Signals Correlation Algorithms For Cheaper Surveys: Using Windowing Functions.

M. T. Atemkeng[1], O. M. Smirnov[12]⋆, C. Tasse[123], G. Foster[12], J. Jonas[12]

[1] *Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown, 6140, South Africa*
[2] *SKA South Africa, 3rd Floor, The Park, Park Road, Pinelands, 7405, South Africa*
[3] *GEPI, Observatoire de Paris, CNRS, Universite Paris Diderot, 5 place Jules Janssen, 92190 Meudon, France*

ABSTRACT

This paper investigates the use of baseline dependent windowing functions in interferometry data to minimize the loss of signal amplitude (smearing) when the correlated data is averaged over wide bandwidth and long time. In radio interferometry smearing is reduced when a cross-correlator averages the correlated data over narrower bandwidth and shorter integration times. Unfortunately, this leads to a huge amount of data to manage and it is becoming a bottleneck for further data processing such as calibration and imaging. With future generation surveys, it is important to investigate the reduction of the output data rate. Therefore, the focus of this paper is on the use of baselines dependent windowing functions to keep smearing down at an acceptable extent and at the same time significantly suppress signals from out field of view sources, while the nominal sensitivity is conserved.

Key words: Instrumentation: interferometers, Methods: data analysis, Methods: numerical, Techniques: interferometric

1 INTRODUCTION

A radio interferometer measures complex quantities called *visibilities*, which, following the van Cittert-Zernike relation (Thompson et al. 2001), correspond to Fourier modes of the sky brightness distribution, corrupted by various instrumental and atmospheric effects. One particular effect, known as *time* and *bandwidth smearing* (or averaging) occurs when the visibilities are averaged over a time and frequency bin of non-zero extent. This unavoidably happens in the correlator (since the correlator output is, by definition, an average measurement over some interval), but also if data is further averaged post-correlation (both for purposes of compression, and to reduce computational cost).

The effect of smearing is mainly a decrease in the amplitude of off-axis sources. This is easy to understand: the visibility contribution of a point source of flux S located in the direction given by the unit vector $\boldsymbol{\sigma}$ is given by

$$V = S \exp\left\{\frac{2\pi i}{\lambda} \boldsymbol{u} \cdot (\boldsymbol{\sigma} - \boldsymbol{\sigma}_0)\right\}, \qquad (1)$$

where $\boldsymbol{u}$ is the baseline vector, and $\boldsymbol{\sigma}_0$ is the phase centre (or fringe stopping centre) of the observation. The complex phase term above rotates as a function of frequency (due to the inverse scaling with λ) and time (due to the fact that $\boldsymbol{u}$ changes with time, at least in an Earth- or orbit-based interferometer). Taking a vector average over a time/frequency bin then results in a net loss of amplitude. The effect increases with baseline length and distance from phase centre. Besides reducing apparent source flux, smearing also distorts the PSF, since different baselines (and thus different Fourier modes) are attenuated differently.

In the era of big interferometers, where computation (and thus data size) becomes one of the main cost drivers, it is in principle desirable to average the data down as much as possible, without compromising the science goals. There are natural limits to this: firstly, we still need to critically sample the *uv*-plane, secondly, we need to retain sufficient spectral resolution, thirdly, we don't want to average (at least pre-calibration) beyond the natural variation of the calibration parameters, and fourthly, we want to keep smearing at acceptable levels in order not to lose too much signal. In this work, we concentrate specifically on the smearing problem. Here, we can identify two regimes:

• In a compact interferometer, the maximum usable field of view (FoV) corresponds to the primary beam (PB) of the antennas; in most cases (but surveys especially) we want the effective FoV to reach this limit. This imposes an upper limit on the size of a time/frequency bin: it must be small enough to keep amplitude loss acceptably low across the entire PB FoV.

⋆ E-mail: m.atemkeng@gmail.com

• In VLBI, smearing is a lot more severe, so the effective FoV is determined by the smallest time/frequency bin size that a correlator can support, and is normally much smaller than the PB. Modern VLBI correlators overcome this by employing a technique called multiple phase centre correlation, where the signal is correlated relative to multiple phase centres simultaneously, thus effectively "tiling" the PB by multiple FoVs. This has a computational cost that scales linearly with the number of phase centres.

On the other hand, smearing also has a useful side effect. In interferometry, anything outside the desired FoV is unwanted signal. However, the PB pattern of any real-life antenna features sidelobes and backlobes that extend across the entire sky, albeit at a relatively faint level. The faintness makes sidelobes useless for imaging any but the brightest sources: the scientifically usable FoV is that given by the main lobe. However, the sum total signal from all the sources in the PB sidelobes, modulated by their PSF sidelobes, contributes an unwanted global background called the *far sidelobe confusion noise* (FSCN). This imposes a fundamental sensitivity limit; in older telescopes and surveys this was well below the achievable thermal noise and therefore not a worry, but modern and future observatories are capable of reaching this limit (**?**). Even in observations well above the FSCN, individual extremely bright radio sources such as Cyg A or Cas A can contribute confusing signal from even the most distant sidelobe: the LOFAR telescope (**?**) has to deal with the so-called "A-team" sources on a routine basis. Since smearing suppresses distant sources, this somewhat alleviates both the FSCN and A-team problems.

When considering a short sequence of visibilities measured by one baseline, we can think of averaging as a convolution of the true visibility by a boxcar function corresponding to the *uv*-extent of the averaging bin, followed by sampling at the centre of each bin. Convolution in the visibility plane corresponds to multiplication of the image by a *tapering function* that is the Fourier transform (FT) of the convolution kernel; the FT of a boxcar is a Jinc-type taper. If we consider the entire *uv*-plane, averaging is only a pseudo-convolution, since the different *uv*-bins (and thus their boxcars) will have different sizes and shapes as determined by baseline length and orientation. Still, we can qualitatively view smearing as some kind of cumulative effect of an ensemble of image-plane tapers corresponding to all the different boxcars[1].

