break down at the precision levels required to measure the temperature of the early Universe. These techniques often presume that certain components are perfectly impedance-matched and that the system remains stable over time. Whilst these assumptions are generally valid, they can potentially lead to errors [30] in the high precision regime required for sky-averaged 21-cm cosmology. Moreover, physical

components like switches, cables, and connectors introduce additional complexities, including extra reflections and standing waves that are challenging to model analytically [30]. These factors can impact the radiometer’s performance, necessitating a calibration approach that can handle such intricacies without depending on potentially incomplete analytical models. As scientists seek fainter signals in the presence of brighter foregrounds, such as the Dark Ages signal [4, 33, 34], and deploy instruments in more demanding environments such as moon surface and orbiting experiments (CosmoCube [16], LuSEE Night [17], LCRT [35] and others [15, 36]), these calibration challenges are expected to become more pronounced. Other challenges in 21-cm modelling include precise antenna beam modelling [37–40]. Furthermore, machine learning methods have been employed in tasks such as forecasting [41] and signal modelling [42–45]. Precise calibration, as discussed in this work, is essential to enable these sophisticated modelling and mitigation techniques.

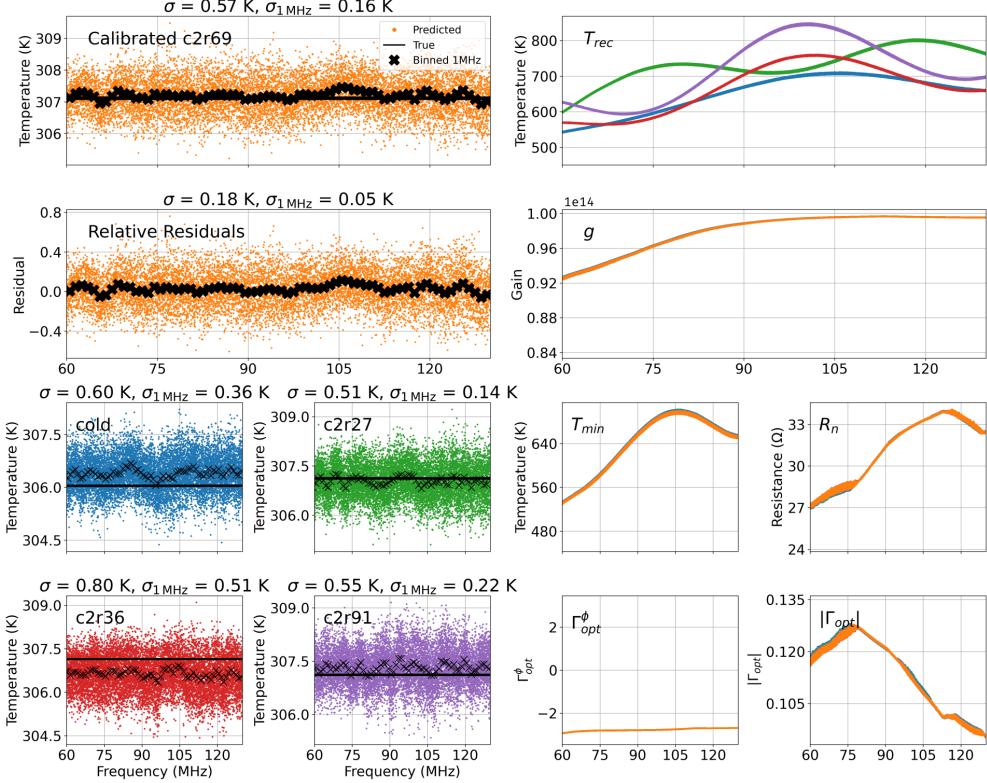
## 2 Results

Our evaluation consists of two tests. First, in Section 2.1, we calibrate the REACH receiver on an internal load with similar properties to the antenna. Second, in Section 2.2, we simulate a full receiver chain (incorporating chromatic antenna beam patterns, realistic foregrounds, and an injected sky-averaged 21-cm signal) to assess the overall performance of the method on a representative science use case.

### 2.1 Calibrating the REACH receiver

We calibrate the REACH receiver [30] and use it to measure the noise temperature of a known source across frequency (a  $69\ \Omega$  resistor on a 2-meter cable) which has similar impedance properties to the REACH dipole antenna. We then compare this result to a direct thermocouple temperature measurement of the resistor to validate the performance of the calibration. We achieve a root-mean-square error (RMSE) on the relative residuals of  $0.05\ K$  at  $1\text{MHz}$  channel width within the frequency band of  $60\text{--}130\text{MHz}$ . Relative residuals are defined as the RMSE error between the predicted and measured temperature values, with the absolute offset accounted for. Calibrating to this level of accuracy and precision on a real antenna would be sufficient to verify or rule out the EDGES signal (mentioned in Section 1) and to detect most possible 21-cm signal models [46]. This would require the subsequent step of calibrating the chromatic effects introduced by the antenna beam, which is independent of the receiver and is therefore beyond the scope of this paper. The methodology predicts the Noise Parameters on individual frequency channels, meaning that it is sensitive to detailed spectral properties at the resolution of the channel width. Some methods commonly used in the field involve fitting polynomials [14, 18, 19] to the model parameters, which can smooth out and obscure fine spectral features.

Once neural network is trained using the default parameters defined in Section 4.2.5, we show that the REACH receiver can be calibrated to the precision required for subsequent antenna calibration and 21-cm signal detection. Figure 3 shows the calibration results for the REACH receiver obtained in the laboratory from a single observation. In this observation, two separate measurements (each with a 600-second



**Fig. 3:** Temperature calibration using internal sources on the REACH receiver. The four panels in the bottom-left quadrant show the measured resistor temperatures (black lines) and the calibrated training solutions (coloured dots). The blue, green, red, and purple dots represent the training data. RMSE errors are shown in the titles. The four panels in the bottom-right quadrant show the predicted Noise Parameters for each source. The top-right two panels show the predicted gain  $g$  (source-independent) and receiver temperature  $T_{rec}$  (source-dependent) for each source, calculated from the predicted Noise Parameters. The two panels in the top-left quadrant show the predictions for the unseen source (a  $69\Omega$  resistor on a 2-meter cable). The “relative residuals” represent the error when the data are centred on zero along the temperature axis. All other plots are “absolute” (i.e., no offset correction applied). The colours of the predicted resistor temperatures on the left correspond to the plots on the right.

integration time) were taken for each calibration source, and the neural network is trained on both simultaneously. Once training is complete, the network predicts the receiver parameters for an unseen test source. During training, whilst the receiver is switched to the internal calibration sources, high-precision temperature probes measure the true temperature of each source. These measurements are then compared to the predicted noise temperatures, and the resulting loss function is minimised leading to accurate Noise Parameter predictions.

The four panels in the bottom-left of Figure 3, display the training data used to train the neural network. In these panels, the flat black lines represent the measured temperatures of the reference sources. This data, together with measurements of the output power and reflection coefficients, forms the complete training set. The coloured points indicate the network’s predictions for each source after training. At the end of training, the calibration sources were calibrated down to 0.2 K, 0.28 K, 0.67 K, and 0.112 K for the cold, c2r27, c2r36, and c2r91 resistors, respectively (where  $c(n)$  represents cable length in meters and  $r(n)$  represents load resistance in Ohms); note that the “cold” load corresponds to a c0r50 source.

