Sean Ting
12/15/05

Removing Ionospheric Corruption from Low Frequency Radio Arrays

Thanks to Shep Doeleman, Colin Lonsdale, and Roger Cappallo of Haystack Observatory for their help in guiding this project in its formative stages over the summer

I. Introduction

The EOR is believed to be the time during which the intergalactic medium (IGM), which consisted entirely of neutral hydrogen, was reionized by the first stellar and quasi-stellar objects. As such, it represents an important era in the history of the universe; however, direct observations of the EOR are currently near impossible. Because of the Gunn-Peterson effect the EOR is difficult to study at wavelengths of less than about one micron (Carilli, 2005). As a result, it is best observed by radio through near-IR wavelengths. One particular candidate for observation, due to the high concentrations of neutral hydrogen prior to the EOR, is the 21-cm emission line of hydrogen. After accounting for the red shifting of the universe during the EOR (estimated to have occurred at some point in $6 \le z \le 17$) this corresponds to an observing frequency on the order of tens to hundreds of MHz, i.e. low frequency radio waves (Carilli, 2005).

This poses a problem because the ionosphere causes large phase delays in low frequency radio waves. In particular, the phase delay of the signal caused by the ionosphere is characterized by the equation:

$$Delay = \frac{40.3}{\nu^2}\int n_e(s)ds \quad (1)\ (Thompson, et\ al\ 2004)$$

υ = *frequency*

$n_e(l)$ = *electron density at length l along path from antenna to top of the ionosphere*

where the integral is taken with respect to the signal's path through the ionosphere. Correspondingly, low frequency radio signals have large phase delays due to the ionosphere and the differential phase delays between two antennae can cause an apparent offset of sources which corrupts the EOR signal.

Thus, in order to map the sky at low frequencies it is necessary to remove the effects of the ionosphere from the data. For radio telescope arrays with small diameters compared to structures in the ionosphere, this can be done by identifying bright calibration sources within the field-of-view. If we know their actual positions from other methods, then by observing their apparent position we can determine the offset induced by the ionosphere for these sources. If the positional offsets for many such sources are known then we can construct a model of how the ionosphere affects areas between calibration sources. In particular the offsets for points between calibration sources can be predicted and so the ionospheric effects can be removed from the data. Note that it is important that the array be smaller than structures in the ionosphere so that the approximation that light traveling from the source goes through the same section of the ionosphere in reaching all antennae holds.

At the present there are no radio telescopes that can remove ionospheric effects to a low enough level for accurate measurements of the EOR because they lack the sensitivity required to observe a large numbers of calibration sources. However, MIT, the CFA, and various Australian groups are constructing the MWA in Western Australia that will have the requisite sensitivity. In order to generate and refine an algorithm for removing ionospheric effects, a simulation of the measurements to be taken was created.

II. Generating a Calibration Algorithm

Currently, the algorithm used to remove ionospheric corruption from low frequency radio observations fits low order Zernike polynomials to the observed offsets (Cotton and Condon, 2002). Because the MWA should have significantly greater sensitivity than the current generation of low frequency radio arrays, it should be able to locate a greater number of bright calibrator sources over the coherence time of the ionosphere. This allows for greater possibilities in the functional fit performed on the offsets. In particular, it should be possible to locate a better space of functions to fit to the data than second-degree Zernike polynomials.

One way to explore these possibilities is by generating a basis of orthogonal functions for a chosen space of functions over the positions of known calibrator sources in the sky-plane. Then using the properties of orthogonal functions, least squares fits to the data can be quickly calculated for various subspaces of the original function space. This idea suggests applying model selection to find the optimal functional subspace. To produce orthogonal functions over a general set of m points:

1) Choose a linearly independent ordered list of functions $\{f_1, f_2, \ldots, f_n\}$, $f_i : \Re^2 \to \Re,\ f_i \in C,\ 1 \le i \le n$ of length less than or equal to the number of calibration sources $\{(x_1, y_1), (x_2, y_2), \ldots, (x_m y_m)\}$.
2) Define an inner-product space over the set of continuous functions $f : \Re^2 \to \Re$ by $\langle f, g \rangle = \sum_{j=1}^{m} f(x_j, y_j) g(x_j, y_j)$. Using this inner-product then we use the traditional definition of orthogonality; two functions are orthogonal if $\langle f, g \rangle = 0$
3) Perform Gram-Schmidt orthogonalization on the functions over the inner-product space to get an orthonormal basis $\{e_1, e_2, \ldots, e_n\}$ for $span\{f_1, f_2, \ldots, f_n\}$

