

Subkilometer scale ionospheric studies at the SKA-Low site, using MWA extended baselines

María J. Rioja^{a,b,c,*} and Richard Dodson^b

^aCSIRO Astronomy and Space Science, Bentley WA, Australia

^bUniversity of Western Australia, ICRAR, M468, Crawley, Western Australia

^cObservatorio Astronómico Nacional (IGN), Alfonso XII, Madrid, Spain

Abstract. The ambitious scientific goals of the square kilometre array (SKA) require a matching capability for calibration of instrumental and atmospheric propagation contributions as functions of time, frequency, and position. The development of calibration algorithms to meet these requirements is an active field of research. We aim to characterize these, focusing on the spatial and temporal structure scales of the ionospheric effects; ultimately, these provide the guidelines for designing the optimum calibration strategy. We used empirical ionospheric measurements at the site where the SKA-Low will be built, using Murchison widefield array (MWA) phase-2 extended baseline observations and the station-based low-frequency excision of atmosphere in parallel (LEAP) calibration algorithm. We did this via direct regression analysis of the ionospheric screens and by forming the full and detrended structure functions. We found that 50% of the screens show significant nonlinear structures at scales >0.6 km that dominate at >2 km and 1% show significant subminute temporal changes, providing that there is sufficient sensitivity. Even at the moderate sensitivity and baseline lengths of MWA, nonlinear corrections are required at 88 MHz during moderate weather and at 154 MHz during poor weather or for high signal-to-noise ratio measurements. Therefore, we predict that improvements will come from correcting for higher order defocusing effects in observations with MWA phase-2 and further with new developments in MWA phase-3. Because of the giant leap in sensitivity, the correction for complex ionospheric structures will be mandatory on SKA-Low, for both imaging and the tied-array beam formation. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JATIS.8.1.011012](https://doi.org/10.1117/1.JATIS.8.1.011012)]

Keywords: atmospheres; interferometers; imaging.

Paper 21113SS received Sep. 6, 2021; accepted for publication Dec. 1, 2021; published online Dec. 29, 2021.

1 Introduction

Low frequency radio astronomy is undergoing a rebirth, with the arrival of the next-generation instruments that will provide two orders of magnitude improvement in sensitivity and an order of magnitude improvement in frequency coverage. In particular, the low frequency square kilometre array (SKA-Low) will provide an instantaneous frequency coverage of 50 to 350 MHz, a collecting area of 0.4 sq. km, spread over 65 km baselines in the phase-1 and the full 1.0 sq. km in phase-2, and exciting prospects of new science. Thus, it is vital that we update the radio-astronomical methods and strategies to match the potential of the radioastronomical instruments. In this paper, we focus on the challenges at the foundation of accurate direction-dependent ionospheric calibration as these are among the most difficult obstacles to achieving the nominal SKA-Low performance.

Directional-dependent (DD) calibration is vital at low frequencies because the field of view (FoV) is intrinsically large and the propagation effects caused by inhomogeneties in the distribution of the ionospheric plasma (ΔI , measured in TECU or 10^{16} electrons per m^2) impose temporal and spatial phase disturbances on the incoming wavefront with magnitude scales that

*Address all correspondence to María J. Rioja, maria.rioja@icrar.org

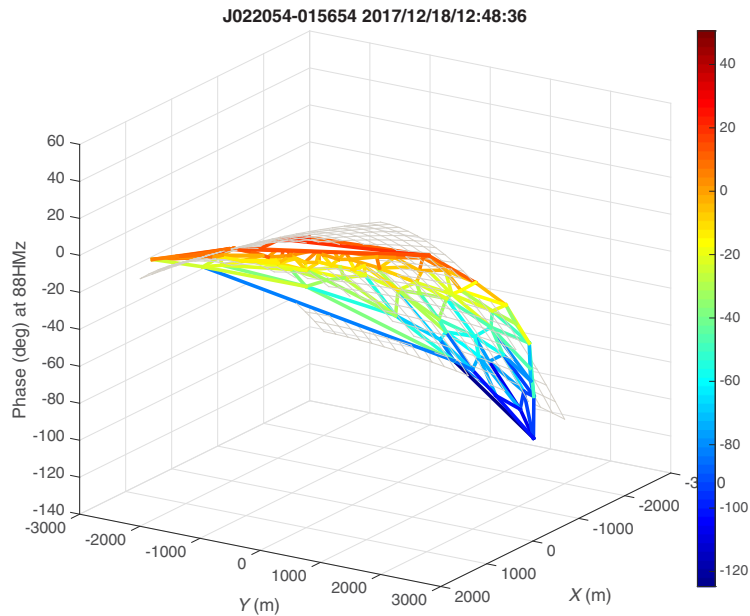


Fig. 1 An example of a LEAP ionospheric phase screen at 88 MHz from a 2-min scan considered in this paper; the span in the sky is 30° – 60° , depending on the height of the ionosphere. Shown as a mesh are the station-based LEAP calibration phases above the MWA Phase-2 extended configuration array for the direction of one LEAP calibrator along a line of sight with an elevation of 65° , color-coded in degrees, with the X and Y axes in meters for the 128 stations (at the nodes of the mesh). The wireframe shows the second-order fit to the data, underlining the high curvature of the surface.

are inversely proportional to the observing frequency ($\phi_{\text{ion}}(\text{deg}) = 480\Delta l\nu_{\text{GHz}}^{-1}$, with ν_{GHz} the observing frequency in GHz^1).

Traditionally, with observations of smaller FoVs and at higher frequencies, these effects could largely be handled with a single station-based correction (or direction-independent (DI) correction), valid for the entire FoV. Lonsdale (Fig. 1)² classified four operating regimes based on the relative sizes of the FoV (V), the ionospheric phase fluctuation disturbances (S), and the projected size of the array (A). Traditionally, with observations of smaller FoVs and at higher frequencies ($V \ll S$), these effects could largely be handled with a single correction per station valid within the FoV. This applies for both compact ($A \ll S$) and extended ($A \gg S$) arrays, which are cases 1 and 2, respectively, for which conventional selfcalibration techniques are suitable.

For modern low-frequency arrays, this is no longer sufficient as the FoV is greater than the scale size of the disturbances ($V \geq S$). Case 3 is for sufficiently compact arrays such that the DD wavefront disturbances over the array can be treated as planar (i.e., linear approximation) and yield coherent apparent source position shifts varying in time and direction. When the linear approximation is no longer valid (i.e., extended array), the presence of nonplanar (i.e., curvature) disturbances additionally leads to source shape deformations; this is case 4, which requires a more complex calibration algorithm, in which an independent solution for each direction and station must be applied. We show here that this is an important issue even for the pathfinders such as Murchison widefield phase-2 array (MWA-2), which has baselines <5 km long.

A prime challenge with DD corrections is in the imaging of the data. The Van-Cittert relationship, which is the basis of all radio-interferometric imaging, requires that the data in the Fourier domain have a uniform calibration applied before transforming it to the image domain. Several research groups are developing solutions for this issue, for example, DDFacet,³ which takes a facetting approach and is widely used for LOFAR LBA imaging, and WSClean⁴ with image domain gridding,⁵ which uses gridding in the image domain to correct for direction-dependent effects and is widely used for MWA imaging.

We looked at efficient methods for generating accurate DD-calibrations, for application in these next-generation imaging tools. We developed a method called LEAP (low-frequency

excision of atmosphere in parallel), which is discriminated from other methods by the fact that every direction can be treated as independent and thus analyzed in parallel. This compares favorably with sequential (peeling) methods (e.g., SPAM⁶ or RTS⁷) that correct for each source in order of strength and simultaneous methods (e.g., SAGECal⁸) that solve for all directions with a very large solution matrix, which do not scale well to large numbers of sources or stations.

LEAP of course has its limitations; the fact that it does not perform source subtraction can elevate the noise floor and the flux density cutoff for calibrator sources. Nevertheless, as it does not require a complete sky model (other than a catalog of LEAP calibrator positions, which are each treated independently) and is very simple and robust, it can play an important role in providing a preliminary solution to the more sophisticated and complicated methods and a unique role in providing the real-time DD calibration for the SKA tied-array beamforming. The latter is used to form multiple simultaneous beams within the FoV, as required in VLBI⁹ and equally in pulsar studies.