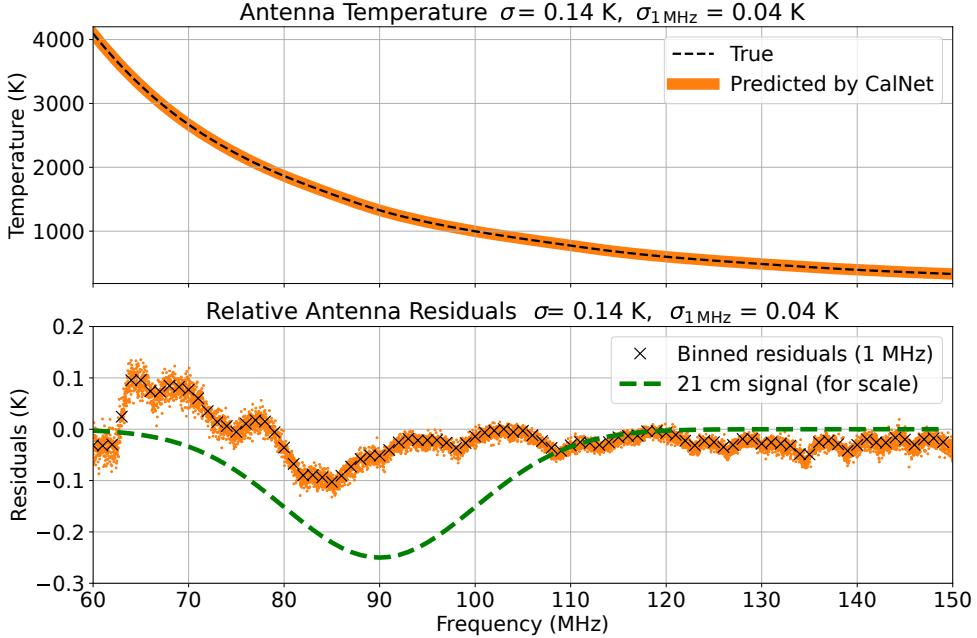
The panel in the top-left of Figure 3 shows the calibration results for an unseen source (in this case, the c2r69 resistor). Here, the orange points represent the temperature predictions made by the neural network on this test data. The RMSE of the residual error on the normalized noise temperature is 0.05 K when evaluated over 1-MHz channels (represented by the black crosses) within the frequency band of 60–130 MHz. For reference, plausible 21-cm signals are expected to exceed 100 mK in amplitude, and may reach up to 500 mK (the approximate amplitude of the EDGES detection). An absolute calibration RMSE of 0.16 K would be sufficient to either confirm or rule out the EDGES detection, place strong constraints on the presence of an excess radio background and to detect many theoretical 21-cm models.

## 2.2 Full chain simulation demonstrating sky-averaged 21-cm signal recovery

To further evaluate the method’s effectiveness once the receiver is connected to the REACH antenna, we simulate a full receiver chain, chromatic antenna beam pattern, realistic foregrounds, and inject a sky-averaged 21-cm signal with properties motivated by current best 21-cm models and constraints to assess whether the method yields residuals sufficiently low to recover a global 21-cm signal under realistic observational conditions. Further details on the simulation methodology are given in Section 4. We show that this signal can be recovered from a realistic simulated system using our machine learning-based framework alongside other current state-of-the-art models in the field.

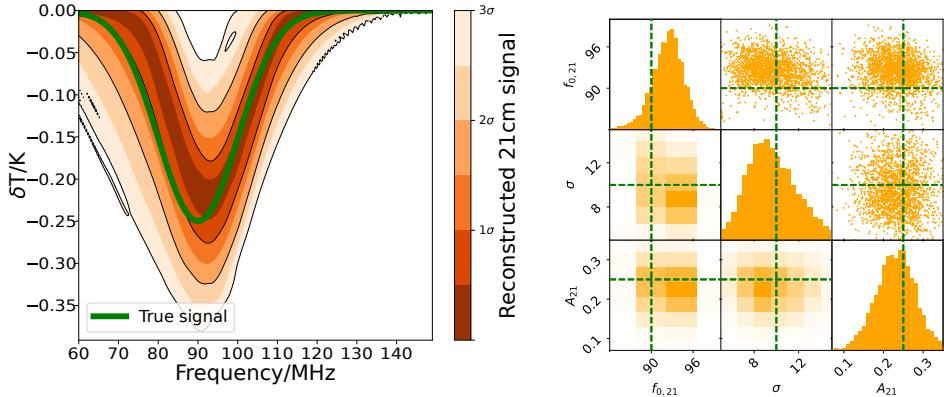
The simulated antenna is calibrated to an RMSE of 0.14 K at a 12-kHz channel width when observed for 100 hours by the simulated receiver chain. When averaged into 1-MHz resolution channels, the signal is recovered with an RMSE of 0.04 K. We bin down to 1 MHz because it allows us to gain sensitivity whilst ensuring that our channel width is much narrower than the spectral features expected from 21-cm signal models. Although the observation time is conservatively large, a shorter observation period would still suffice to recover a 21-cm signal of this magnitude.

Figure 4 shows that calibrating to this residual noise level should, in principle, be sufficient to extract a sky-averaged 21cm signal. Notably, the absolute calibration error and the relative calibration error are the same, indicating that the calibration successfully recovered both the spectral structure and the offset. In contrast, traditional relative calibration methods recover only the spectral structure and require an additional correction for the offset at a later stage.



**Fig. 4:** The top panel shows the predicted antenna temperature in orange, with the true temperature overlaid in black dashes. The true signal is the simulated signal, including the foregrounds, beam effects, and simulated 21-cm signal. The predicted temperature is that recovered by the machine learning calibration methodology after the simulated temperatures have passed through the receiver simulation pipeline. The bottom panel shows the relative calibration residuals when the predicted temperatures are subtracted from the true temperatures. The sky-averaged 21-cm signal injected into the simulated antenna temperature is shown in green for scale. The black crosses represent the orange points (originally at a resolution of 12 kHz) binned to 1-MHz channels.  $\sigma$  and  $\sigma_{1\text{MHz}}$  are the RMSE errors when the orange points are subtracted from the true temperature at 12 kHz and 1 MHz, respectively.

The calibrated antenna temperature is then passed through the REACH data analysis pipeline, which models the chromatic antenna beam pattern and foregrounds using the methods described in [37, 47]. The functional posterior recovered is shown in Figure 5 and the distributions over the predicted model parameters are shown in Figure 5. We use the REACH data analysis pipeline for this simulation because it is an advanced forward model that includes beam, chromaticity, horizon, moon and ionosphere modelling [47–49], and Bayesian anomaly mitigation [50, 51]. Using this, the injected sky-averaged 21-cm signal parameters are recovered as shown in Figure 5.



**Fig. 5:** Inferred sky-averaged 21-cm signal from the calibrated antenna temperature shown in Figure 4. Left: The reconstructed sky-averaged 21-cm signal, showing the inferred signal’s functional posterior in orange and the initially injected true signal in green. Right: Posterior distributions for the parameters of the recovered sky-averaged 21-cm signal, with the true values indicated in green.

### 3 Discussion

In this work, we introduce a novel machine learning-based calibration framework for precision radiometry. Furthermore, we demonstrate its capabilities for application in 21-cm radio cosmology experiments aiming to measure the temperature of the early Universe through the detection of the sky-averaged 21-cm signal. We train a neural network to model the complex, non-linear system behaviour of a precision radiometer. This method, an alternative to traditional calibration approaches, addresses some of the challenges faced by those methods, such as reliance on ideal impedance matching assumptions, limited ability to capture non-linear system behaviour, and difficulties in simultaneously recovering both spectral structure and absolute offsets.

We apply the method to data from the receiver of the REACH radio telescope, and we successfully calibrate most instrumental systematics and reduce the instrumental noise to the noise floor on an internal source chosen for its impedance properties similar to the antenna. On this system, we achieve a relative RMSE of 0.05 K when the predicted temperatures are compared to the measured source temperature. This level of precision, if achieved on the antenna, would be sufficient to verify or rule out the EDGES signal (amplitude  $0.5^{+0.42}_{-0.18}$  K [22]) and to detect most theoretical models [46] of the sky-averaged 21-cm signal.