What if we were to employ weighted averaging instead of simple averaging (whether in the correlator, or in post-processing)? This would correspond to a pseudo-convolution of the *uv*-plane by some ensemble of *windowing functions* (WFs), different from boxcars, which would obviously yield different image-plane tapers, and thus result in different

smearing response. Filter theory suggests that a WF can be tuned to achieve some desired tapering response. An optimal taper would be one that was maximal across the desired FoV, and minimal outside it. In this work, we apply filter theory to derive a set of correlator WFs (CWFs) that approximate this more optimal smearing behaviour. The trade-off is an increase in thermal noise, since minimum noise can only be achieved with unweighted averaging. We show that this effect can be partially mitigated through the use of *extended WFs*.

Cite Offringa and LOFAR.

In the era of the Square Kilometre Array (SKA) and its pathfinders, where dealing with the huge data volumes is one of the main challenges, use of CWFs potentially offers additional leverage in optimizing radio observations. Decreased smearing across the FoV allows for more agressive data averaging, thus reducing storage and compute costs. The trade-off is a loss of sensitivity, which pushes up observational time requirements. However, sinc

N and A-team signal could, conceivably, make up for the loss in nominal sensitivity. In the VLBI regime, use of CWFs potentially offers an increase in effective FoV at a given correlator dump rate, or equivalently, the ability to tile the PB FoV with fewer phase centres, allowing for smaller correlators.

2   OVERVIEW AND PROBLEM STATEMENT

The following formalism deals with visibilities both as functions (i.e. entire distributions on the *uv*-plane), and single visibilities (i.e. values of those functions at a specific point). To avoid confusion between functions in functional notation and their values, we will use $\mathcal{V}$ or $\mathcal{V}(u,v)$ to denote functions, and V to denote individual visibilities. Likewise, $\mathcal{I}(l,m)$ denotes a function on the *lm*-plane i.e. an image. The symbol δ always denotes a function, that is a delta-function.

Depending on whether we want to consider polarization or not, $\mathcal{V}$ can be taken to represent either scalar (complex) visibilities, or 2×2 complex visibility matrices as per the radio interferometer measurement equation (RIME) formalism (Smirnov 2011). Likewise, $\mathcal{I}$ can be treated as a scalar (total intensity) image, or a 2×2 brightness matrix distribution. The derivations below are valid in either case.

We shall use the symbols $\mathbf{u} = (u,v)$ or $\mathbf{u} = (u,v,w)$ to represent baseline coordinates in units of wavelength, and $\mathbf{u}^{\mathrm{m}}$ for units of metres, with $\mathbf{u} = \mathbf{u}^{\mathrm{m}}/\lambda = \mathbf{u}^{\mathrm{m}}\nu/c$

2.1   Visibility and relation with the sky

An interferometer array measures the quantity $\mathcal{V}(u,v,w)$, known as the visibility function. Here, the coordinates u, v and w are vector components in units of wavelength, describing the distance between two antennas p and q, called the *baseline*. The w axis is oriented towards the *phase centre* of the observation, while u points East and v North. Given a sky distribution $\mathcal{I}(l,m)$, where l, m are the direction cosines, the nominal observed visibility is given by the van Cittert-Zernike theorem (Thompson 1999) as

$$\mathcal{V}^{\mathrm{nom}}(u,v) = \iint\limits_{lm} \frac{\mathcal{I}(l,m)}{\sqrt{1-l^2-m^2}}\, e^{-2\pi i\phi(u,v,w)} dl\, dm, \quad (2)$$

---

[1] For completeness, we should note that this "smearing taper" is not the only tapering effect at work in interferometric imaging. Firstly, antennas have a non-zero physical extent: a measured visibility is already convolved by the aperture illumination functions (AIFs) of each pair of antennas. The resulting image-plane taper is exactly what the PB is. Secondly, most imaging software employs convolutional gridding followed by an FFT, which produces an additional taper that suppresses aliasing of sources from outside the imaged region.

where $\phi(u,v,w) = ul + vm + w(n-1)$, and $n = \sqrt{1 - l^2 - m^2}$ (the $n-1$ term comes about when fringe stopping is in effect, i.e. when the correlator introduces a compensating delay to ensure $\phi = 0$ at the centre of the field, otherwise the term is simply n).

Given a pair of antennas p and q forming a baseline $\boldsymbol{u}_{pq} = (u_{pq}, v_{pq}, w_{pq})$, and taking into account the *primary beam* patterns $\mathcal{E}_p(l,m)$ and $\mathcal{E}_q(l,m)$ that define the directional sensitivity of the antennas, this becomes

$$\mathcal{V}_{pq}(u,v) = \iint_{lm} \frac{\mathcal{E}_p \mathcal{I} \mathcal{E}_q^H}{\sqrt{1 - l^2 - m^2}} \, e^{-2\pi i \phi(u,v,w)} dl dm. \qquad (3)$$

Assuming a small field of view ($n \to 1$) and/or a coplanar array ($w = 0$), this becomes a 2D Fourier transform (FT):

$$\mathcal{V}_{pq}(u,v) = \iint_{lm} \mathcal{E}_p \mathcal{I} \mathcal{E}_q^H \, e^{-2\pi i (ul + vm)} dl dm. \qquad (4)$$

The effect of the primary beam can alternatively be expressed in terms of a convolution with its FT, the *aperture illumination function* (AIF) $\mathcal{A}_p(u,v)$. In functional form:

$$\mathcal{V}_{pq} = \mathcal{A}_p \circ \mathcal{V}_{pq}^{\mathrm{nom}} \circ \mathcal{A}_q^H. \qquad (5)$$

2.2 Imaging, averaging and convolution

Earth rotation causes the baseline to rotate in time, which we can denote by $\boldsymbol{u}_{pq}^{\mathrm{m}} = \boldsymbol{u}_{pq}^{\mathrm{m}}(t)$. The baseline in units of wavelength can be treated as a function of frequency and time:

$$\boldsymbol{u}_{pq}(t,\nu) = \boldsymbol{u}_{pq}^{\mathrm{m}}(t)\nu/c. \qquad (6)$$