Traditional MWA DD-calibration strategies, as used by GLEAM,¹⁰ for example, assume that observations fall in “case 3,” and ignore the potential defocusing effect arising from nonlinear disturbances in the wavefront above the array. The ionospheric calibration is carried out in the image domain as a rubber sheet-type correction to the individual 2-min long scan of images before mosaicking. This approach worked perfectly well for GLEAM (with a maximum MWA phase-1 baseline of 2 km) and has had some success with GLEAM-X (with a maximum MWA phase-2 baseline of 5 km). However, as we will show, for the increased sensitivity of SKA-Low, this is going to be completely inadequate, even for this range of short baseline lengths.

The ionospheric calibration data provide a means to gain information on the physical nature and spatial and temporal structure scales of the propagation media.¹¹ Previous MWA large-scale studies have been performed using image-based analysis;^{12–14} however, these cannot probe the fine scale.

In this paper, we use the LEAP station-based calibration data, in the visibility domain, to quantify the presence of higher order spatial and temporal phase structure over the MWA phase-2 array imprinted in the incoming wavefront as it propagates through the ionospheric plasma irregularities, for directions within the FoV. As LEAP measures the phase variations within the baselines formed by the array, we are able to study the small-scale spatial and temporal structures, that is, at scales from 30 m to 5 km. Our particular interest is to discover at what level the planar distortions approximations breakdown, entering the calibration regime 4, as that is crucial information for the planning of the SKA-Low calibration schemes.

1.1 LEAP Primer

We quickly recap the discussions in Ref. 15 (hereafter LEAP-I) to summarize the operations of LEAP. The core insight of the LEAP algorithm is that the extremely wide fractional frequency bandwidth of current and further low-frequency arrays allows for the use of frequency smearing to suppress all sources away from the phase center. LEAP processing only requires as input the directions for calibrator sources that are sufficiently strong to dominate the phase observable after the other directions are suppressed by frequency averaging. For this, we used the GLEAM-I catalog,¹⁰ so the positions are accurate for the frequencies being used. This simple approach is remarkably effective as the contributions from other directions are suppressed by a factor $\propto \frac{\Delta\nu}{\nu} \theta_0$, where $\Delta\nu$ is the bandwidth, ν is the central frequency, and θ_0 is the source distance from the center in beamwidth units. The fractional bandwidth is very high for modern low-frequency instruments. SKA-Low, for example, has a frequency span ratio close to 1; MWA has one as high as 1:3, and MWA phase-3 (MWA-3) will be even higher. We require the input data to be, at a minimum, bandpass corrected so that the frequency channels can be averaged. For example, if using *a priori* conventional DI selfcalibration (i.e., a flux-weighted average solution across the FoV), LEAP will correct for the DD ionospheric residuals with respect to the flux-weighted DI calibration direction; these DD residuals are small, so we can then determine the DD dispersive delay unambiguously from the phase term only.

An example of the phase screens (hereafter ionospheric screen) over the array observed with the MWA-2, for a given direction, is presented in Fig. 1.

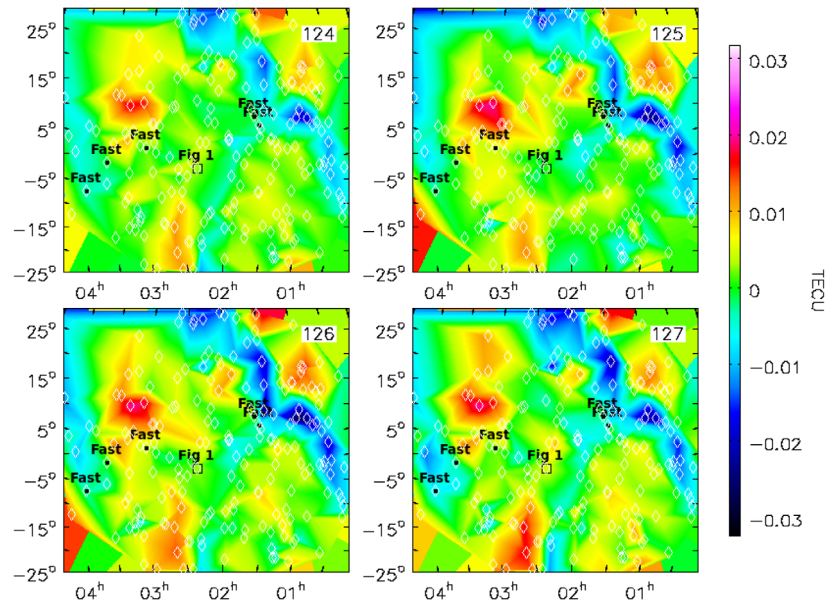


Fig. 2 Example of the ionospheric phase screens as would be provided to WSClean for DD wide-FoV imaging, from a 2-min scan at 88 MHz in our dataset. Each plot represents the ionospheric phase solutions for a given station across the wide MWA FoV, for four stations in the outer region of the extended configuration (i.e., long baselines). The TEC values have been interpolated from the LEAP station-based phase solutions toward the ~ 200 calibrator directions, which are shown with white diamonds. The location of the calibrator shown in Fig. 1 is marked with a black square (labeled Fig. 1), and the location of the few fast changing sources (labeled Fast) are marked with a star.

The interpolation and smoothing of these solutions over the array, and over different pointings, provide an input into, for example, WSClean for DD calibration and imaging. An example is shown in Fig. 2, where the interpolated TEC surface for all directions on the sky is projected onto a regular grid for each station.

2 Observations and Analysis

We used two datasets of GLEAM^{10,16} observations with the MWA phase-2 long baseline configuration¹⁷ at two of the lowest frequency bands, with an instantaneous bandwidth of 30 MHz for the explorations presented for this paper.

One dataset consisted of a series of interleaving 2-min long scans, centered at 88 and 154 MHz, on December 18, 2017, under moderate weather conditions. The interval between observations at the same frequency was 10 min, spanning an hour and a half.

The other dataset includes a small number of 2-min long scans carried out on June 2018, with the strong source 3C444 (J221425-170140) near the phase center, under a variety of weather conditions, that is, good weather on June 3rd and poor weather on June 12th.¹⁸ The latter was flagged to be rejected and reobserved based on severe coherent artifacts (residual sidelobes from several strong sources caused by ionospheric distortions) that contaminated the whole of the image, following the standard GLEAM analysis.

All observations were calibrated using LEAP, following the calibration and imaging procedure described in Sec. 1.1 as part of a wider-scope project to advance the effective imaging implementation (e.g., a single gridding and imaging inversion) of LEAP DD calibration along with a comparative study of “refocusing” calibration in wide FoV imaging using MWA observations (Dodson et al., 2022, in prep).

However, the aim of this paper is to provide answers to some of the questions about the local spatial and temporal scales of the ionospheric disturbances pertinent for the SKA and to identify

the requirements for an optimum strategy for the removal of systematic errors depending on the observing frequency and under different ionospheric conditions.

Of special interest for this work are the LEAP calibration by-products, namely the measured station-based phase solutions that define the so-called ionospheric phase screens above the array, along the directions of the stronger LEAP calibrators (i.e., those with reduced thermal noise error contribution). They are a measure of the wavefront disturbances introduced by its propagation through the ionosphere. Hence, each screen conveys information of the spatial (and temporal) structure of the DD ionospheric disturbances, at scales smaller than the array size (i.e., <5 km and potentially down to 30 m). They are the main observable for the ionospheric analysis in the visibility domain followed here.

The station-based phase observables are obtained with the script discussed in LEAP-I, where *casa* was used to: call WSClean for phase rotation to calibrator source directions and imaging, for analysis and validation and to solve for gain solutions. As part of ICRAR bridging contributions to the SKA science processing, a GPU version that can efficiently process multiple directions in parallel has been written.¹⁹ This addresses the issue of parallel reading of the data; however, the processing here predated the release of that tool.

We followed two methods of analysis of the individual ionospheric phase screens. First was a regression analysis fitting using linear and second-order surfaces.

As a result, we measured the planar “slope” and orientation from the first-order two-dimensional (2D) linear fits, along with the curvature (i.e., obtained as the quadratic mean of the three second-order fit coefficients).