Notably, this methodology operates on a frequency-by-frequency basis, making it sensitive to detailed spectral properties that are sometimes missed by methods relying on fitted polynomials [14, 18]. Furthermore, the method is effective on live, in-situ calibration data and does not require any laboratory measurements. This is critical for highly precise radiometry, where system drifts (demonstrated in Supplementary

Information B, Section 6) could make laboratory measurements taken prior to instrument deployment unreliable. The method also predicts receiver parameters directly, rather than using methods such as Dicke-switching [7], which can lose important gain information and rely on impedance matching assumptions (outlined in Supplementary Information A, Section 5) that may not hold at the precision and sensitivity levels required for sky-averaged 21-cm cosmology. It should also be noted that the REACH receiver used in this experiment is more complex than just the LNA, incorporating a second stage of amplification, filtering, an RF-to-optical link, and a high-precision spectrometer. In practice, REACH plans to combine this type of machine learning calibration with other more traditional methods, capable of providing better insight into the physics behind the calibration solution.

Our initial efforts focus on calibrating the REACH receiver using internal reference sources, to ensure the system is ready for precise and accurate antenna measurements upon completion of commissioning. Through extensive and highly detailed simulations incorporating realistic models of the antenna [52], receiver chain [53], chromatic beam patterns [37, 54], and foregrounds [55], we have also demonstrated that our machine learning-based calibration framework can calibrate receiver instrumental systematics sufficiently well that the sky-averaged 21-cm signal could be recovered from a simulated realistic antenna. These simulations include an injected sky-averaged 21-cm signal model based on current best 21-cm models and constraints [46].

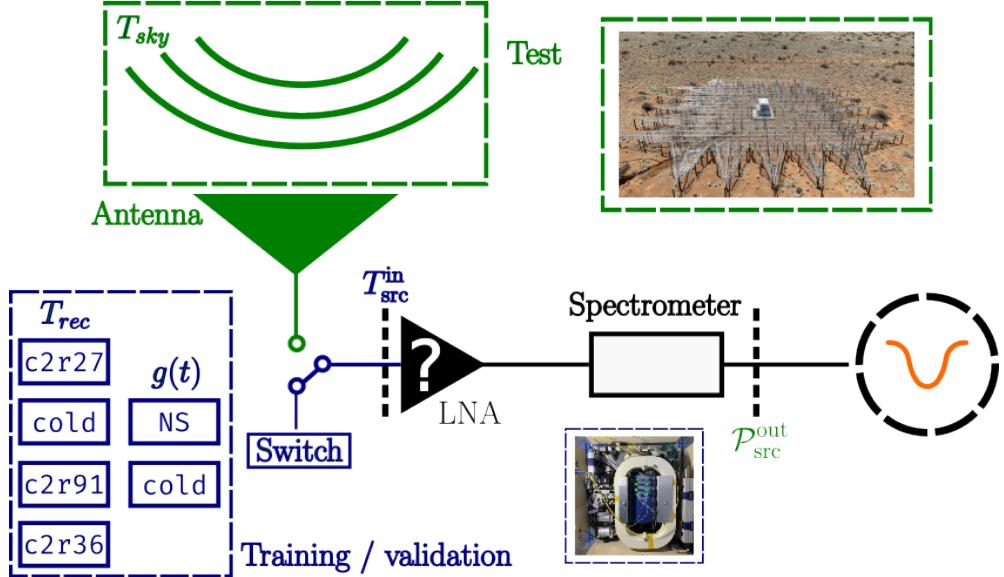
Furthermore, the machine learning-based calibration framework not only addresses current challenges in radiometer calibration for sky-averaged 21-cm cosmology but also opens new avenues for future research and instrumentation. By accurately modelling complex, non-linear system behaviours without relying on simplifying assumptions, this approach can be adapted to other instruments and observational setups. This adaptability is particularly significant as the field moves toward more ambitious projects, such as deploying radiometers in space [15–17] or on the lunar surface [35, 36, 56]. There may also be applications beyond cosmology, such as airborne radiometry (where alternative machine learning-based methods have been used for radiometer calibration [11, 57]) and other areas of astrophysics requiring highly sensitive measurements. Future work will focus on incorporating larger and more diverse datasets, particularly large sets of observations spanning many nights that can be fitted simultaneously.

## 4 Methods

In this section, we explain our machine learning-based radiometer calibration methodology in detail.

### 4.1 Radiometer calibration

Radiometer calibration is the process of mapping measured power spectra onto input source temperatures. The goal of “calibration” is thus to build a model that links the input temperature  $T_{\text{src}}^{\text{in}}$  to the output power  $\mathcal{P}_{\text{src}}^{\text{out}}$ . Typically, radiometers are calibrated by exposing the antenna to a known source, such as an internal black-body [58, 59] or a well-characterised external source [60, 61]. However, the latter is not possible



**Fig. 6:** A physical system is built to generate data that represents the impedance properties of the antenna. The system generates training data from the calibration sources then switches to the antenna to generate test data where predictions can be made. In Section 2.1, a characterised internal source is used in place of the antenna and in Section 2.2 the antenna is simulated.

for sky-averaged experiments as the antenna beam covers the whole sky and is not directive.

For sky-averaged experiments, the calibration methodology proposed by Dicke [7] is commonly used to calibrate the non-linear, time-dependent gain of the LNA. This technique involves comparing the source temperature ( $T_{\text{src}}^{\text{in}}$ ) with internal reference loads whose properties are known: a noise source ( $T_{\text{NS}}^{\text{in}}$ ) and a load ( $T_{\text{L}}^{\text{in}}$ ). We use resistors because they emit well-understood Johnson-Nyquist noise, which is proportional to physical temperature, so their true power output can be inferred from a temperature measurement. It should be noted that these methods either assume that the reference loads are impedance-matched to each other, which is difficult to achieve in reality, or apply a first-order correction [14, 18, 19]. Although this assumption and subsequent first-order correction are generally reasonable, they can potentially lead to non-negligible errors (see Supplementary Information A, Section 5) at the fine scales required for sky-averaged 21-cm cosmology. Other methods assume “ideal” open and short cables [62, 63], which, although also generally reasonable, may lead to errors in this high-precision regime.

In addition to the non-linear gain  $g$  introduced by the amplifier, an impedance mismatch between the source of interest and the LNA modulates the amplifier’s intrinsic

noise, resulting in standing waves along the cable that contribute an additional temperature component,  $T_{\text{rec}}$ , to the output signal. Therefore, as shown in Equation 1,  $T_{\text{src}}^{\text{in}}$  can be related to  $\mathcal{P}_{\text{src}}^{\text{out}}(\nu)$ , where we also define the mismatch factor:

$$M = \frac{(1 - |\Gamma_{\text{src}}|^2)(1 - |\Gamma_{\text{rec}}|^2)}{|1 - \Gamma_{\text{src}}\Gamma_{\text{rec}}|^2}, \quad (2)$$

which preserves the source independence of the gain  $g$  [32].

The Noise Parameter formalism introduced by [64] accounts for the effect of these standing waves, so that Equation 1 becomes:

$$\mathcal{P}_{\text{src}}^{\text{out}} = gM \left( T^{\text{in}}_{\text{src}} + T_{\text{min}} + T_0 \frac{4R_N}{Z_0} \frac{|\Gamma_{\text{src}} - \Gamma_{\text{opt}}|^2}{(1 - |\Gamma_{\text{src}}|^2)|1 + \Gamma_{\text{opt}}|^2} \right), \quad (3)$$

where  $\Gamma_{\text{rec}}$  and  $\Gamma_{\text{src}}$  are the measurable reflection coefficients of the receiver and source, respectively, as measured by a Vector Network Analyser (VNA), and  $Z_0$  and  $T_0$  are measurable properties of the LNA. It is important to note that these two values are constants and are degenerate with  $R_N$ . Whilst their true values are measured to ensure unit consistency, they could be set to 1, and the fit would still yield equivalent results. The remaining parameters— $|\Gamma_{\text{opt}}|$ ,  $\Gamma_{\text{opt}}^\phi$ ,  $R_N$ , and  $T_{\text{min}}$ —must be predicted to characterize  $T_{\text{rec}}$ . Uncalibrated values of  $\mathcal{P}_{\text{src}}^{\text{out}}$  are measured by a spectrometer, which determines the intensity of electromagnetic radiation at various wavelengths.