This, in turn, allows us to express the visibility in eq. (4) as a continuous function of t, ν:

$$\mathcal{V}_{pq}(t,\nu) = \iint_{lm} \mathcal{E}_p \mathcal{I} \mathcal{E}_q^H \, e^{-2\pi i (u_{pq}(t)l + v_{pq}(t)m)} dl dm. \qquad (7)$$

Synthesis imaging recovers a so-called "dirty image" as the inverse Fourier transform of some measured visibility distribution $\mathcal{V}^{(\mathrm{m})}$. This is sampled by a number of baselines pq at specific time/frequency points. We can express the imaging process as

$$\mathcal{I} = \mathcal{F}^H \{ \mathcal{W} \cdot \mathcal{V}^{(\mathrm{m})} \} \qquad (8)$$

where $\mathcal{W}$ is the (weighted) sampling function – a "bed-of-nails" function that is non-zero at points where we are sampling a visibility, and zero elsewhere. Designating each baseline as pq, and each time/frequency point as t_k, ν_l, we can represent $\mathcal{W}$ by a sum of "single-nail" functions $\mathcal{W}_{pqkl}$:

$$\mathcal{W} = \sum_{pqkl} \mathcal{W}_{pqkl} = \sum_{pqkl} W_{pqkl} \delta_{pqkl}, \qquad (9)$$

where δ_{pqkl} is a delta-function shifted to the uv-point being sampled:

$$\delta_{pqkl}(\boldsymbol{u}) = \delta(\boldsymbol{u} - \boldsymbol{u}_{pq}(t_k, \nu_l)) \qquad (10)$$

and W_{pqkl} is the associated weight. The Fourier transform being linear, we can rewrite eq. (8) as

$$\mathcal{I} = \sum_{pqkl} W_{pqkl} \mathcal{F}^H \{ \mathcal{V}_{pqkl}^{(\mathrm{m})} \}, \qquad (11)$$

where

$$\mathcal{V}_{pqkl}^{(\mathrm{m})} = \delta_{pqkl} V_{pqkl}^{(\mathrm{m})} \qquad (12)$$

i.e. is a visibility distribution corresponding to the single visibility sample $pqkl$. We can rewrite eq. (8) again as

$$\mathcal{I}^{\mathrm{D}} = \sum_{pqkl} W_{pqkl} \mathcal{I}_{pqkl}^{\mathrm{D}}, \quad \mathcal{I}_{pqkl}^{\mathrm{D}} = \mathcal{F}^H \{ \mathcal{V}_{pqkl}^{(\mathrm{m})} \}, \qquad (13)$$

which shows that the dirty image $\mathcal{I}^{\mathrm{D}}$ is a weighted sum of dirty images corresponding to the individual visibility samples $pqkl$ (each of which is essentially a single fringe pattern).

Were we to measure instantaneous visibility samples, we would have the simple relation of

$$V_{pqkl}^{(\mathrm{m})} = \mathcal{V}_{pq}(t_k, \nu_l), \qquad (14)$$

where the right-hand side is given by eq. (7). This results in what we'll call the *ideal* dirty image $\mathcal{I}^{DI}$:

$$\mathcal{I}^{\mathrm{DI}} = \sum_{pqkl} W_{pqkl} \mathcal{I}_{pqkl}^{\mathrm{DI}}, \quad \mathcal{I}_{pqkl}^{\mathrm{DI}} = \mathcal{F}^H \{ \delta_{pqkl} \mathcal{V}_{pq} \}, \qquad (15)$$

However, an interferometer necessarily measures the average visibility over a rectangular time-frequency bin given by the *time* and *frequency sampling intervals* $\Delta t, \Delta \nu$, which we'll call the *sampling bin*

$$\mathsf{B}_{kl}^{[\Delta t \Delta \nu]} = \left[t_k - \frac{\Delta t}{2}, t_k + \frac{\Delta t}{2} \right] \times \left[\nu_l - \frac{\Delta \nu}{2}, \nu_l + \frac{\Delta \nu}{2} \right], \quad (16)$$

This, ideally, can be represented by an integration:

$$V_{pqkl}^{(\mathrm{m})} = \frac{1}{\Delta t \Delta \nu} \iint_{\mathsf{B}_{kl}^{[\Delta t \Delta \nu]}} \mathcal{V}_{pq}(t,\nu) d\nu dt. \qquad (17)$$

Inverting the relation of eq. (6), we can change variables to express this as an integration over the corresponding bin $\mathsf{B}_{pqkl}^{[uv]}$ in uv-space:

$$V_{pqkl}^{(\mathrm{m})} = \frac{1}{\Delta t \Delta \nu} \iint_{\mathsf{B}_{pqkl}^{[uv]}} \mathcal{V}_{pq}(u,v) \left| \frac{\partial(t,\nu)}{\partial(u,v)} \right| du dv, \qquad (18)$$

where $\mathsf{B}_{pqkl}^{[uv]}$ is the corresponding bin in uv-space. Note that the sampling bins in $t\nu$-space are perfectly rectangular and do not depend on baseline (assuming baseline-independent averaging), while the sampling bins in uv-space are slightly curved, and do depend on baseline (hence the extra pq index). Assuming a bin small enough that the fringe rate $\partial \boldsymbol{u}/\partial t$ is approximately constant over the bin, we then have

$$V_{pqkl}^{(\mathrm{m})} \sim \iint_{\mathsf{B}_{pqkl}^{[uv]}} \mathcal{V}_{pq}(\boldsymbol{u}) d\boldsymbol{u}, \qquad (19)$$

Now, let us introduce a *normalized boxcar windowing function*, $\Pi^{[t\nu]}$

$$\Pi^{[t\nu]}(t,\nu) = \left\{ \begin{array}{cc} \frac{1}{\Delta t \Delta \nu}, & |t| \leqslant \Delta t/2, \quad |\nu| \leqslant \Delta \nu/2 \\ 0, & \text{otherwise,} \end{array} \right. \qquad (20)$$

using which we may re-write eq. (17) as

$$V_{pqkl}^{(\mathrm{m})} = \iint_{\infty} \mathcal{V}_{pq}(t,\nu) \Pi^{[t\nu]}(t - t_k, \nu - \nu_l) dt d\nu, \qquad (21)$$

which can also be expressed as a convolution:

$$V_{pqkl}^{(\mathrm{m})} = [\mathcal{V}_{pq} \circ \Pi^{[t\nu]}](t_k, \nu_l), \qquad (22)$$