At this stage it is important to note that in step 2 we have not actually defined an inner-product space but rather something that closely resembles an inner product space. Given a vector space V over a field F we generate an inner-product space by defining a real-valued mapping over $V \times V$ satisfying the following properties:

$$i)\ \langle v,v\rangle \geq 0\ for\ v \in V$$
$$ii)\ \langle v,v\rangle = 0 \Leftrightarrow v = 0$$
$$iii)\ \langle u+v,w\rangle = \langle u,w\rangle + \langle v,w\rangle\ for\ all\ u,v,w \in V$$
$$iv)\ \langle av,w\rangle = a\langle v,w\rangle\ for\ all\ a \in F\ and\ v,w \in V$$
$$v)\ \langle w,v\rangle = \langle v,w\rangle\ for\ all\ v,w \in V$$

However the mapping we have defined does not satisfy condition ii); any function f that has zeros at all of the calibration points will satisfy $\langle f,f\rangle = 0$ although we do not necessarily have $f = 0$. This can pose a problem during Gram-Schmidt orthogonalization which will be addressed later, but for the moment assume that none of the new functions $e_i, 1 \leq i \leq n$ produced during Gram-Schmidt orthogonalization satisfy $\| e_i \| = 0$ where $\| \bullet \|$ is the norm derived from the inner-product, $\| f \| = \langle f,f\rangle^{1/2}$. Then it is a basic theorem in linear algebra, one whose proof does not depend on condition ii) of an inner-product holding, that given an inner-product space V, a vector $v \in V$, and a subspace $U \subseteq V$ the value $\| v-u \|\ where\ u \in U$ is minimized by taking u equal to the projection of v on to U. In our case we have $V = C$ and so assume that the phase screen can be represented by some continuous function g that takes values on $\{(x_1,y_1),(x_2,y_2),\ \ldots,(x_m y_m)\}$ corresponding to the measured offsets at those points. Then this theorem yields the result that

$$\| g-e \| = \langle g-e, g-e\rangle^{1/2} = \sqrt{\sum_{j=1}^{m}[g(x_j,y_j)-e(x_j,y_j)]^2}$$ where

$e \in span\{e_1,e_2,\ \ldots,e_n\} = span\{f_1,f_2,\ \ldots,f_n\}$ is achieved by letting e equal the projection of g onto $span\{e_1,e_2,\ \ldots,e_n\}$. In other words we can find best least-squares fit to the offsets within any function space by finding an orthonormal basis for that space under this pseudo-inner-product space and projecting the values of the known offsets onto this space. The nice property of an orthonormal basis is that finding the projection of a vector onto the span of the basis is very efficient. More concretely, having found an orthonormal basis the best-fit function is defined by:

$$g = \sum_{j=1}^{n}\langle e_j, g\rangle e_j$$

As is mentioned above we have not quite produced an inner-product space because we can have $\| f \| = 0$ even if we do not have $f = 0$. This can cause problems in the Gram-Schmidt orthogonalization process because at one step functions are divided by their norms, which results in an undefined function if the norm is equal to 0. Although we have not been able to fully characterize for what distributions of points this occurs, it seems that this is only a problem in points with an underlying symmetry. For instance this regularly occurs if the points all lie exactly on grid points. However, when random distributions of points were used no problem was found, thus, since sources in the sky are distributed randomly this algorithm should not run into problems.

3. Applying Model Selection

Using the above algorithm, it is quick and efficient to generate a functional space of degree approximately fifty, or on the order of the expected number of calibrator sources for the MWA over an 8 x 8-degree field of view. This project focused on finding an optimal polynomial subspace with which to model the ionospheric corruption. Although other functions may improve the fit, polynomials have traditionally been used for calibrating radio arrays and have performed well at low orders. However, because the only empirical results on polynomial fits to ionospheric offsets have been performed with far fewer calibration sources than the MWA should see (likely on the order of three to five times less), it is an open question as to how the MWA should be calibrated. K-fold cross validation provides a logical means for addressing this question for several reasons. First, it can be run separately on the x-offsets and the y-offsets so that if the characteristics between those two sets are different, k-fold cross validation can recognize this and provide a better fit than would occur if a single order were forced on both polynomials. Second, although the number of calibration sources will be increased from in past situations, training examples are still relatively scarce and so k-fold cross validation makes more efficient use of the data than other hold-out cross validation.

In order to analyze the effectiveness of adding automatic model selection to radio array calibration the observation simulation developed at Haystack Observatory in Westford, MA was used. It allows the user to specify an input sky, ionosphere, and radio array and then generates the image the radio array observes after ionospheric corruption. Then, using the original sky map, the positional offsets caused by the ionosphere can be determined. By passing the calibration algorithm the position and offset of the n brightest, sources, a model of how well the algorithm performs as a function of the number of calibrator sources can be generated. Here, the performance of the algorithm can be measured in absolute terms by examining the root-mean-square of the positional offsets with no calibration and then after the fit. It is also important to measure the performance relative to polynomial fits of fixed order.