The Noise Parameter formalism is proposed in the context of sky-averaged 21-cm cosmology in [63]. An alternative approach, used more commonly in this context, represents the system using Noise Wave parameters, as introduced in [65], more recently applied to astronomy in [13], and applied directly to sky-averaged 21-cm cosmology in [14, 18, 22]. The Noise Wave and Noise Parameter formalisms are mathematically equivalent and both (assuming matching assumptions are correct or lead to negligible error) leave four unknown source independent parameters to be computed. For a more detailed explanation see Supplementary Information A, Section 5 and we recommend the aforementioned literature.

## 4.2 Machine learning

A neural network is composed of a series of interconnected layers of nodes. Nodes represent mathematical functions called activation functions that act on signals passing through them. Nodes are connected by weights that vary the amplitude of the signal as it passes through. Given specific inputs (data), the weights can be adjusted so that the outputs become more desirable relative to the inputs by differentiating a goodness-of-fit measure, known as the “loss,” with respect to the weights. By adjusting the weights to minimise the loss, neural networks can make accurate predictions based on complex non-linear relationships between inputs and outputs. A neural network as described above is known as a multi-layer perceptron.

At its core, radiometer calibration requires mapping output powers  $\mathcal{P}_{\text{ant}}^{\text{out}}$  onto input temperatures of interest,  $T_{\text{ant}}^{\text{in}}$ . We characterise this relationship using a neural network trained to predict Noise Parameters based on thermocouple data, power spectral density (PSD) measurements, and Vector Network Analyser (VNA) measurements. Machine learning datasets are typically divided into three sets: *training*, *validation*

and *test*. The neural network regresses over the training data, cross-validates on the validation data, and the test data remains unseen to provide an unbiased assessment. These sets are normally generated by simply splitting a broader dataset into three according to some ratio.

#### 4.2.1 Training data

The neural network trains on carefully designed and selected internal sources and then tests on data collected when the system physically switches to the target source, allowing it to make predictions, as shown in Figure 6. Since it is not possible to build a single source representing the antenna’s complex impedance, the system switches between a series of electrical components, each with different impedance properties that sample the overall region of the Smith chart occupied by the antenna [30]. A Smith chart is a graphical tool used to represent complex impedance and reflection coefficients. If the calibration sources are representative of the antenna, the neural network should be able to learn how to predict the Noise Parameters for the calibrators and subsequently predict them for the antenna.

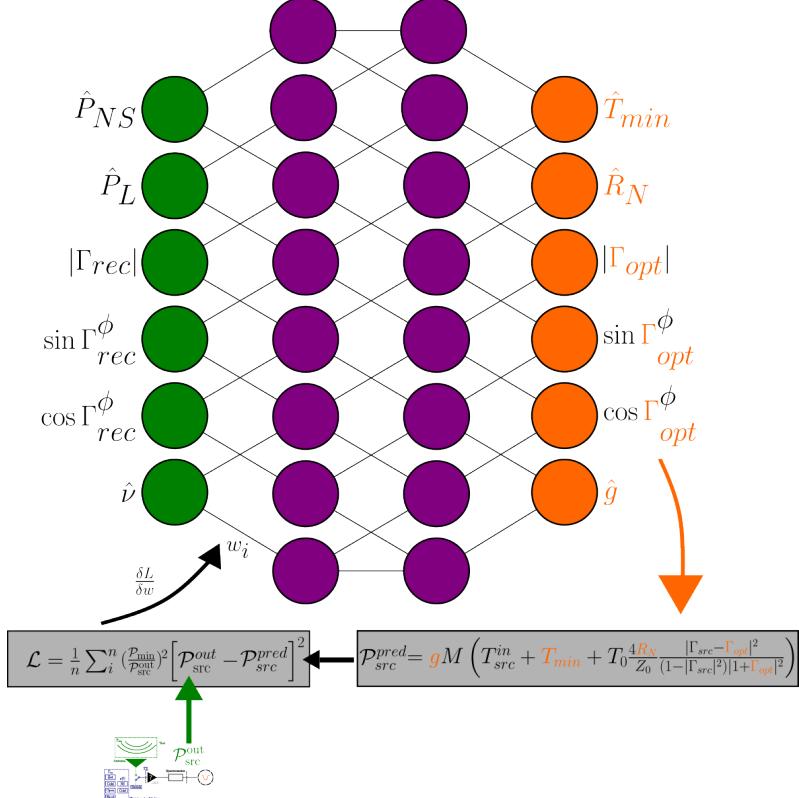
#### 4.2.2 Architecture

We aim to train a neural network to characterise the receiver such that, based on thermocouple data and VNA measurements, the true antenna temperature can be accurately predicted on a frequency-by-frequency basis. This was done in [10] by training on and predicting antenna temperatures directly. However, this is not possible using current sky-averaged 21-cm cosmology instrumentation, as the sky temperature reaches up to 5000 K in the frequency band of interest, and current instruments are not built with sources (or training data) at these temperatures. This is due to the practical constraints of heating a resistor to 5000 K. Furthermore, unlike in [66], sky-averaged antennas are not steerable and cannot target known sources for training. Neural networks are built to interpolate, not extrapolate. Therefore, we must design our architecture such that the network predicts and trains on data that are informative and cover a similar range for both the training data (the calibration sources) and the test data (the antenna). Conveniently, the source-independent Noise Parameter formalism described in Section 4.1 provides us with this.

Therefore, the neural network must be designed to predict the Noise Parameters given thermocouple data, PSD data, and VNA measurements. The Noise Parameters predicted by the neural network can then be used in a single analytical step to predict  $T_{\text{src}}^{\text{in}}$ . We use a multi-layer perceptron, with the number of layers and neurons to be determined by a hyper-parameter optimisation methodology described in Section 4.2.5.

#### 4.2.3 Loss function

Given the network architecture described in Section 4.2.2, a loss function must be defined that quantifies the “goodness-of-fit” of the predicted Noise Parameters. The true values of the Noise Parameters are not known so cannot be regressed over directly. However, the true temperature of the calibration sources can be measured. Equation 3 links output power to input temperature given Noise Parameters and other measurable



**Fig. 7:** Depicts network architectures, showing the flow from predicted Noise Parameters, to output power calculation, to loss, to weights optimisation. The inputs from the PSD measurements and VNA measurements are shown in green and the outputs in orange. See Figure 6 for an enlarged view of the instrument, which outputs  $\mathcal{P}_{src}^{out}$ . Values with a are normalised.

properties. As such, we can predict the Noise Parameters and use them alongside other measured parameters to regress over the mean squared error between the measured power (acting as ground truth) and the predicted power. The predicted power is calculated using the neural network to predict the Noise Parameters, followed by an analytical step using Equation 3 to compute  $\mathcal{P}_{src}^{out}$ . This yields the loss function

$$\mathcal{L} = \frac{1}{n} \sum_i^n \left( \frac{\mathcal{P}_{\min}}{\mathcal{P}_{src}^{out}} \right)^2 \left[ \mathcal{P}_{src}^{out} - g M \left( T_{src}^{in} + T_{min} + T_0 \frac{4R_N}{Z_0} \frac{|\Gamma_{src} - \Gamma_{opt}|^2}{(1 - |\Gamma_{src}|^2)(1 + |\Gamma_{opt}|^2)} \right) \right]^2, \quad (4)$$

where the term  $(\frac{\mathcal{P}_{\min}}{\mathcal{P}_{src}^{out}})^2$  is included to account for the fact that the noise in the system is radiometric, not Gaussian, as is often assumed when using a mean-squared error loss function.  $\mathcal{P}_{\min}$  is the lowest measured power across the training set.