Also we measured the root mean squared residual errors (RMSE) to the first- and second-order fits (RMSE₁ and RMSE₂, respectively) and computed the RMSE fractional improvement, defined for a given screen as

$$\Delta\text{RMSE} = \frac{(\text{RMSE}_1 - \text{RMSE}_2)}{\text{RMSE}_1}.$$

Second, we compute the second-moment structure function (hereafter full structure function or SF) of the ionospheric screens (ϕ_{ion}) over the range of (projected) baselines in the MWA array, r , following the approach in, for example, Ref. 20:

$$D(r) = \langle (\phi(r') - \phi(r' + r))^2 \rangle,$$

where $\langle \dots \rangle$ stands for a statistical average.

The SF analysis is traditionally used to measure the scaling features of the electron density fluctuations in the ionosphere to understand the nature of the physical processes (e.g., turbulences) at the origin of the measured signal.

Here, we use the ability of the SF as a tool to characterize complex surfaces, particularly with nonuniformly sampled and nonperiodic data, using the measured phase values that form the ionospheric screens.^{11,21,22} Of particular interest here is the capacity to detect deviations from linearity. We calculated the SFs using the baseline phase values, deduced from the station solutions. In addition, we formed the detrended structure function (DSF) from the residual phase values, after subtracting the corresponding planar fits from the station solutions. This eliminates the overall dominant influence of a planar slope in the data set.

We expected $D(r)$, in a noise-free environment, to follow

$$D(r) = C^2 r^\beta,$$

where C is the slope of the ionospheric phase screen (i.e., change of phase per unit baseline distance) and β is the so-called scaling exponent, which allows us to characterize the scaling nature of the signal under investigation (in our case, the ionospheric disturbances); this value is equal to 5/3 for pure thin-screen Kolmogorov turbulence or 2 for perfectly planar surfaces. This behavior holds over the inertial region, which is the approximately linear regime in log-log space between the outer scale at which power is injected and the inner scale at which power is dissipated in the Kolmogorov theory of turbulence.

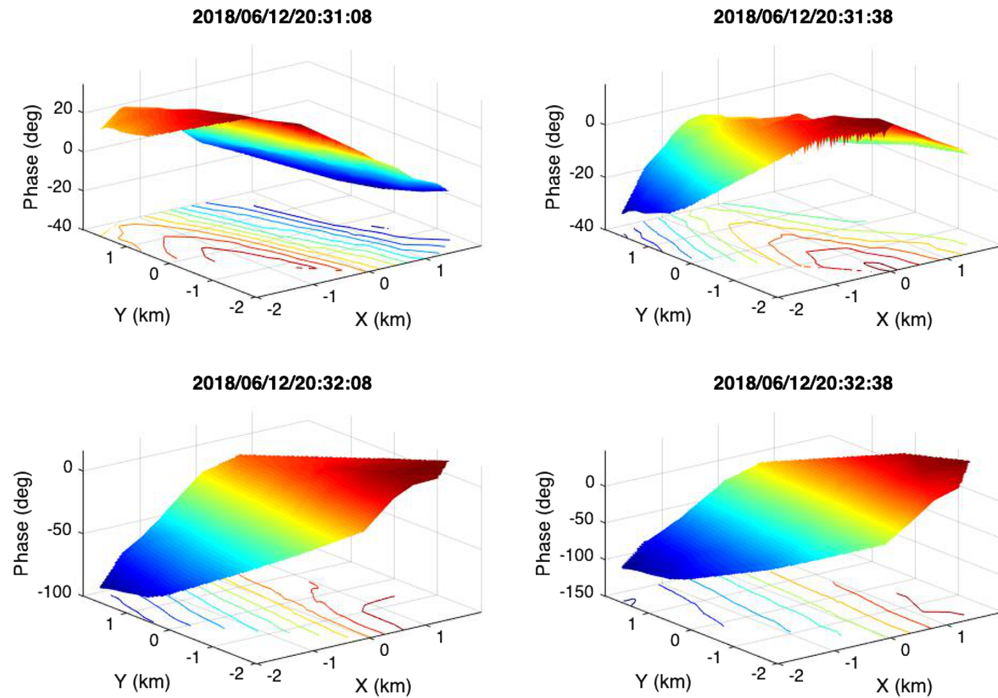


Fig. 3 A highly time variable phase screen, with 30 s sampling, observed on a nominally good weather day (2018/06/12) on 3C444 at 154 MHz. A compact (<1 km) knot of plasma can be seen passing through the line of sight of the array. About 1% of the sources with SNR > 6 have high temporal variability.

We also generated spatial structure functions using simulated datasets of synthetic phase surfaces based on our MWA observations, i.e., ionospheric toy-models with known properties, to test the applicability of SF analysis to the characterization of ionospheric phase screens and to help with interpretation of our empirical findings in terms of ionospheric behavior. The toy models consisted of a family of surfaces defined by 2D linear and quadratic polynomial functions with and without added noise, for a range of values for the polynomial coefficients and noise parameters, compatible with our findings from the regression analysis.

Dodson et al. (2022, in prep.) will present a complementary analysis of the ionospheric disturbances based on the comparison of pre- and postcalibration images. The latter has the capacity to additionally correct for defocusing artifacts in the image, such as changes in the source shape and residual sidelobe patterns, which arise from the small-scale ionospheric disturbances (Fig. 3).

3 Results

Figure 1 shows an example of LEAP ionospheric phase screens above the MWA array in the direction of a strong LEAP calibrator for a 2-min scan solution. For weaker sources, it is expected that the ionospheric signature will be diluted and ultimately buried under the increasing thermal noise. The results presented here comprise the analysis of the ionospheric screens in the directions of all (~200) LEAP calibrators (with signal to noise ratio (SNR) > 2, per station) within the MWA FoV in each 2-min scan, for the set of scans and the analysis described in Sec. 2.

3.1 Regression Analysis

Figures 4(a) and 4(b) show the outcomes from the linear 2D regression analysis of the unwrapped station-based phase solutions in an ionospheric screen, for eight consecutive 2-min scans, at 88 and 154 MHz, respectively. Unwrapping, that is taking into account full wraps of phase by enforcing continuity, is essential to accurately fit the phase slopes on the longer

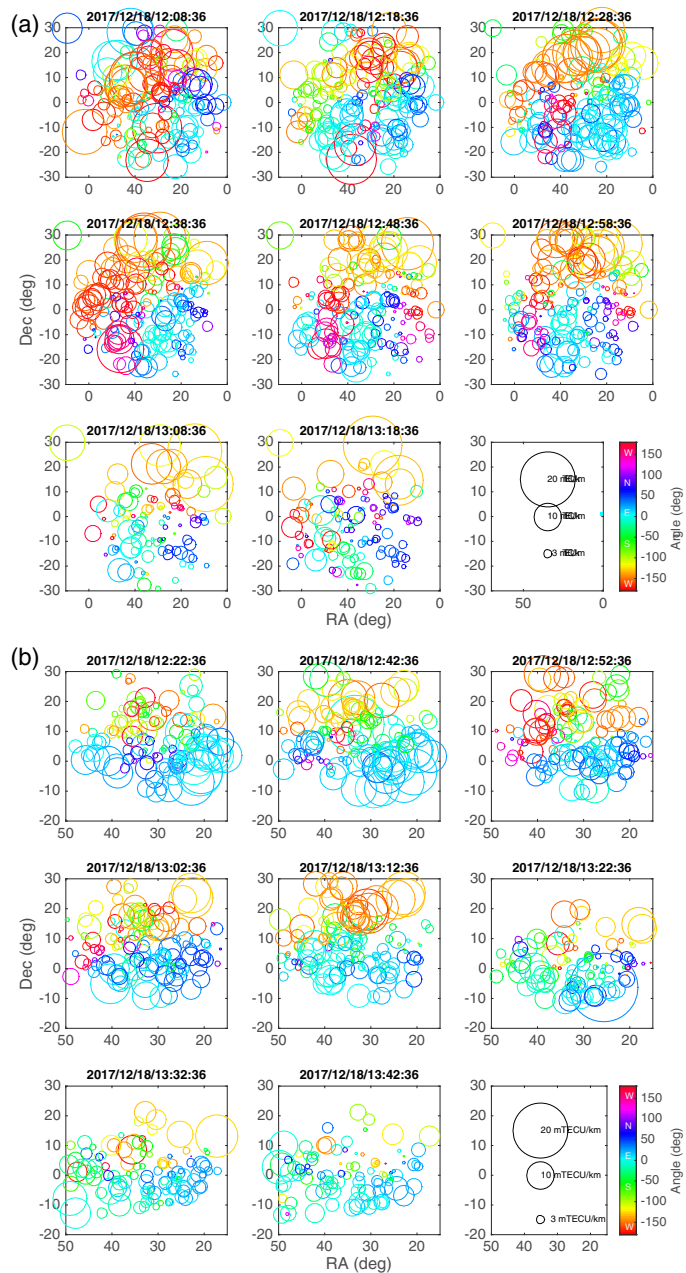


Fig. 4 Outcomes of 2D linear regression analysis of LEAP ionospheric screens: fitted values of slope (circle size, see last frame) and orientation angle (east through north, cyclic color scale) across the array for the RA and DEC directions of LEAP calibrators with SNR > 2 per station. (a) 88 MHz and (b) 154 MHz. Each subplot is for a 2-min scan at the time indicated in the title. The results for any snapshot at either frequency are spatially correlated, even though the ionospheric screens in each direction are measured independently.