Likewise, eq. (18) can also be rewritten as a convolution in uv-space:

$$V_{pqkl}^{(\mathrm{m})} = [\mathcal{V}_{pq} \circ \Pi_{pqkl}^{[uv]}](\boldsymbol{u}_{pq}(t_k, \nu_l)), \qquad (23)$$

where $\Pi_{pqkl}^{[uv]}$ is a boxcar-like WF that corresponds to bin $\mathsf{B}_{pqkl}^{[uv]}$ in uv-space (and also includes the determinant term of eq. 18). This makes it explicit that each averaged visibility is drawn from a convolution of the underlying visibilities with a boxcar-like WF.

Note what eq. (23) does and does not say. It does say that each individual averaged visibility corresponds to convolving the true visibilities by some WF. However, this WF is different for each baseline pq and time/frequency sample t_k, ν_l (which is emphasized by the subscripts to $\Pi^{[uv]}$ in the equations above). Averaging is thus not a "true" convolution, since the convolution kernel changes at every point in the uv-plane. We'll call this process a *pseudo-convolution*, and the kernel being convolved with ($\Pi_{pqkl}^{[uv]}$) an example of a *baseline-dependent windowing function* (BDWF). In subsequent sections we will explore alternative BDWFs.

In actual fact, a correlator (or an averaging operation in post-processing) deals with averages of discrete and noisy samples, rather than a continuous integration. Ignoring the complexities of correlator implementation (where the sampled quantities are voltages rather than visibilities), let us cast this process in terms of a simple averaging operation. That is, assume we have a set of *hi-res* or *sampled visibilities* on a high-resolution time/frequency grid t_i, ν_j:

$$V_{pqij}^{(\mathrm{s})} = \mathcal{V}_{pq}(t_i, \nu_j) + \mathcal{N}[\sigma_{pqij}^{(\mathrm{s})}], \qquad (24)$$

where $\mathcal{V}_{pq}$ is given by eq. (7), and $\mathcal{N}$ represents the visibility noise term, which is a complex scalar or complex 2×2 matrix with the real and imaginary parts being independently drawn from a zero-mean normal distribution with the indicated r.m.s. (**see noise chapter in ?**).

The *lo-res* or *averaged visibilities* are then a discrete sum:

$$V_{pqkl}^{(\mathrm{m})} = \frac{1}{n} \sum_{ij \in \mathsf{B}_{kl}} V_{pqij}^{(\mathrm{s})}, \qquad (25)$$

where B_{kl} is the set of sample indices ij corresponding to the *resampling bin*, i.e.

$$\mathsf{B}_{kl} = \left\{ ij : \ t_i \nu_j \in \mathsf{B}_{kl}^{[\Delta t \Delta \nu]} \right\}, \qquad (26)$$

and $n = n_t \times n_\nu$ is the number of samples in the bin. Using the BDWF definitions above, this becomes a conventional discrete convolution (assuming a regular $t\nu$ grid):
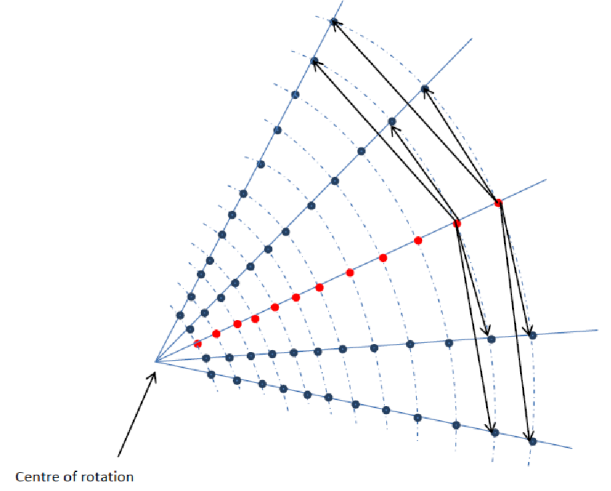
$$V_{pqkl}^{(\mathrm{m})} = \sum_{i,j=-\infty}^{\infty} V_{pqij}^{(\mathrm{s})} \Pi^{[t\nu]}(t_i - t_k, \nu_j - \nu_l). \qquad (27)$$

In uv-space, this becomes a discrete convolution on an irregular grid (the $\boldsymbol{u}_{ij}$ grid being schematically illustrated by Fig. 1):

$$V_{pqkl}^{(\mathrm{m})} = \sum_{i,j=-\infty}^{\infty} V_{pqij}^{(\mathrm{s})} \Pi_{pqkl}^{[uv]}(\boldsymbol{u}_{ij} - \boldsymbol{u}_{kl}), \qquad (28)$$

2.3 Effect of averaging on the image

In the limit of $\Delta t, \Delta \nu \to 0$, averaging becomes equivalent to sampling. An interferometer must, intrinsically, employ a

Figure 1. Schematic of uv-coverage for regularly spaced time-frequency samples.

finitely small averaging interval. The Fourier phase component $2\pi\phi(u, v, w)$ is a function of frequency and time, with increasing variation over the averaging interval for sources far from the phase centre. The average of a complex quantity with a varying phase then effectively "washes out" amplitude, the effect being especially severe for wide FoVs (for an extensive discussion, see Bregman 2012). This effect is often referred to as *time* and *bandwidth smearing*.