Visualised in Figure 7, this loss function allows the estimation of how close the predicted Noise Parameters are to their “true” values without explicitly knowing these true values. The gradient of the high-dimensional loss landscape with respect to the network weights can be approximated using automatic differentiation (we use TensorFlow). From this, the loss can be adjusted towards its minima. Thus, the true Noise Parameters can be estimated and subsequently  $\mathcal{P}_{\text{src}}^{\text{out}}$  calculated, which can be inverted to yield  $T_{\text{src}}^{\text{in}}$ , which in the context of cosmology will be  $T_{\text{sky}}^{\text{in}}$ .

#### 4.2.4 Data normalisation

Neural networks work most efficiently and optimally on values of order unity, as this prevents the saturation of activation functions. We normalise the input powers and frequencies to within this range using min-max normalisation [67]. Training on complex numbers is slightly more intricate, as the system must be designed to interpret the wrapped nature of such parameters.  $0\pi$  represents the same point in the phase cycle as  $2\pi$ , but in a physical scenario these two positions in phase may have different practical implications. We can encode this into the data directly using phase wrapping. As shown in Figure 7, we achieve this by feeding the network  $|\Gamma_{\text{rec}}|$ ,  $\sin \Gamma_{\text{rec}}^\phi$ , and  $\cos \Gamma_{\text{rec}}^\phi$ .

The system we aim to predict is physical, but many degenerate non-physical solutions form troughs in the loss landscape. This can slow or prevent convergence on the physical solution. To solve this, we encode physics into our network architecture where possible by constraining  $R_N$ ,  $g$ , and  $T_{\min}$  to be positive, scaling them between zero and one based on prior knowledge of their maximum scale, and fixing their output activation to a sigmoid function.  $\Gamma_{\text{opt}}$  is complex, and we predict its absolute value and the unwrapped phase as described for the inputs. Although convergence is achieved with or without this scaling, it can be significantly accelerated by constraining network outputs to tighter ranges based on loose prior knowledge of the Noise Parameters.

#### 4.2.5 Hyper-parameter optimisation

The system hyperparameters are Network structure, learning rate, dropout (proportion of nodes randomly “turned off” during training, for regularisation), and batch size are optimised using a Bayesian optimisation algorithm. We use Hyperopt [68], a Python library, which implements Bayesian optimisation through the Tree-structured Parzen Estimator (TPE) algorithm [69]. This method systematically improves upon candidate solutions based on prior evaluations, utilising a probabilistic model to predict the performance of new hyperparameter combinations. Based on these results, we can select the optimal network structure and hyperparameter combination. Throughout the training process we implement early stopping when the validation loss does not improve for 1000 epochs, with the total number of epochs being set to 5000. An epoch is 1 training loop over the full training set.

Optimal values depend on data size and complexity, so further optimisation may be required for future use. The optimal parameters determined using the REACH receiver data were: batch size of 256, dropout uniformly sampled between 0 and 0.5 on all nodes, learning rate of 0.0005, 60 neurons in layer 1, 100 neurons in layer 2, and 80 neurons in layer 3.

### 4.3 Simulated data

To validate our machine learning-based calibration framework, we assess extensive simulations mimicking the operational conditions of a sky-averaged 21-cm experiment, as detailed in Section 2.2. In this section, we explain how the data for these physical simulations were generated. These simulations incorporated realistic models of the REACH antenna, receiver chain, chromatic beam patterns, and foreground emissions. We utilised the methodologies described in [53] to simulate the receiver chain and those in [37, 47] to generate synthetic datasets closely resembling the data expected from the REACH receiver.

The simulated antenna parameters were derived to be similar to the characteristics of the proposed REACH antenna, including impedance properties and beam patterns [37]. A realistic sky model was incorporated, accounting for foreground emissions from our galaxy and extragalactic sources as outlined in [47]. The receiver chain was modelled using Noise Parameter measurements obtained from laboratory tests of the REACH receiver components [30].

To achieve a comprehensive simulation of the entire calibration process, our workflow proceeded as follows:

1. **Antenna Temperature Simulation:** Beam simulations and realistic foreground models (as detailed in [37, 47, 52, 53]) were used to generate a realistic sky-averaged 21cm antenna temperature.
2. **Receiver Chain Simulation:** The simulated antenna temperature was passed through a comprehensive simulation of the receiver, which incorporates laboratory measurements from the REACH receiver, to produce the corresponding simulated output power.
3. **Reference Load Simulation:** In parallel, output power for simulated reference loads (mimicking the calibration sources) was generated.
4. **Neural Network Training:** The simulated measurements from the reference loads were used to train a neural network to accurately predict the receiver's Noise Parameters.
5. **Antenna Temperature Recovery:** The trained neural network was subsequently applied to recover the antenna temperature from the simulated receiver output.
6. **Signal Extraction:** A model was fitted to the recovered antenna temperature to extract the underlying sky-averaged 21-cm signal, based on the most current constraints [46].

A theoretical model of the sky-averaged 21-cm signal was injected into these simulations to assess the capability of our calibration framework to recover the signal amidst instrumental effects, systematics, and environmental factors that might impact the observations.

### 4.4 REACH receiver data

For the practical implementation of our calibration framework, we used data from the REACH (Radio Experiment for the Analysis of Cosmic Hydrogen) radiometer. The

REACH instrument is designed specifically for precision measurements of the sky-averaged 21-cm signal, featuring a custom-built antenna and receiver system optimised for low-noise observations.

We collected the calibration data by switching the receiver between multiple internal reference sources with known impedance properties, as well as an internal source with properties similar to the antenna and measurable ground truth temperature for evaluation. The system design and Smith chart coverage for this configuration are outlined in [30]. The methodology for data acquisition and instrument configuration follows the procedures detailed in [9, 30]. The internal reference sources are carefully designed to sample the impedance space occupied by the antenna, thereby providing a comprehensive training set for the neural network.

Temperature measurements from thermocouples attached to the calibration sources were recorded to obtain accurate reference temperatures. Additionally, Vector Network Analyser (VNA) measurements were performed to characterise the reflection coefficients of the receiver and sources. These measurements served as crucial inputs to the neural network during the training phase.

## Contributions

S.A.K. Leeney led the research, development, and preparation of the manuscript. H.T.J. Bevins supervised the research and proposed the initial concept for the study. E. de Lera Acedo supervised the project and provided guidance on the fundamentals of radiometer calibration. W.J. Handley supervised the project, offering advice on machine learning and data analysis. C. Kirkham advised on simulated data generation and preprocessing for calibration. R. S. Patel configured and ran the observations on the REACH receiver. J. Zhu developed the initial calibration simulation pipeline. D. Molnar provided foundational guidance on radiometer calibration during the early stages of the project. D. Anstey helped configure the REACH data analysis pipeline for the analysis of simulated data, and J. Cumner contributed to the generation of simulated antenna temperatures. The remaining authors, from K. Artuc to M. Spinelli, contributed to the REACH telescope and are therefore on the REACH builders list. They also contributed to the manuscript's preparation and review. These authors are credited in alphabetical order.