baselines. Shown are the magnitude of the “slope” (circle size, mTECU/km) and the orientation angle (circle color, degrees) of the fitted plane. The sky locations of the symbols correspond to that of the LEAP calibrators across the FoV for each scan. The FoV at 154 MHz is smaller than for 88 MHz, and each subplot is taken at a different moment in time (although all of those shown have a common phase center). Thus, the station gains vary between each subplot, and the selected sources vary between scans.

Figure 5 corresponds to the second-order 2D regression analysis, showing the magnitude of the fitted curvature (circle color, mTECU/km²) and slope (circle size with same scale as Fig. 4,

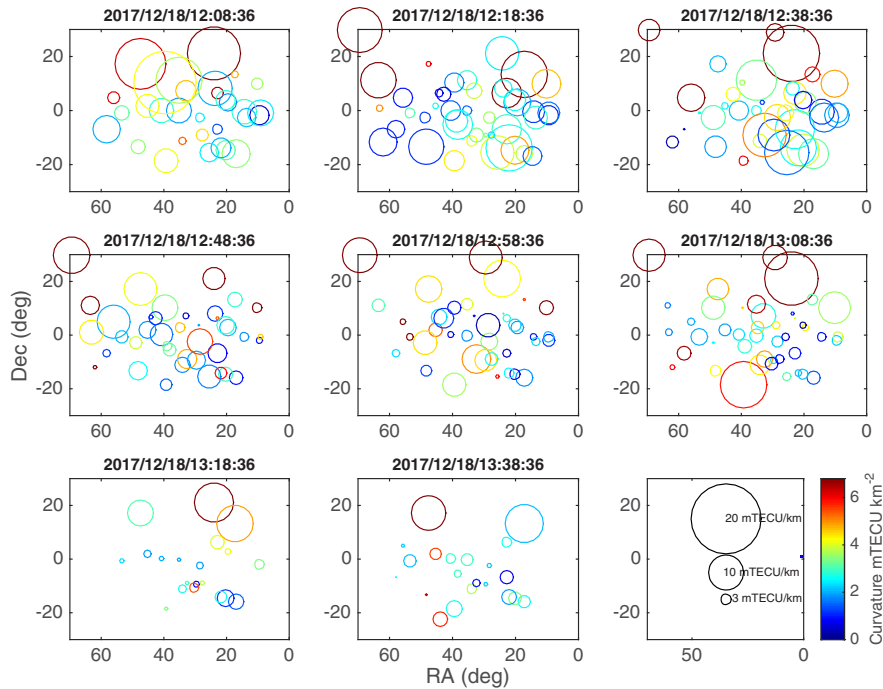


Fig. 5 Outcomes of 2D second-order regression analysis of LEAP ionospheric screens: fitted values of slope (circle size, see last frame and identical to Fig. 4) and curvature (color) across the array for the RA and DEC directions of LEAP calibrators with SNR > 6 per station, at 88 MHz. Each subplot is for a 2-min scan at the time indicated in the title. The curvature shows significantly less spatial correlation than the slope.

in mTECU/km), at 88 MHz, with SNR > 6. Because a higher signal-to-noise ratio (SNR) is required to measure the curvature, there are significantly fewer data points. Furthermore, we do not show the corresponding figure for 154 MHz; there were too few data points per scan to show the spatial correlation.

The median value for the ionospheric slope is 5 ± 3 mTECU/km, across the ~ 5 km span of the array for all LEAP calibrator directions at both 88 and 154 MHz. The median value for the curvature is 3 ± 2 mTECU/km² at both frequencies for SNR > 6.

For each screen, the linear and second-order RMSE errors from the regression analysis are combined to form the fractional Δ RMSE quantity, which would show a significant improvement of one fit above the other as expected if deviations from linearity are significant, that is, cases in which the curvature signature is large enough that its magnitude is comparable to or greater than the thermal noise contribution and results in a large fractional reduction of the RMSE.

Figure 6 shows a histogram of the fractional Δ RMSE values corresponding to the screens shown in Fig. 4, binned by SNR. This shows that the ionospheric curvature signature, if present, appears more significant toward the directions of strong LEAP calibrators, where the thermal noise is reduced.

Some limited insights in the temporal behavior also come from the regression analysis. We searched the solutions from all of the stronger sources for indications of significant variation on time spans shorter than the full scan length of 2 min. Three of the more than 300 source directions with a SNR > 6 per station, for all scans, at each frequency, showed significant temporal variability. Their location is marked in Fig. 2. Figure 3 shows a similar plot to Fig. 1 for the example with the best SNR, with solutions every 30 s. Clear changes in their spatial and temporal structure are easily seen, even by visual inspection.

3.1.1 Structure Functions

Structure function analysis is the traditional method for inferring the nature of a propagation media. We formed the spatial structure functions of the ionospheric screens corresponding

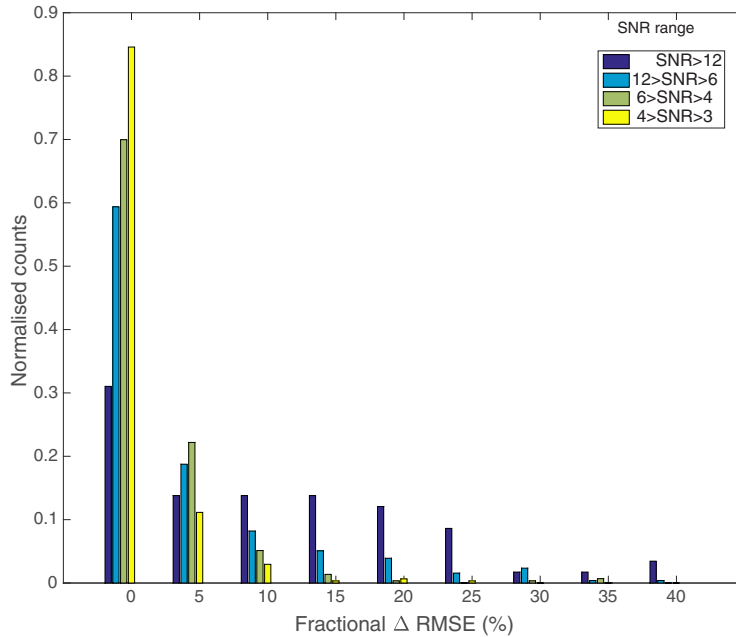


Fig. 6 The fraction of LEAP ionospheric screens showing significant curvature increases with the SNR per station (derived from the RMSE of the 2D fit residuals). The metric for significant curvature is that the fractional difference of the RMSE values from the linear and second-order 2D fits is greater than the expectation values from the low signal-to-noise cases. Here, we show the normalized counts of directions with ionospheric screen curvatures for SNR ranges indicated with color codes at 88 MHz. We conclude that, for SNR > 6 (or RMSE < 10°), most screens are poorly described by a planar surface, even with the small size of MWA-2 array. Therefore, the incidence of curvature is expected to grow larger, in more directions, with increased sensitivity, such as MWA-3, and much more for SKA-Low, which includes much longer baselines in addition to a giant leap in sensitivity.

to the strongest sources in our data across the limited MWA baseline range. For the “full” SF, we use the station-based phase values measured with LEAP in the direction of the source, and the values after subtracting the linear plane from the linear regression analysis, for the “detrended” SF, respectively.