The discussion above provides an alternative way to look at smearing. Combining eqs. (8–15) with (23), and using the Fourier convolution theorem, we can see that the dirty image is formed up as

$$\mathcal{I}^{\mathrm{D}} = \sum_{pqkl} W_{pqkl} \mathcal{T}_{pqkl} \mathcal{I}_{pqkl}^{\mathrm{DI}}, \qquad (29)$$

where the baseline-dependent *tapering function* $\mathcal{T}_{pqkl}$ is the inverse FT of the BDWF:

$$\mathcal{T}_{pqkl} = \mathcal{F}^H\{\Pi_{pqkl}^{[uv]}\}. \qquad (30)$$

In other words, the dirty image made from averaged visibilities is a weighted average of the per-visibility ideal dirty images, each one multiplies by its own taper. The FT of a boxcar-like function is a sinc-like function, schematically illustrated in one dimension by Fig. 2. Time and bandwidth smearing represents the average effect of all these individual tapers. Shorter baselines correspond to smaller boxcars and wider tapers, longer baselines to larger boxcars and narrower tapers, and are thus more prone to smearing.

Figure 3 (produced by simulating a series of high time-frequency resolution observation using MeqTrees, and applying averaging) shows the attenuation of a 1 Jy source as a function of distance from phase centre, for a set of different time and frequency intervals. The simulations correspond to JVLA in the C configuration, with an observing frequency of 1.4 GHz. At this frequency, the first null of the PB is at $r \approx 36'$, and the half-power point is at $\sim 16'$, thus we can consider the "conventional" FoV (i.e. the half-power beam width, or HPBW) to be about 0.5° across. Note that the sensitivity of the upgraded JVLA, as well as improvements in calibration techniques ?, allow imaging to be done in the
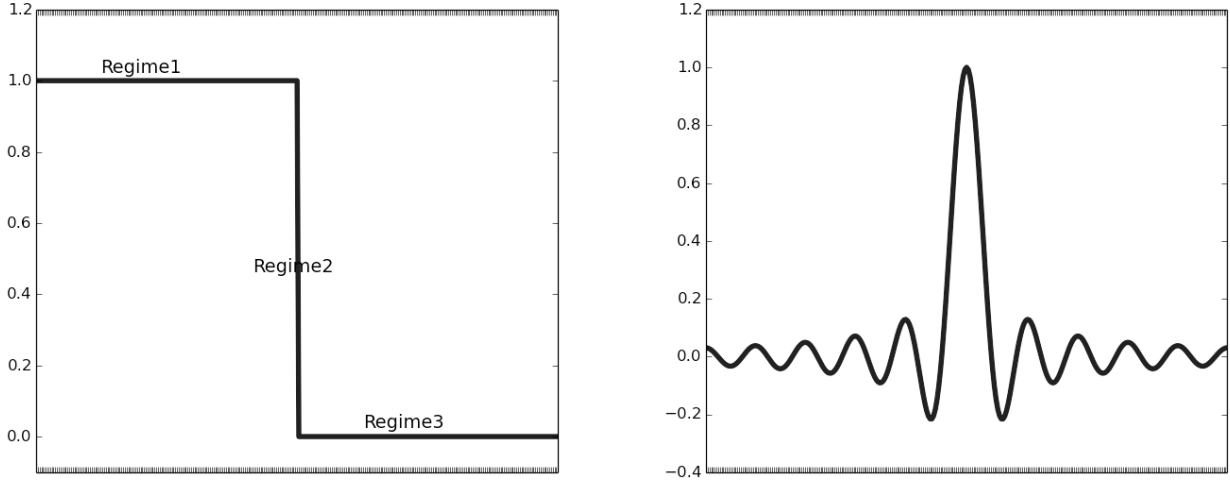
Figure 2. Left: boxcar response. In the uv-plane, this represents the windowing function corresponding to normal averaging of visibilities. In the image plane, this represents the ideal image-plane tapering function. Right: Sinc response. In the image plane, this represents the tapering function corresponding to a boxcar WF in the uv-plane. In the uv-plane, this represents the ideal WF.
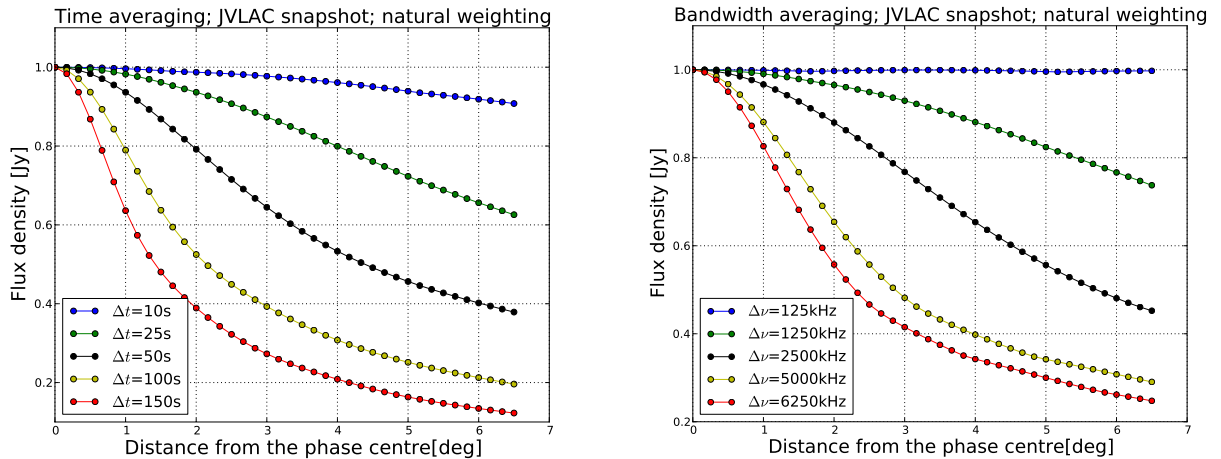
Figure 3. Effects of time and frequency averaging: the apparent intensity of a 1 Jy source, as seen by JVLA-C at 1.4 GHz, as a function of distance from phase centre. (Left) Frequency interval fixed at 125 kHz, time interval varies; (right) time interval fixed at 1s, frequency interval varies.

first PB sidelobe as well (and in fact it may be necessary for deep pointings, if only to deconvolve and subtract side-lobe sources), so we could also consider an "extended" FoV extending out to the second null of the PB at $r \approx 1.25°$. Whatever definition of the FoV we adopt, Fig. 3 shows that to keep amplitude losses across the FoV to within some acceptable threshold, say 1%, the averaging interval cannot exceed some critical size, say 10s and 1 MHz. Conversely, if we were to adopt an aggressive averaging strategy for the purposes of data compression, say 50s and 5 MHz, the curves indicate that we would suffer substantial amplitude loss towards the edge of the FoV.