Finally, combining the above expressions produces a linear relationship between the *uncalibrated* temperature and the final, *calibrated* temperature. The reader should note that throughout such treatments, several assumptions are made: (i) low-reflection internal references, (ii) negligible mismatch between the internal references, and (iii) parametric modelling of noise waves via polynomials in frequency ( $T_{\text{unc}}, T_{\text{cos}}, \dots$ ).

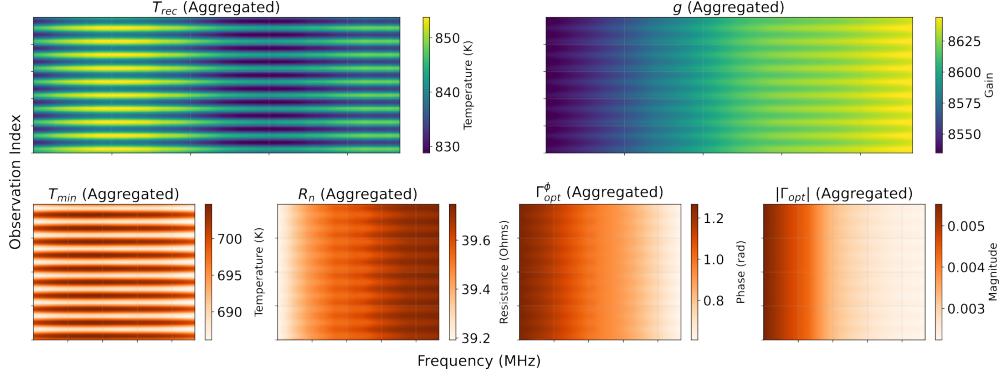
In experiments where the required precision presses well beyond the validity regime of these approximations (as is often the case for sky-averaged 21-cm experiments), systematic errors may be introduced. By contrast, our machine learning approach predicts the necessary parameters (specifically, the Noise Parameters) without *a priori* assumptions of perfect impedance matching or negligible reflections in the internal references. Consequently, our framework avoids introducing potential modelling biases inherent in the traditional scheme, providing a more comprehensive solution for high-precision sky-averaged 21-cm radiometry.

## 6 Supplementary Information B: Training on a Temporally Unstable System

To demonstrate how the method operates on an unstable system, we generated a simulated dataset designed to mimic a temporally unstable system, reflecting the complexities encountered in real-world observational environments. Similar to the process described in Section 2.2 for generating simulated data, we simulated a realistic model of the receiver chain. However, in this instance, we extended the simulation to cover multiple observations, accumulating a total of 1000 observations, each lasting 30 seconds, for every calibration source. This dataset allows us to test our calibration framework in the temporal domain, predicting three-dimensional Noise Parameter *surfaces* (parameter value as a function of frequency and time), rather than just two-dimensional frequency-dependent curves.

A key distinction in this simulation was the introduction of temporal variability into one of the system's Noise Parameters. Specifically, the Noise Parameter  $T_{\min}$  was varied sinusoidally by a fixed offset over the course of the observations, fluctuating by  $\pm 5\%$  around its nominal value. This variation was implemented to simulate the gradual drifts and changes that can occur in the electronic components of a receiver due to environmental factors or ageing, representing a significant challenge for traditional calibration techniques. It is important to note that many conventional calibration methods rely on the assumption that receiver Noise Parameters remain stable over the observation period, with the exception of the overall gain  $g$ . Such temporal instabilities, if not properly accounted for, can introduce systematic errors that corrupt the final scientific measurements.

We trained the neural network using this temporally dynamic dataset. The network successfully learned to characterise the unstable system, demonstrating its ability to adapt to and model time-varying instrumental effects by predicting the full three-dimensional Noise Parameter surfaces. By processing the sequence of observations, the model effectively captured the underlying sinusoidal variation in  $T_{\min}$  within its predicted surface, whilst simultaneously determining the surfaces for the other stable Noise Parameters and the gain. It is worth noting here that there is some degeneracy



**Fig. 8:** Heatmaps illustrating the temporal evolution of the inferred receiver Noise Parameters over the simulated multi-night observation period. Each panel corresponds to one of the four Noise Parameters ( $T_{\min}$ ,  $R_N$ ,  $|\Gamma_{\text{opt}}|$ ,  $\Gamma_{\text{opt}}^\phi$ ) and the gain  $g$ . The y-axis represents observation order (ascending from first to last), and the x-axis represents frequency. Colour intensity indicates the inferred parameter value. The sinusoidal variation deliberately introduced into  $T_{\min}$  (bottom-left panel) is clearly visible and accurately recovered by the neural network, demonstrating the model’s ability to characterise a temporally unstable system. Note that each heatmap contains aggregated data from all calibrators used in the simulation.

between  $T_{\min}$  and the other noise parameters, causing a light ripple to appear in the time axis of the other heatmaps. This does not affect the overall calibration result.

Figure 8 illustrates the network’s performance in tracking the evolving Noise Parameters over the simulated observation period. The heatmaps display the inferred values for each Noise Parameter as a function of time. Notably, the heatmap corresponding to  $T_{\min}$  clearly shows the sinusoidal oscillation in the time axis introduced into the simulation.

This result signifies an important extension of radiometer calibration into the temporal domain, potentially for the first time. Our machine learning framework moves beyond static, two-dimensional system characterisation (parameter vs. frequency) to predict three-dimensional Noise Parameter surfaces (parameter vs. frequency and time), learning how the physical properties of the instrument evolve. This capability to model temporal variations is particularly valuable as radio astronomy pushes towards deploying instruments in increasingly remote and challenging environments, both on Earth and in space (e.g., lunar missions like LuSEE-Night [17] or potential orbital concepts such as CosmoCube [16]).

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Code availability

The code that supports the findings of this study is available from the corresponding author upon reasonable request.

## Competing interests

The authors declare no competing interests.

## References

- [1] Nyquist, H.: Thermal agitation of electric charge in conductors. *Physical review* **32**(1), 110 (1928)
- [2] Wyatt, C.: Radiometric Calibration: Theory and Methods. Elsevier, ??? (2012)
- [3] Durrer, R.: The Cosmic Microwave Background. Cambridge University Press, ??? (2020)
- [4] Pritchard, J.R., Loeb, A.: Constraining the unexplored period between the dark ages and reionization with observationsj? format?\_ of the global 21 cm signal. *Physical Review D—Particles, Fields, Gravitation, and Cosmology* **82**(2), 023006 (2010)
- [5] Battaglia, A., Kollias, P., Rysavy, Y., Roy, R.J.: A w-band radar–radiometer system for accurate and continuous monitoring of clouds and precipitation. *Journal of Atmospheric and Oceanic Technology* **34**(11), 2421–2437 (2017)
- [6] Arunachalam, K., Behdad, N., Gundala, S., Riazi, A., Webster, J.G.: Characterization of a digital microwave radiometry system for noninvasive thermometry using a temperature-controlled homogeneous test load. *Physics in Medicine and Biology* **53**, 883–901 (2008)
- [7] Dicke, R.H.: The measurement of thermal radiation at microwave frequencies. *Review of scientific instruments* **17**(7), 268–275 (1946)
- [8] Corbella, I., Camps, A., Torres, F., Bará, J.: Analysis of noise-injection networks for interferometric-radiometer calibration. *IEEE Transactions on Microwave Theory and Techniques* **48**(4), 545–552 (2000)
- [9] Lera Acedo, E.: Reach: Radio experiment for the analysis of cosmic hydrogen. In: 2019 International Conference on Electromagnetics in Advanced Applications (ICEAA), pp. 0626–0629 (2019). IEEE
- [10] Ogut, M., Bosch-Lluis, X., Reising, S.C.: A deep learning approach for microwave and millimeter-wave radiometer calibration. *IEEE Transactions on Geoscience and Remote Sensing* **57**(8), 5344–5355 (2019)