The clearest example, for the direction toward the strongest source 3C444 with a point source flux of 44 Jy at 154 MHz, is shown in Fig. 7, for the full (solid lines) and detrended (dotted lines) SF analysis, for observations on 2 days with “good” (red for 154 MHz and green for 88 MHz) and “poor” (blue for 154 MHz) weather conditions. In this case, the SFs increase with baseline length with gradients of $1.8^{+0.2}_{-0.2}$, $1.7^{+0.2}_{-0.2}$, $1.82^{+0.04}_{-0.04}$, respectively. The noise floor is not reached until very short baselines, below a few hundred meters. The detrended SF analysis shows a rise of the residual signal over the noise at baselines greater than 0.6 km, which is clearer for the “poor” weather observations and lower frequencies.

Figure 8 shows the behavior of the spatial structure functions [(a) full; (b) detrended] for simulated 2D ionospheric screens (i.e., toy models) defined by ionospheric slope and curvature values comparable with the measurements from our regression analysis in the 2017 datasets (i.e., a slope of 5.1 mTECU/km and curvature of 6.8 mTECU/km²) to four times greater for moderate (blue dots) and poor (yellow cross) weather, respectively; for comparison, the case for a linear screen (i.e., no curvature and a slope double the minimum) is shown with green squares. In addition, we tested the effect of measurement noise in the SF by adding a random noise signal with an RMSE of up to 10° per station (i.e., ~ 2 mTECU at 88 MHz) to these toy models (red circles and purple crosses for moderate and poor weather, respectively). Figure 8 (lower) shows the behavior of the gradient of the structure functions across the limited range of MWA baseline lengths as a function of the added noise signals to the toy models.

Figure 9 shows a compilation of the full structure functions of the ionospheric screens toward strong (SNR > 20) sources, for each individual 2-min measurement set and source direction

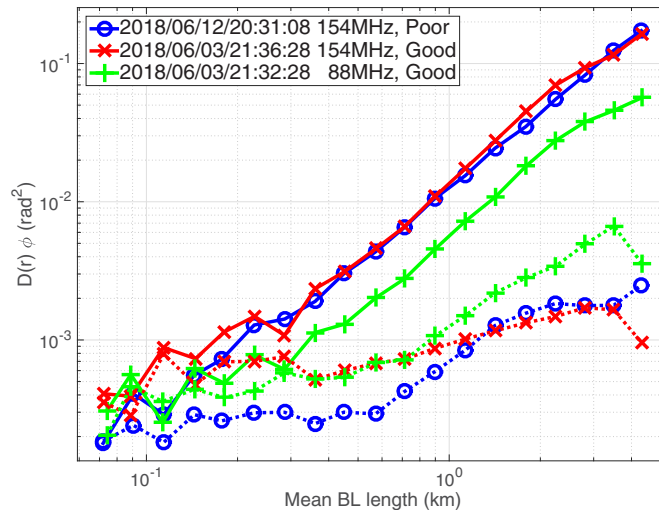


Fig. 7 Full (solid lines) and detrended (dotted lines) spatial structure function $D(r)$ of the LEAP ionospheric screens measured in the direction of the strong source 3C444 (J221425-170140) from MWA-2 observations carried out on June 2018 under good (red crosses) and poor (blue circles) weather conditions at 154 MHz and under good weather conditions (green pluses) at 88 MHz. The full structure functions follow a near-constant gradient across the baseline range, here shown in units of radians² for comparison to Mevius et al. 2016. The DSFs have been calculated after removal of the 2D linear fit from the ionospheric screens. In this case, the transition from noise dominated domain to baseline length-dependent domain occurs at about 0.6 km, which would be our recommended scale for the introduction of higher order DD-calibrations. To scale the y axis to units of mTECU², for comparison with Fig. 9, multiply the 88/154 MHz data by 110/338, respectively.

(dashed lines) and using multiple scans and directions to form the combined SF (black solid line, in bold) at 88 MHz (upper left) and 154 MHz (upper right). A histogram of the structure function gradients over the range of 0.5 to 5 km baselines is shown (lower left), along with the histogram of the ionospheric slopes [measured on the input ionospheric screens with the linear regression analysis] (lower right). Figure 10 shows the DSFs for the same datasets as Fig. 9 at 88 MHz (left) and 154 MHz (right); here, after the contribution from a nominal planar surface is subtracted out, the detrended SF is dominated by the presence of nonlinear structures in the ionospheric screens.

4 Discussions and Conclusions

4.1 Regression Analysis

We used the LEAP measurements of the ionospheric phase surfaces over the MWA-2 array to make a detailed analysis of the spatial structure of the ionospheric distortions imposed on the incoming wavefronts toward ~ 200 simultaneous directions across the FoV for 20 datasets spanning a range of different weather conditions, at 154 and 88 MHz. LEAP, being a station-based phase calibration, is sensitive to higher order ionospheric phase structural changes at small scales, less than the size of the array. Alternative calibration strategies that use image-domain apparent source position shifts to measure the ionospheric phases only measure a simple gradient.¹²⁻¹⁴ However, LEAP analysis requires observations of stronger sources as we now are determining antenna-based corrections¹⁵ rather than corrections from the array-averaged data in the image. Nevertheless, ignoring higher order effects results in severe artifacts in the images and introduces an intractable bias in studies such as EoR.^{23,24} Here, we discuss the results of two approaches to measure spatial small-scale phase deviations that are undetectable to analysis based on the image-domain position shifts.

We are confident that the LEAP calibration method is providing measurements of only the ionospheric surface. Standard practice in MWA analysis is to assume that the beam response is