Finally, note that the curves corresponding to acceptably low values of smearing across the FoV (i.e. up to 25s

and up to 1.25 MHz) have a very gentle slope, with very little suppression of sources *outside* the FoV.

2.4 The case for alternative BDWFs

The tapering response induced by normal averaging (Fig. 3) is far from ideal: it either suppresses too much within the FoV, or too little outside the FoV, or both. The optimal tapering response would be boxcar-like, as in Fig. 2(left). The BDWF that would produce such a response is sinc-like, as in Fig. 2(right). The problem with a sinc is that it has infinite support; applying it over finite-sized bins necessarily means a *truncated* BDWF that results in a suboptimal taper. The problem of optimal filtering has been well studied in signal processing (usually assuming a true convolution rather than

Signal processing	BDWFs
Frequency (freq) domain	Image plane
Time domain	Fourier plane or uv plane
Spectral response	
or freq response	Image plane response (IPR)
Time response	Fourier plane response
Cut-off time interval	
or time pass band	uv averaging bin
Cut-off freq interval	
or freq pass band, or main lobe	FoV
Time stop band	Outside of the uv-bin
Freq stop band	Outside of the FoV
Octave	Doubling in size
Normalized freq	Distance from phase centre
Band-limited	
(applied to visibilities)	restricted FoV

Table 1. Mapping of terminology between signal processing and BDWFs.

the pseudo-convolution we deal with here), and we shall apply these lessons below.

The derivations above make it clear that by using a different BDWF in place of the conventional boxcar-like $\Pi^{[uv]}$ could in principle yield more optimal tapering response. The obvious catch is a loss in sensitivity. Each visibility sample is subject to an independent Gaussian noise term in the real and imaginary part; the noise of the average of a set of samples is minimized when the average is naturally weighted (or unweighted, if the noise is constant across visibilities). Thus, any deviation from a boxcar WF must necessarily increase the noise in the visibilities. Below we will study this effect both theoretically and via simulations, to establish whether this trade-off is sensible, and under which conditions.

3 OVERVIEW OF WINDOWING FUNCTIONS

In signal processing, a WF is a mathematical function with limited support (i.e. zero outside some interval). Conventionally, a time series is convolved with a WF to produce some desired response in the frequency domain. Applying this to our problem can lead to quite some confusion in terminology. Table 1 provides a mapping between the terms commonly used in SP, and their conceptual equivalent in BDWFs.

WF – or rather their corresponding image-plane response (IPR) – can be characterized in terms of various metrics. Some common ones are the peak sidelobe level (PSL), the main lobe width (MLW) and the sidelobes roll-off (SLR) rate. In terms of the "ideal" IPR (Fig. 2, left), these correspond to the following desirable traits:

• Maximally conserve the signal within the FoV ("regime 1" in the figure), and make the transition in "regime 2" as sharp as possible. Both of these correspond to larger MLW.
• Attenuate sources outside the FoV ("regime 3"): this corresponds to a lower PSL and higher SLR.

Below we provide an overview some common (one-dimensional) WFs employed in signal processing.
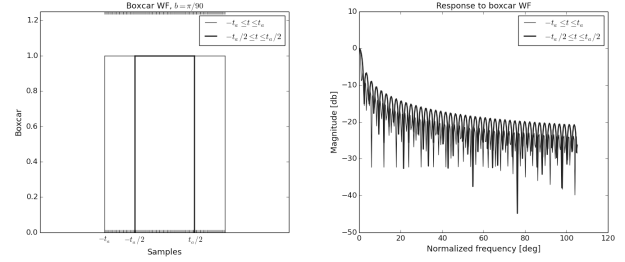
Figure 4. Boxcar windowing function and its tapering response.
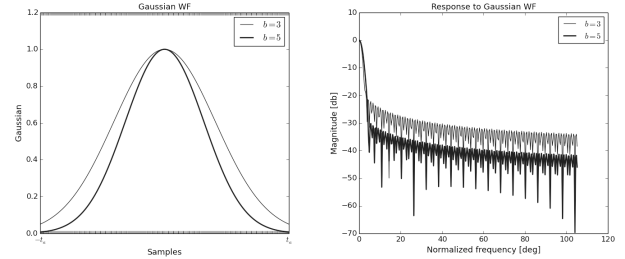
Figure 5. Gaussian windowing function and its tapering response.

3.1 Boxcar window

The boxcar window for a cut-off time interval of $[-t_a, t_a]$ is defined as:

$$\Pi(t) = \begin{cases} 1 & -t_a \leqslant t \leqslant t_a \\ 0 & \text{otherwise} \end{cases} \qquad (31)$$

Fig. 4 shows a plot of $\Pi(t)$ and its response. The blue and red curves correspond to cut-off time intervals of $[-t_a, t_a]$ and $[-t_a/2, t_a/2]$ respectively. Note that when the cut-off time interval is larger, the MLW is narrower, and the sidelobes are lower.

The other WFs given below are all multiplied with a boxcar to ensure a cut-off interval of $[-t_a, t_a]$.

3.2 Gaussian window

A Gaussian WF centred at zero with a standard deviation of σ_1 is given by:

$$G(t) = \Pi(t)e^{-bt^2}, \qquad (32)$$

where $b = (2\sigma_1^2)^{-1}$. The FT of the Gaussian term is given by $\mathcal{F}\{G\} = \sqrt{\frac{b}{\pi}}e^{-cl^2}$, where $c = \pi^2/b$, i.e. is also a Gaussian with standard deviation of $\sigma_2 = (2\pi\sigma_1)^{-1}$.