- [11] Alam, A.M., Kurum, M., Ogut, M., Gurbuz, A.C.: Microwave radiometer calibration using deep learning with reduced reference information and 2-d spectral features. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **17**, 748–765 (2023)
- [12] Pritchard, J.R., Loeb, A.: 21 cm cosmology in the 21st century. *Reports on Progress in Physics* **75**(8), 086901 (2012)
- [13] Rogers, A.E., Bowman, J.D.: Absolute calibration of a wideband antenna and spectrometer for accurate sky noise temperature measurements. *Radio Science* **47**(06), 1–9 (2012)
- [14] Monsalve, R.A., Rogers, A.E., Bowman, J.D., Mozdzen, T.J.: Calibration of the edges high-band receiver to observe the global 21 cm signature from the epoch of reionization. *The Astrophysical Journal* **835**(1), 49 (2017)
- [15] Yamasaki, Y., Matsumoto, T., Iguchi, S., Onishi, T., Yamada, T., Yamauchi, D., Tsuchiya, F., Takahashi, K., Isobe, N., Iwata, T., *et al.*: A 1–50 mhz short dipole antenna for observing the 21 cm global signal on the moon. In: *Space Telescopes and Instrumentation 2024: Optical, Infrared, and Millimeter Wave*, vol. 13092, pp. 2342–2349 (2024). SPIE
- [16] Artuc, K., Acedo, E.d.L.: The spectrometer development of cosmocube, lunar orbiting satellite to detect 21-cm hydrogen signal from cosmic dark ages. arXiv preprint arXiv:2406.10096 (2024)
- [17] Bale, S.D., Bassett, N., Burns, J.O., Jones, J.D., Goetz, K., Hellum-Bye, C., Hermann, S., Hibbard, J., Maksimovic, M., McLean, R., *et al.*: Lusee’night’: The lunar surface electromagnetics experiment. arXiv preprint arXiv:2301.10345 (2023)
- [18] Roque, I., Handley, W., Razavi-Ghods, N.: Bayesian noise wave calibration for 21-cm global experiments. *Monthly Notices of the Royal Astronomical Society* **505**(2), 2638–2646 (2021)
- [19] Murray, S.G., Bowman, J.D., Sims, P.H., Mahesh, N., Rogers, A.E., Monsalve, R.A., Samson, T., Vydula, A.K.: A bayesian calibration framework for edges. *Monthly Notices of the Royal Astronomical Society* **517**(2), 2264–2284 (2022)
- [20] Furlanetto, S.R., Oh, S.P., Briggs, F.H.: Cosmology at low frequencies: The 21 cm transition and the high-redshift universe. *Physics reports* **433**(4-6), 181–301 (2006)
- [21] Barkana, R.: Possible interaction between baryons and dark-matter particles revealed by the first stars. *Nature* **555**(7694), 71–74 (2018)
- [22] Bowman, J.D., Rogers, A.E., Monsalve, R.A., Mozdzen, T.J., Mahesh, N.: An absorption profile centred at 78 megahertz in the sky-averaged spectrum. *Nature*

- 555**(7694), 67–70 (2018)
- [23] Hills, R., Kulkarni, G., Meerburg, P.D., Puchwein, E.: Concerns about modelling of the edges data. *Nature* **564**(7736), 32–34 (2018)
- [24] Sims, P.H., Poher, J.C.: Testing for calibration systematics in the edges low-band data using bayesian model selection. *Monthly Notices of the Royal Astronomical Society* **492**(1), 22–38 (2020)
- [25] Cang, J., Mesinger, A., Murray, S.G., Breitman, D., Qin, Y., Trotta, R.: The edges measurement disfavors an excess radio background during the cosmic dawn. arXiv preprint arXiv:2411.08134 (2024)
- [26] Singh, S., Subrahmanyam, R., Shankar, N.U., Rao, M.S., Fialkov, A., Cohen, A., Barkana, R., Girish, B., Raghunathan, A., Somashekar, R., et al.: Saras 2 constraints on global 21 cm signals from the epoch of reionization. *The Astrophysical Journal* **858**(1), 54 (2018)
- [27] Singh, S., Subrahmanyam, R., Shankar, N.U., Rao, M.S., Girish, B., Raghunathan, A., Somashekar, R., Srivani, K.: Saras 2: a spectral radiometer for probing cosmic dawn and the epoch of reionization through detection of the global 21-cm signal. *Experimental Astronomy* **45**, 269–314 (2018)
- [28] Girish, B., Srivani, K., Subrahmanyam, R., Udaya Shankar, N., Singh, S., Jishnu Nambissan, T., Sathyanarayana Rao, M., Somashekar, R., Raghunathan, A.: Saras cd/eor radiometer: design and performance of the digital correlation spectrometer. *Journal of Astronomical Instrumentation* **9**(02), 2050006 (2020)
- [29] Singh, S., Nambissan T, J., Subrahmanyam, R., Udaya Shankar, N., Girish, B., Raghunathan, A., Somashekar, R., Srivani, K., Sathyanarayana Rao, M.: On the detection of a cosmic dawn signal in the radio background. *Nature Astronomy* **6**(5), 607–617 (2022)
- [30] Razavi-Ghods, N., Roque, I.L., Carey, S.H., Ely, J.A., Handley, W., Magro, A., Chiello, R., Huang, T., Alexander, P., Anstey, D., et al.: Receiver design for the reach global 21-cm signal experiment. arXiv preprint arXiv:2307.00099 (2023)
- [31] Corbella, I., Gasiewski, A.J., Klein, M., Leuski, V., Francavilla, A.J., Piepmeyer, J.R.: On-board accurate calibration of dual-channel radiometers using internal and external references. *IEEE transactions on microwave theory and techniques* **50**(7), 1816–1820 (2002)
- [32] Pozar, D.M.: *Microwave and RF Design of Wireless Systems*. John Wiley & Sons, ??? (2000)
- [33] Mondal, R., Barkana, R.: Prospects for precision cosmology with the 21 cm signal from the dark ages. *Nature Astronomy* **7**(9), 1025–1030 (2023)

- [34] Fialkov, A., Gessey-Jones, T., Dhandha, J.: Cosmic mysteries and the hydrogen 21-cm line: bridging the gap with lunar observations. *Philosophical Transactions of the Royal Society A* **382**(2271), 20230068 (2024)
- [35] Bandyopadhyay, S., McGarey, P., Goel, A., Rafizadeh, R., Delapierre, M., Arya, M., Lazio, J., Goldsmith, P., Chahat, N., Stoica, A., et al.: Conceptual design of the lunar crater radio telescope (lcrt) on the far side of the moon. In: 2021 IEEE Aerospace Conference (50100), pp. 1–25 (2021). IEEE
- [36] Burns, J., Hallinan, G., Chang, T.-C., Anderson, M., Bowman, J., Bradley, R., Furlanetto, S., Hegedus, A., Kasper, J., Kocz, J., et al.: A lunar farside low radio frequency array for dark ages 21-cm cosmology. arXiv preprint arXiv:2103.08623 (2021)
- [37] Anstey, D., Cumner, J., Lera Acedo, E., Handley, W.: Informing antenna design for sky-averaged 21-cm experiments using a simulated bayesian data analysis pipeline. *Monthly Notices of the Royal Astronomical Society* **509**(4), 4679–4693 (2022)
- [38] Hibbard, J.J., Rapetti, D., Burns, J.O., Mahesh, N., Bassett, N.: Fitting and comparing galactic foreground models for unbiased 21 cm cosmology. *The Astrophysical Journal* **959**(2), 103 (2023)
- [39] Mittal, S., Kulkarni, G., Anstey, D., Lera Acedo, E.: Impact of extragalactic point sources on the low-frequency sky spectrum and cosmic dawn global 21-cm measurements. *Monthly Notices of the Royal Astronomical Society* **534**(2), 1317–1328 (2024)
- [40] Sims, P.H., Bowman, J.D., Mahesh, N., Murray, S.G., Barrett, J.P., Cappallo, R., Monsalve, R.A., Rogers, A.E., Samson, T., Vydula, A.K.: A bayesian approach to modelling spectrometer data chromaticity corrected using beam factors–i. mathematical formalism. *Monthly Notices of the Royal Astronomical Society* **521**(3), 3273–3297 (2023)
- [41] Gessey-Jones, T., Handley, W.: Fully bayesian forecasts with evidence networks. *Physical Review D* **109**(12), 123541 (2024)
- [42] Bevins, H., Handley, W., Fialkov, A., Lera Acedo, E., Javid, K.: Globalemu: a novel and robust approach for emulating the sky-averaged 21-cm signal from the cosmic dawn and epoch of reionization. *Monthly Notices of the Royal Astronomical Society* **508**(2), 2923–2936 (2021)
- [43] Shen, E., Anstey, D., Lera Acedo, E., Fialkov, A.: Flexknot as a generalized model of the sky-averaged 21-cm signal at  $z \sim 6$ –30 in the presence of systematics. *Monthly Notices of the Royal Astronomical Society* **529**(2), 1642–1653 (2024)