4. Results and Future Directions

Adding in an automatic model selection mechanism to the algorithm yielded disappointing results. The algorithm was evaluated by looping over twenty randomly generated skies with 150 sources each over an 8 x 8 degree field of view as well as five ionospheres simulating various levels of turbulence and total electron content that the array might face. Initially, because of the scarcity of training examples, leave-one-out cross validation was used. However, this produced results significantly worse than did manually fixing the degree of the fit to be a polynomial of third, fourth, or fifth degree (See Fig. 1 for summary of important results). Examining the order of polynomials averaged into the final fit using LOOCV indicated that this was because when only one point was used to test each fit, there were a large number of zero or one polynomials being averaged, which have little predictive power, and a large number of polynomials of order greater than 5, which over fit the data.

This suggested two additional refinements that could be made. Removing the highest degree polynomials from the functional space should remove the tendency of the model selection to over fit and performing k-fold cross validation for large k, but with k less than n which will give the algorithm more points to test each fit on, again reducing the tendency to over or under fit. Implementing these refinements did improve the algorithm, however it still performed worse than a fixed order fit. Several additional tweaks failed to improve the algorithm significantly.

While, the attempts to employ model selection to improve calibration for low frequency radio arrays was unsuccessful, it is likely that other aspects of machine learning can be applied. After the failure of the bulk of my work on applying model selection I considered the possibility that with the additional points locally weighted linear regression could effectively model the corruption. And initial tests on this have indicated that it performs significantly better than do polynomial fits. Although this method suffers from its inability to produce a single analytic function, this seems like a very fruitful future direction and I will present the results to my advisors at Haystack Observatory this holiday break. There is a further possibility that if supervised learning can be applied to determine the best order fit under different ionospheric conditions, i.e. solar minimum vs. solar maximum, presence of traveling ionospheric disturbances, night vs. day, and for different radio frequencies. Thus, while the initial attempts at applying machine learning to radio array calibration was unsuccessful, it has suggested further avenues of study.

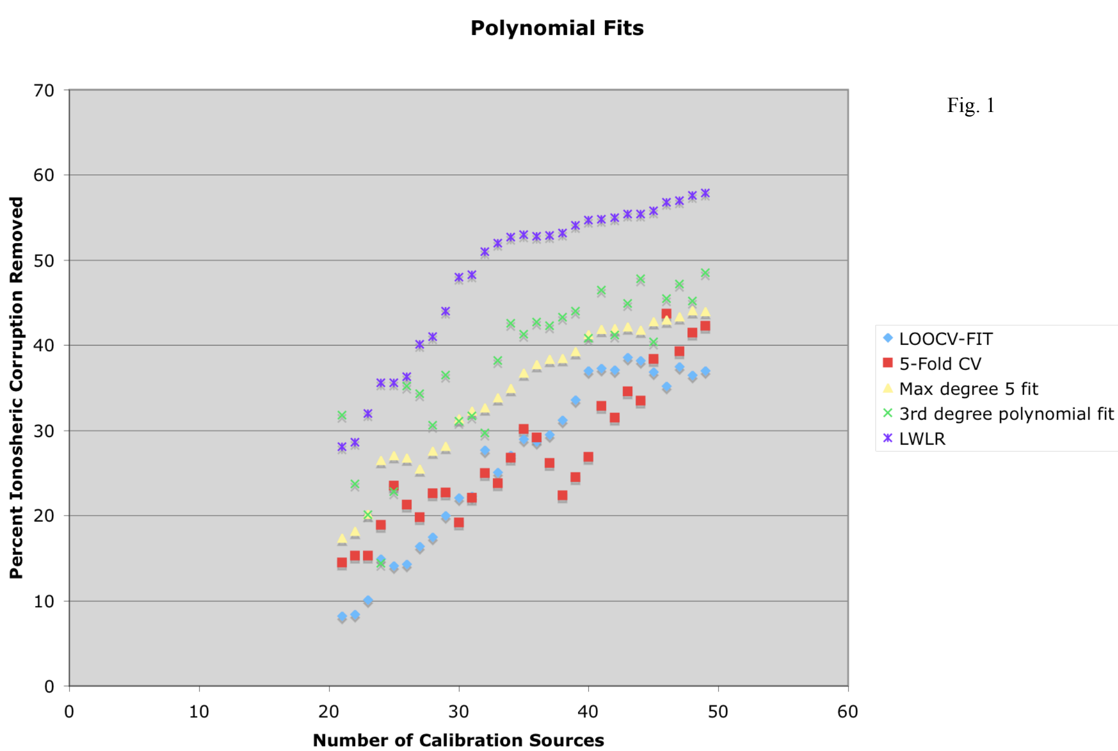

Fig. 1

References

Carilli, C.L., "Radio astronomical probes of cosmic reionization and the first luminous sources: probing the 'twilight zone'." ASP Conference Series, 2004.

Cotton, W. D. and Condon, J. J., "Calibration and imaging of 74 MHz data from the Very Large Array" in Proceedings of URSI General Assembly, 17-24 Aug. 2002, MAAstricht, The Netherlands, paper 0944, pp. 1-4, 2002.

Thompson, R., Moran, J., and Swenson, G. "Interferometry and Synthesis in Radio Astronomy", 1991, New York: Cambridge University Press.