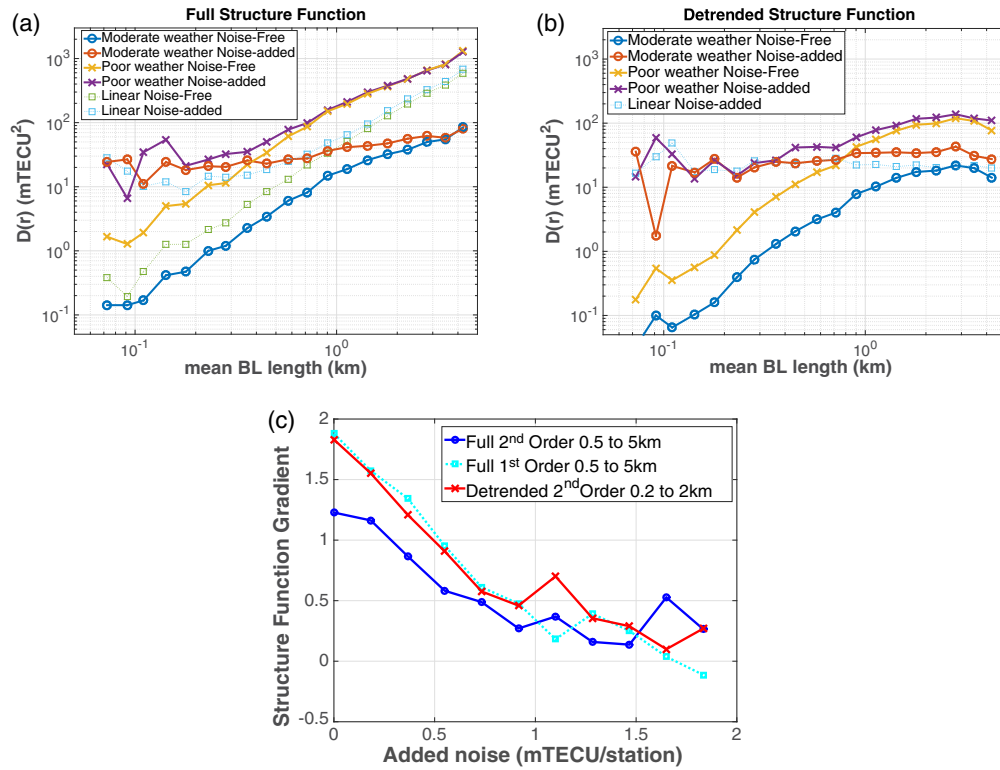


Fig. 8 Spatial structure functions calculated for simulated toy ionospheric models at 88 MHz, based on the ionospheric screens measured with MWA-2. (a) The traditional full structure function (SF) for a 2D-polynomial model with parameters for curvature and slope typical for our moderate-weather observations (circles), noise-free and with added-noise equivalent to a per station SNR of 20. Plotted with crosses are the SF for poor weather with parameters approximately four times greater, with and without noise. Finally, with dotted lines and squares are the SF calculated for a planar model, with and without noise. For a given weather condition, the effect of added noise limits the SF measurements at short baselines and shifts the SF upward, compared with the noise-free case; however, the values and gradients of the SFs are preserved at longer baselines. (b) The detrended SF, calculated after a 2D linear model was subtracted from the data described above, showing the signature of the residual nonlinear ionospheric structures. We consider any excess signal above the noise-floor region at the shortest baselines as the indicator for the presence of nonlinear screen structure. The turn-over (or potentially turn up with even higher order terms) toward the longest baselines indicates the end of the inertial region. (c) The gradients measured from the SF plots versus added noise for the full SF (blue circles and cyan squares fitted over the range of 0.5 to 5 km and for the detrended SF red crosses fitted over 0.2 to 2 km baseline lengths). The blue line is for a second-order screen model, and the cyan line is for a planar model. For the latter, the noise-free gradient should be exactly 2. This shows how the gradient is a sensitive function of the noise-level, particularly with the narrow range of MWA-2 baselines.

the same for all stations. If there were deviations from this assumption, random distortions would appear in the phase surfaces. Such behavior would be expected mainly on the short baselines, where the stations can interfere with each other. Given that we do not see such behavior, i.e., the surfaces are smooth, we can take that the assumption is valid, particularly for this extended baseline configuration.

For the regression analysis, we fitted first- and second-order polynomials to the surfaces to estimate the ionospheric screen parameters: slope, direction, and curvature. These display large-scale coherent behavior in the slopes (Fig. 4) over the FoV, as previously reported.^{13–15} The RMSE values for the second-order fit correlates well with the inverse of the LEAP calibrator flux densities and the expected thermal SNR, that is, the stronger sources have smaller fitted parameter errors, as expected from high signal-to-noise ratio measurements. The coherent behavior in the curvatures, derived in the second-order fit, over the sky is less clear because of the

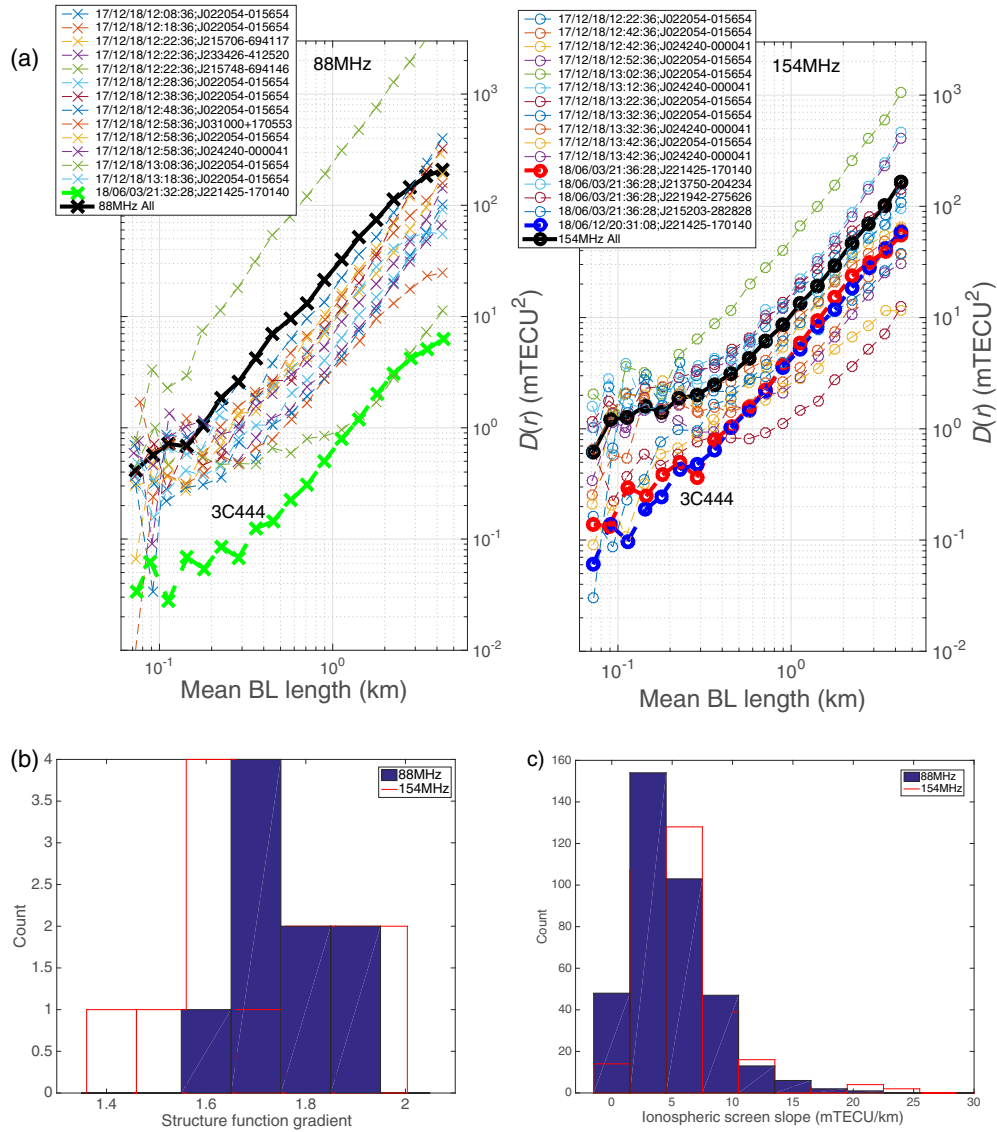


Fig. 9 Upper plots show the empirical spatial structure functions calculated from the MWA-2 LEAP ionospheric screens in the directions of sources with SNR > 20 (per station) versus projected baseline bins at 88 (left) and 154 MHz (right). The intermittent lines are for measurements of an individual screen, color coded to identify the time of the 2-min scan and source direction. The bold thick black line comprises the measurements from the ensemble of all screens, that is, multiple scans and directions, to calculate the structure function. The bold colored lines, labeled with 3C444, are repeats of the data from Fig. 7. The structure functions are shown in units of mTECU² to allow for direct comparison between 88 and 154 MHz. For the case of lowest measurement noise individual screens (toward 3C444, with SNR > 50, shown with the same bold colors as in Fig. 7), the SF extends in a quasilinear fashion to baselines of ~100 m. (b) A histogram of the SF gradients in (a) measured between 0.5 and 5 km, plotted in blue solid for 88 MHz and red outlines for 154 MHz. The median gradient is 1.72 ± 0.15 , which is consistent within errors to that of Mevius et al. (2016)¹¹ and the expectations for a 2D Kolmogorov spectrum (1.89 ± 0.1 and 1.67 , respectively). (c) A histogram of the empirical MWA-2 LEAP ionospheric screen slopes C for all of the LEAP calibrators from the regression analysis, in mTECU/km, at 88 and 154 MHz. The median slope is 5 ± 3 mTECU/km at both frequencies.

sparsity of strong calibrator signals, but it does indicate spatial coherence (Fig. 5) at the lower frequency.

The fractional RMSE change between the first- and second-order fits provides a useful metric for the significant detection of curvature in the presence of measurement noise. When

considering all observations, the RMSE values from the first- and second-order fits are not significantly different for the majority of cases. This is to be expected as the small collecting area of the MWA stations and the low sensitivity result in a thermal noise contribution larger than the ionospheric phase signature for the majority of LEAP calibrator sources. Figure 6 shows that, when introducing LEAP calibrator flux density cutoffs, the impact of the curvature is increasingly visible for higher SNRs with 20% and 40% of cases showing significant ($> \sim 15\%$) fractional RMSE improvements for $6 < \text{SNR} < 12$ and $\text{SNR} > 12$, respectively.