Fig. 5 shows a plot of $G(t)$ and its response. The WF is truncated at $[-t_a, t_a]$, with $b = 3$ for the blue curve and $b = 5$ for the red curve. Its response is characterized by extremely low sidelobes, but a narrow main lobe.

3.3 Butterworth window

A Butterworth WF is flat in the time pass band, and rolls off towards zero in the time stop band and it is characterized by two independent parameters, the cut-off time time $[-t_a, t_a]$
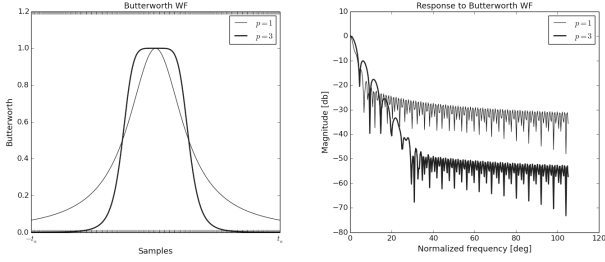
Figure 6. Butterworth windowing function and its tapering response.
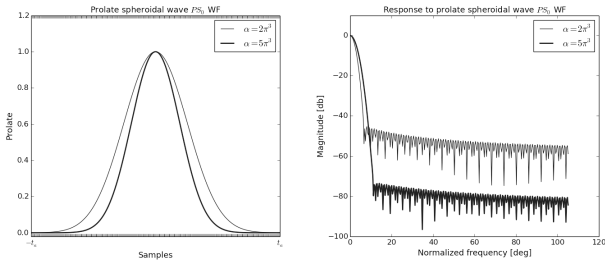
Figure 7. Prolate spheroidal wave (of oder zeros) windowing function and its tapering response.

and the order p. The two parameters control the FoV and sidelobes attenuation. The Butterworth WF is given by:

$$\mathrm{BW}(t) = \Pi(t)\Big(1 + (t/t_a)^{2p}\Big)^{-1}. \qquad (33)$$

Figure 6 shows Butterworth WFs for the same cut-off interval $[-t_a, t_a]$, with orders of $p = 1, 3$. Note that increasing the order p conserves the MLW, and dramatically lowers distant sidelobes, at the cost of pushing up the near-in sidelobes

3.4 Prolate spheroidal window

This WF is given by a prolate spheroidal wave function of sequence zero ($n = 0$) characterized by two independent parameters, the cut-off time $[-t_a, t_a]$ and the order α (Landau & Pollak 1978; Delsarte et al. 1985). The two parameters control the FoV and sidelobes attenuation. The prolate spheroidal wave function PS_0 is the eigenfunction and solution of the integral:
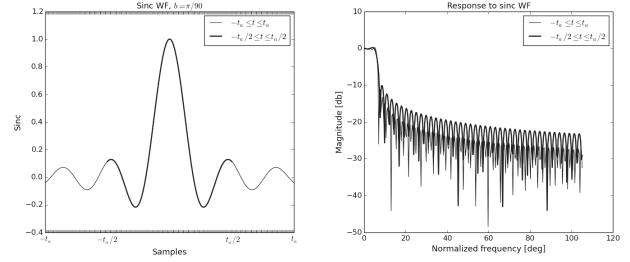
$$\int_{-t_a}^{t_a} \mathrm{PS}_0(\xi) \frac{sin(\frac{\alpha}{\pi}(t-\xi))}{\frac{\alpha}{\pi}(t-\xi)} d\xi = \lambda_{n=0,\alpha,t_a} \mathrm{PS}_0(t), \qquad (34)$$

where $\lambda_{n=0,\alpha,t_a}$ is the coresponding eigenvalue. Fig. 7 shows prolate spheroidal WFs for the same cut-off interval $[-t_a, t_a]$, with orders of $\alpha = 2\pi^3, 5\pi^3$. Note that increasing the order α increases the MLW, and dramatically lowers sidelobes.

3.5 Sinc window

The sinc WF is defined as

$$\mathrm{Sinc}(t) = \Pi(t) \frac{\sin(\pi bt)}{\pi bt}. \qquad (35)$$

Figure 8. Sinc windowing function and its tapering response.

Fig. 8 shows $\mathrm{Sinc}(t)$ for a fixed value of b, and cut-off intervals given by $[-t_a, t_a]$ and $[-t_a/2, t_a/2]$. Note that the response to a sinc WF is almost perfectly flat in the main lobe (more so for larger intervals). The sidelobe response is relatively poor, but better for larger intervals.

3.6 Bessel J_0 window

This WF is given by a Bessel function of the first kind of order zero. Using a power series expansion, we have:

$$J_0(t) = \Pi(t) \sum_{k=0}^{\infty} \frac{(-1)^k (t/2)^{2k}}{(k!)^2} \qquad (36)$$

Fig. ?? shows $J_0(t)$ and its response, with $J_0(t)$ truncated at time intervals $[-t_a, t_a]$ and $[-t_a/2, t_a/2]$. The performance of J_0 is somewhat worse than the sinc within the main lobe, and somewhat better in the sidelobes.

3.7 J_0-Hamming, J_0-Hanning and J_0-Blackman windows

OMS: Marcellin, this is contradictory. If you're passing the signal through two filters as you say, that means two convolutions. But here you write the WF as a product of J_0 with another function. That's convolution with a product, not two convolutions. So which one is correct? Also, please provide citations for the Hanning, Hamming and Blackman filters. Also, in the figures the WFs look absolutely identical. Is this an error, or are they really so similar? If the latter, then we really shouldn't waste space...

A $J_0(t)$ WF can be used to filter out the input signal, providing an intermediate signal. The intermediate signal can then be passed through a Hamming, Hanning or Blackman filter to further increase the pass band (regime 1 or FoV) and increase the stop band (regime 3 or outside the FoV) attenuation. The process can be resume as follow:

$$J_0^{\mathrm{X}} = J_0(t)X(t) \qquad (37)$$

where $X(t)$ is either the Hamming (Hm), Hanning (Hn) or Blackman (B) WFs known in the litterature **OMS: citation please**. Fig. ??, Fig. ?? and Fig. ?? show the $J_0^{\mathrm{Hm}}, J_0^{\mathrm{Hn}}, J_0^{\mathrm{B}}$ WFs and their responses. Compared to the $J_0(t)$ WF, they show lower PSL and higher SLR.