- [44] Saxena, A., Meerburg, P.D., Weniger, C., Acedo, E.d.L., Handley, W.: Simulation-based inference of the sky-averaged 21-cm signal from cd-eor with reach. *RAS Techniques and Instruments* **3**(1), 724–736 (2024)
- [45] Hibbard, J.J., Nhan, B.D., Rapetti, D., Burns, J.O.: Medea: A new model for emulating radio antenna beam patterns for 21 cm cosmology and antenna design studies. *The Astrophysical Journal* **975**(1), 36 (2024)
- [46] Bevins, H.T., Heimersheim, S., Abril-Cabezas, I., Fialkov, A., Lera Acedo, E., Handley, W., Singh, S., Barkana, R.: Joint analysis constraints on the physics of the first galaxies with low-frequency radio astronomy data. *Monthly Notices of the Royal Astronomical Society* **527**(1), 813–827 (2024)
- [47] Anstey, D., Lera Acedo, E., Handley, W.: A general bayesian framework for foreground modelling and chromaticity correction for global 21 cm experiments. *Monthly Notices of the Royal Astronomical Society* **506**(2), 2041–2058 (2021)
- [48] Pattison, J.H., Anstey, D.J., Lera Acedo, E.: Modelling a hot horizon in global 21-cm experimental foregrounds. *Monthly Notices of the Royal Astronomical Society* **527**(2), 2413–2425 (2024)
- [49] Shen, E., Anstey, D., Spinelli, M., Acedo, E.d.L., Fialkov, A.: Bayesian data analysis for sky-averaged 21-cm experiments with contamination from linearly polarised foreground. arXiv preprint arXiv:2412.02552 (2024)
- [50] Leeney, S., Handley, W., Lera Acedo, E.: Bayesian approach to radio frequency interference mitigation. *Physical Review D* **108**(6), 062006 (2023)
- [51] Anstey, D., Leeney, S.A.: Enhanced bayesian rfi mitigation and transient flagging using likelihood reweighting. *RAS Techniques and Instruments* **3**(1), 372–384 (2024)
- [52] Cumner, J., Lera Acedo, E., Villiers, D.I., Anstey, D., Kolitsidas, C.I., Gurdon, B., Fagnoni, N., Alexander, P., Bernardi, G., Bevins, H.T., *et al.*: Radio antenna design for sky-averaged 21 cm cosmology experiments: the reach case. *Journal of Astronomical Instrumentation* **11**(01), 2250001 (2022)
- [53] Sun, S., Lera Acedo, E., Wu, F., Yue, B., Zhu, J., Chen, X.: Calibration error in 21-cm experiments global spectrum experiment (2024)
- [54] Cumner, J., Pieterse, C., De Villiers, D., Lera Acedo, E.: The effects of the antenna power pattern uncertainty within a global 21 cm experiment. *Monthly Notices of the Royal Astronomical Society* **531**(4), 4734–4745 (2024)
- [55] Anstey, D., Leeney, S.A.: Enhanced bayesian rfi mitigation and transient flagging using likelihood reweighting. arXiv preprint arXiv:2310.02146 (2023)

- [56] Reising, S.C., Kangaslahti, P., Schlecht, E., Jiang, J., Bosch-Lluis, X., Ogut, M., Goncharenko, Y., Padmanabhan, S., Cofield, R., Chahat, N., *et al.*: Tropospheric water and cloud ice (twice) millimeter and submillimeter-wave radiometer instrument for 6u-class nanosatellites. In: 2016 41st International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz), pp. 1–2 (2016). IEEE
- [57] Bosch-Lluis, X., Reising, S.C., Kangaslahti, P., Tanner, A.B., Brown, S.T., Padmanabhan, S., Parashare, C., Montes, O., Razavi, B., Hadel, V.D., *et al.*: Instrument design and performance of the high-frequency airborne microwave and millimeter-wave radiometer. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **12**(11), 4563–4577 (2019)
- [58] Kendall, J., Berdahl, C.M.: Two blackbody radiometers of high accuracy. *Applied Optics* **9**(5), 1082–1091 (1970)
- [59] Sapritsky, V.: Black-body radiometry. *Metrologia* **32**(6), 411 (1995)
- [60] Adam, R., Ade, P., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A., Barreiro, R., Bartolo, N., *et al.*: Planck 2015 results-viii. high frequency instrument data processing: Calibration and maps. *Astronomy & Astrophysics* **594**, 8 (2016)
- [61] Noordam, J.E.: Lofar calibration challenges. In: *Ground-based Telescopes*, vol. 5489, pp. 817–825 (2004). SPIE
- [62] Price, D., Greenhill, L., Fialkov, A., Bernardi, G., Garsden, H., Barsdell, B., Kocz, J., Anderson, M., Bourke, S., Craig, J., *et al.*: Design and characterization of the large-aperture experiment to detect the dark age (leda) radiometer systems. *Monthly Notices of the Royal Astronomical Society* **478**(3), 4193–4213 (2018)
- [63] Price, D.C., Tong, C.-Y.E., Sutinjo, A.T., Greenhill, L.J., Patra, N.: Measuring noise parameters using an open, short, load, and  $\lambda/8$ -length cable as source impedances. *IEEE Transactions on Microwave Theory and Techniques* **71**(3), 1102–1111 (2022)
- [64] Haus, H., Atkinson, W., Branch, G., Davenport, W., Fonger, W., Harris, W., Harrison, S., McLeod, W., Stodola, E., Talpey, T.: Representation of noise in linear twoports. *Proceedings of the IRE* **48**(1), 69–74 (1960)
- [65] Meys, R.: A wave approach to the noise properties of linear microwave devices. *IEEE Transactions on Microwave Theory and Techniques* **26**(1), 34–37 (1978)
- [66] Ogut, M., Bosch-Lluis, X., Reising, S.C.: Deep learning calibration of the high-frequency airborne microwave and millimeter-wave radiometer (hammr) instrument. *IEEE Transactions on Geoscience and Remote Sensing* **58**(5), 3391–3399 (2019)

- [67] Géron, A.: Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow. ” O'Reilly Media, Inc.”, ??? (2022)
- [68] Bergstra, J., Komorowski, B., Eliasmith, C., Yamins, D., Cox, D.D.: Hyperopt: a python library for model selection and hyperparameter optimization. Computational Science & Discovery **8**(1), 014008 (2015)
- [69] Watanabe, S.: Tree-structured parzen estimator: Understanding its algorithm components and their roles for better empirical performance. arXiv preprint arXiv:2304.11127 (2023)