The presence of curvature in the ionospheric screens over the array, if uncorrected, results in increased residual sidelobes after deconvolution, particularly for stronger sources. This underlines the importance of station-based calibration to reduce the residual sidelobes in the cleaned images. This will be fully discussed in Dodson et al. (2022 in prep.), where we show the improvement in the recovered source peak flux densities in the images.

We have limited information for the discussion of high temporal variability of the phase screens, other than noting that three out of the about three hundred strong ($\text{SNR} > 6$) sources in all of the datasets showed significant variation at short timescales. One example is plotted in Fig. 3. Because of the few lines of sight toward strong sources, it is impossible to track the nature of this behavior, which will affect the more sensitive SKA observations.

Significant small-scale deviations from linearity in the spatial ionospheric phase distributions are detected from the regression analysis, even with the limited sensitivity and size of MWA phase-2. This indicates that MWA phase-2 observations reside in the Lonsdale regime 4. Under these conditions, the performance of image-based apparent source position shift ionospheric calibration degrades; thus there will be benefits from using higher order calibration with MWA phase-2. These effects are expected to become more significant in observations with MWA phase-3 and even more with SKA, as discussed below.

4.2 Structure Function Analysis

Because of the excellent ground coverage of the MWA stations, our observations are highly applicable for SF analysis to characterize the fine scale structure of the ionospheric wavefront distortions above the array toward multiple viewing directions and to probe the underlying physical nature of the distorting media.

Because of the denser station coverage, we are able to reconstruct the fine structure much better than the similar study on LOFAR,¹¹ while the latter explores a much larger range of baseline lengths. The observational differences lead to slightly different methods; for example, the tracking LOFAR observations allow for the subtraction of the temporal mean phase, whereas for the snapshot MWA observations, we subtracted a DI calibration. Because of this, in cases in which one very strong source dominates the DI calibration, the residual ionospheric screen slope in the direction of that source (e.g., 3C444) can be significantly lower than the median slope over all directions. An example of this can be seen in Fig. 7, where the intercept at 1 km (that corresponds to the slope of the ionospheric screen) of the SF at 88 MHz is less than the intercept of the 154 MHz SF, where prima-facie one would expect the reverse. This occurs because of the greater dominance of this source in the DI calibration at the lower frequency. However, the log-log SF gradients are preserved. The benefits from high SNR measurements are the potential to probe smaller scales, increased precision of ionospheric calibration, and tighter constraint of the upper limit of the inner scale of turbulence, in the absence of other systematic errors.

MWA is also free of potential clock errors, which were one of the contributions to the systematic noise floor limits in the LOFAR analysis; for 3C444, we reach a noise floor of 0.1, 0.3, and 0.05 mTECU for our three datasets in poor and good weather at 154 MHz and good weather at 88 MHz, as measured in the detrended data over the baseline range of 50 to 100 m. Nevertheless, the two approaches reach similar results as to the SF gradients being most compatible with Kolmogorov turbulence. Mevius et al. used the “refractive scale” R_{diff} as a physically relevant measurement (and a proxy of ionospheric weather quality), whereas we use the screen slope (mTECU/km), which is independent of the SF gradient. These two quantities are trivially convertible and comparable between the two studies. Our median slope of 5 ± 3 mTECU/km is equivalent to R_{diff} being 2 to 12 km, and R_{diff} values of more than 5 km are

considered suitable for EoR observations;¹¹ this of course assumes that the impact of severe weather manifested as a high ionospheric slope is also turbulent and ignores the potential for a high-quality DD-calibration to turn a nominally “bad weather day” into a good weather day. In our comparison of the range of R_{diff} , we note the differences in the methods; LOFAR is a pointed tracking observation with a constant ionospheric wedge over the array subtracted, whereas this analysis is for a wide-FoV experiment with many lines of sight where a global DI model has been subtracted. These differences will affect the deduced R_{diff} , making exact comparison difficult.

We used toy surface models to improve our interpretations of the SF method. These show that, in the presence of noise, interpreting a gradient of $\sim 5/3$ as proof of Kolmogorov turbulence is risky, particularly with the short baseline range of the MWA as noise flattens the gradient in the full structure function. We investigated the impact of joint fitting of the noise floor plus the noise-free structure function (i.e., $C^2 r^\beta + \sigma^2$) but found that, for the MWA, the terms β and σ are highly correlated. We found our best results by limiting the measurement of the gradient to the range of baselines least affected by the noise (that is 0.5 to 5 km). Finally, the signature of the toy model’s curvature is hard to see in the SF as it is dominated by the linear component. However, the DSF does allow for the robust detection of nonlinear distortions, the spatial scales for the onset of these effects, and the end of the inertial-like regime.

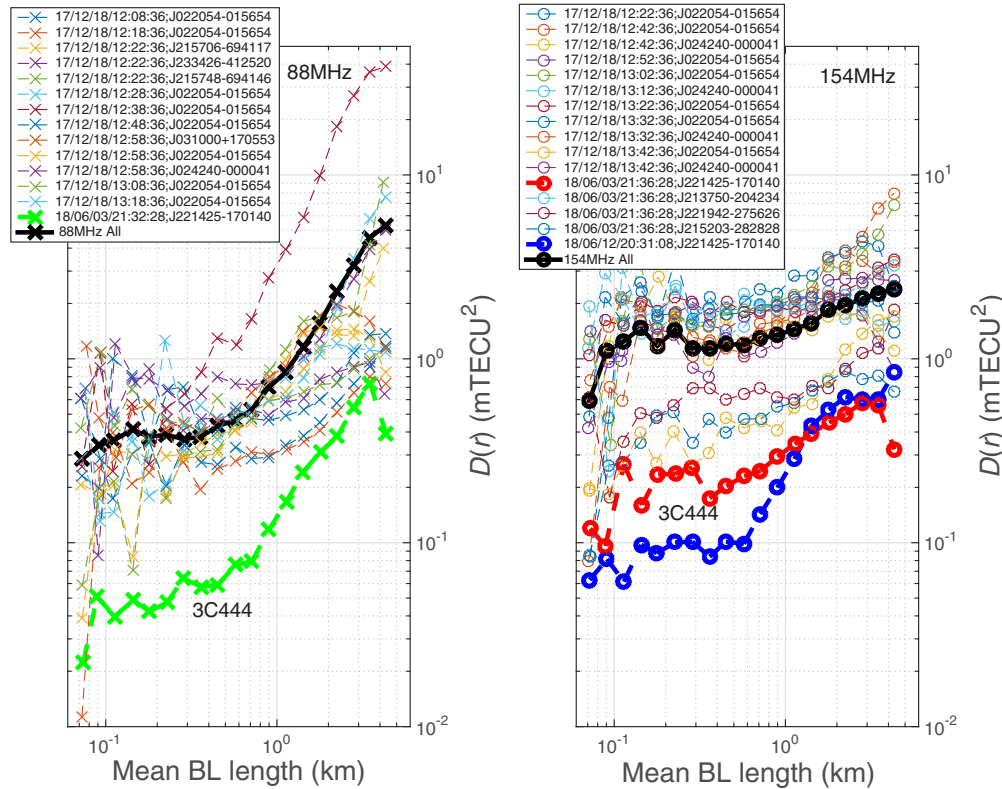


Fig. 10 Empirical detrended spatial structure functions calculated from the MWA-2 LEAP ionospheric screens in the directions of sources with SNR >20 (per station) after subtracting the planar fit versus projected baseline bins at 88 (left) and 154 MHz (right). Any excess signal above the noise unambiguously exposes the presence of the “nonlinear” components in the ionospheric screens. The DSF for individual screens in the direction of 3C444 are shown in bold (green, blue, red) matching the colors in Fig. 7 and for the ensemble over time and directions, shown in bold and black. The bold colored lines, labeled with 3C444, are repeats of the data from Fig. 7. The ensemble DSF at 154 MHz does not show significant nonlinear terms greater than the noise, although individual datasets do show such behavior. On the other hand, both the ensemble and individual DSFs at 88 MHz indicate the presence of significant nonlinear structure at small scales, i.e., baseline lengths greater than 0.6 km.

For the strong source 3C444 (Fig. 7), we can track the structure functions down to the shortest baselines without being overcome by the thermal noise; the gradient in this case is 1.8 at 154 MHz in both good and poor weather days and 1.7 at 88 MHz between 0.5 and 5 km. Thus, at both frequencies, the best fit interpretation would agree with there being Kolmogorov turbulence, but we do not consider this proven given the difficulty of separating the gradient from the noise. Using DSF (i.e., after subtracting a linear surface), the detectable divergence from the linear fit occurs at less than 1 km, which we can confidently interpret as the first indications of nonlinear behavior above the noise. At the longest baselines, there is a change in the DSF gradient, which is characteristic of curvature. In summary, we would recommend that nonlinear corrections are applied for baselines greater than 1 km.

Similar conclusions come from the study of the ensemble of directions and times to the strongest sources in all of the datasets; the gradients are consistent with Kolmogorov turbulence, being 1.6 ± 0.2 and 1.8 ± 0.1 for the 88 and 154 MHz datasets (Fig. 9). But we stress that these measured gradients could be underestimated because of the impact of the thermal noise. The detrended SF (Fig. 10) allows us to better detect deviations from a planar surface. These show that in general at 88 MHz, even in moderate weather, higher order ionospheric corrections should be used for baselines greater than 0.6 km and must be used for baselines greater than 2 km. The combined SF at 154 MHz is more consistent with planar solutions, except for strong sources (i.e., as shown in Fig. 7) and/or poor weather, when higher order corrections are required over the same baseline ranges. This implies that MWA-2 falls in Lonsdale's regime 4.

4.3 Implications for SKA

The SKA-Low stations will be ~ 10 times larger, thus having close to a hundred times larger collecting area (per station). Furthermore, the instantaneous bandwidth will be 10 times larger. Therefore, the sensitivity will be much higher than for MWA, and the importance of precise ionospheric calibration (and other systematic errors) will be even more significant. The expected continuum baseline sensitivity is 3 mJy in 1 min,²⁵ from which we predict more than 100 sources with $\text{SNR} > 6$ in the FoV, using the TREC models.²⁶

We found that for MWA at 88 MHz, even in moderate weather, the influence of higher order terms is detectable for baselines greater than 0.6 km and is significant for baselines greater than 2 km. On the other hand, for most sources at 154 MHz in moderate weather, planar solutions should be sufficient, but for strong sources (i.e., as shown in Fig. 7) and/or poor weather (or both), higher order corrections are required. These conclusions will be most applicable for the SKA, in which the increase in sensitivity means that many sources will be at similar or higher SNR than the few limited examples with the MWA. Thus, we conclude that station-based DD calibration solutions are required for both the final imaging and the tied-array beam forming, even for \sim kilometer-long baselines. The VLBI beamforming, for example, is assuming that all stations within a radius of 20 km (or as a fall-back position 4 km) will be included in the formed tied-array beams.⁹ In this case, real-time station-based DD calibration will be required; this can be provided by LEAP, as we are currently working on demonstrating.

In summary, the opportunities opened up by the amazing potential of SKA-Low to revolutionize low-frequency astronomy require that the calibration of the data is equally accurate. Thus station-based DD corrections will be needed for all baselines and operations.

References

1. O. J. Sovers, J. L. Fanselow, and C. S. Jacobs, "Astrometry and geodesy with radio interferometry: experiments, models, results," *Rev. Mod. Phys.* **70**, 1393–1454 (1998).
2. C. J. Lonsdale, "Configuration considerations for low frequency arrays," in *From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science*, N. Kassim et al., Eds., Astronomical Society of the Pacific Conference Series, Vol. 345, p. 399 (2005).
3. C. Tasse et al., "Faceting for direction-dependent spectral deconvolution," *Astron. Astrophys.* **611**, A87 (2018).
4. A. R. Offringa et al., "WSCLEAN: an implementation of a fast, generic wide-field imager for radio astronomy," *Mon. Not. R. Astron. Soc.* **444**, 606–619 (2014).

5. S. van der Tol, B. Veenboer, and A. R. Offringa, “Image domain gridding: a fast method for convolutional resampling of visibilities,” *Astron. Astrophys.* **616**, A27 (2018).
6. H. T. Intema et al., “Ionospheric calibration of low frequency radio interferometric observations using the peeling scheme. I. Method description and first results,” *Astron. Astrophys.* **501**, 1185–1205 (2009).
7. D. A. Mitchell et al., “Real-time calibration of the Murchison widefield array,” *IEEE J. Sel. Top. Signal Process.* **2**, 707–717 (2008).
8. S. Yatawatta et al., “Data multiplexing in radio interferometric calibration,” *Mon. Not. R. Astron. Soc.* **475**, 708–715 (2018).
9. C. Garcia-Miro, “WP10: VLBI with the SKA,” tech. rep., JIVE (2019).
10. N. Hurley-Walker et al., “GaLactic and extragalactic all-sky Murchison widefield array (GLEAM) survey—I. A low-frequency extragalactic catalogue,” *Mon. Not. R. Astron. Soc.* **464**, 1146–1167 (2017).
11. M. Mevius et al., “Probing ionospheric structures using the LOFAR radio telescope,” *Radio Sci.* **51**, 927–941 (2016).
12. S. T. Loi et al., “Power spectrum analysis of ionospheric fluctuations with the Murchison widefield array,” *Radio Sci.* **50**, 574–597 (2015).
13. C. H. Jordan et al., “Characterization of the ionosphere above the Murchison radio observatory using the Murchison widefield array,” *Mon. Not. R. Astron. Soc.* **471**, 3974–3987 (2017).
14. J. F. Helmboldt and N. Hurley-Walker, “Ionospheric irregularities observed during the GLEAM survey,” *Radio Sci.* **55**, e07106 (2020).
15. M. J. Rioja, R. Dodson, and T. M. O. Franzen, “LEAP: an innovative direction-dependent ionospheric calibration scheme for low-frequency arrays,” *Mon. Not. R. Astron. Soc.* **478**, 2337–2349 (2018).
16. N. Hurley-Walker et al., “GaLactic and extragalactic all-sky MWA-eXtended (GLEAM-X) survey: pilot observations,” MWA Proposal id.2017A-11 (2017).
17. R. B. Wayth et al., “The phase II Murchison widefield array: design overview,” *Publ. Astron. Soc. Aust.* **35**, e033 (2018).
18. N. Hurley-Walker, “Characterisation of the ionosphere via temporal variation of antenna gains,” MWA memo (2019).
19. C. Gray, “CUDA port of LEAP algorithm,” <https://jira.skatelescope.org/browse/SP-994> (2021).
20. S. Van der Tol, “Bayesian estimation for ionospheric calibration in radio astronomy,” PhD Thesis, University of Delft (2009).
21. P. De Michelis et al., “Looking for a proxy of the ionospheric turbulence with Swarm data,” *Sci. Rep.* **11**, 6183 (2021).
22. R. A. Antonia and C. W. V. Atta, “Structure functions of temperature fluctuations in turbulent shear flows,” *J. Fluid Mech.* **84**(3), 561–580 (1978).
23. H. K. Vedantham and L. V. E. Koopmans, “Scintillation noise in widefield radio interferometry,” *Mon. Not. R. Astron. Soc.* **453**, 925–938 (2015).
24. C. M. Trott et al., “Assessment of ionospheric activity tolerances for epoch of reionization science with the Murchison widefield array,” *Astrophys. J.* **867**, 15 (2018).
25. R. Braun et al., “Anticipated performance of the square kilometre array—phase 1 (SKA1)” (2019).
26. A. Bonaldi et al., “The tiered radio extragalactic continuum simulation (T-RECS),” *Mon. Not. R. Astron. Soc.* **482**, 2–19 (2019).

María J. Rioja is a senior research fellow with joint appointments at the University of Western Australia, Space And Astronomy, CSIRO (AU) and the Observatorio Astronómico Nacional (IGN-OAN, ES) with more than 60 journal papers. Her research focus is mainly in astrophysical applications of high-precision astrometry and thus the correction of distortions on long baselines to radio interferometric data. She is focusing on the development of the next-generation of astronomical methods to allow for the delivery on the power of the next-generation of radio telescopes.

Richard Dodson is a senior research fellow at the University of Western Australia (AU). His research focus is on the observational issues that the SKA must address to deliver on the potential of this ground-breaking instrument, thus the systematic errors that will be the true limitations. These include those that come from the ionosphere for SKA-Low, the atmosphere for SKA-VLBI, and imperfections in imaging methods for deep spectral line imaging.