3.8 Relative performance

Table 2 summarizes the performance of the different WFs. From this it is clear that the sinc, the Butterworth (BW)
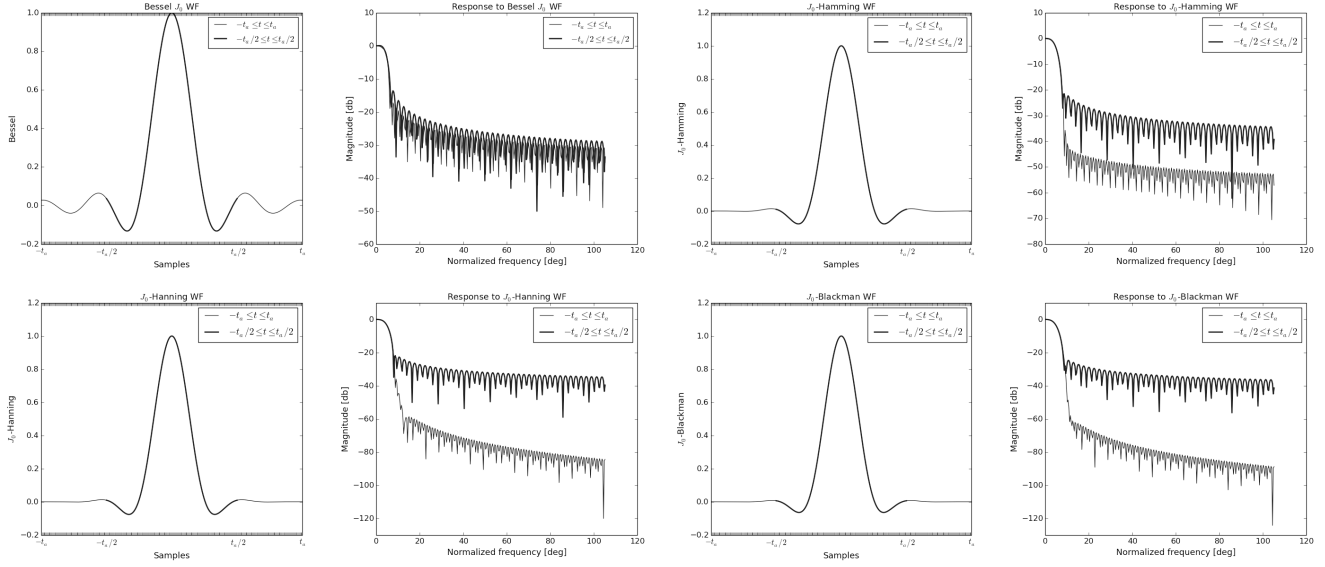
Figure 9. Bessel J_0 windowing and it tapering response, as well as the Bessel-Hamming, Bessel-Hanning and Bessel-Blackman combinations.

WFs		Windows MLW (deg and at -3db)	response PSL (db)	SLR (db/oct)
$\Pi(t)$	$t \in \lvert t_a \rvert$	~ 1.406	-6.663	-12.089
	$t \in \lvert t_a/2 \rvert$	~ 2.812	-6.671	-11.065
$\mathrm{Sinc}(t)$	$t \in \lvert t_a \rvert$	~ 12.304	-10.889	-12.661
	$t \in \lvert t_a/2 \rvert$	~ 12.304	-13.241	-11.447
$J_0(t)$	$t \in \lvert t_a \rvert$	~ 9.140	-14.553	-12.011
	$t \in \lvert t_a/2 \rvert$	~ 9.140	-13.614	-11.794
$G(t)$	b=3	~ 2.109	-21.535	-9.589
	b=5	~ 2.812	-30.211	-9.091
$BW(t)$	p=1	~ 2.109	-13.718	-12.581
	p=3	~ 4.218	-10.145	-27.330
$PS_0(t)$	$\alpha = 2\pi^3$	~ 3.515	-45.302	-7.424
	$\alpha = 5\pi^3$	~ 4.218	-73.597	-6.375
$J_0^{\mathrm{Hm}}(t)$	$t \in \lvert t_a \rvert$	~ 9.140	-35.724	-11.948
	$t \in \lvert t_a/2 \rvert$	~ 9.140	-22.670	-19.527
$J_0^{\mathrm{Hn}}(t)$	$t \in \lvert t_a \rvert$	~ 9.140	-35.954	-41.684
	$t \in \lvert t_a/2 \rvert$	~ 9.140	-27.765	-46.233
$J_0^{\mathrm{B}}(t)$	$t \in \lvert t_a/2 \rvert$	~ 9.140	-48.660	-37.274
	$t \in \lvert t_a/2 \rvert$	~ 9.140	-24.723	-7.972

Table 2. Comparative performance of different windowing functions. **OMS: Marcellin, it is not clear what exact variation of the WF these numbers correspond to. Maybe add a few rows for different variations (i.e. different values of p, b)? In fact, it's not even clear if the comparison is systematic. For example, the FoV realized by different WFs is different.** *Marcellin anwser, thank you I noticed that when I was doing the new plots, I'll do it explicitly and I want to change the scaling of the windows so that we can easily see what you are asking.*

and the Bessel (J_0) WFs provide the more optimal tapering response. It is these WFs that will serve as the basis of BDWFs developed in the rest of this work.

To construct two-dimensional BDWFs from one-dimensional WFs, we will use the following definitions:

$$
\begin{aligned}
S(u,v) &= S(u)S(v), \\
J_0(u,v) &= J_0(r), \\
BW(u,v) &= BW(r), \quad r = \sqrt{u^2 + v^2}. \quad (38)
\end{aligned}
$$

OMS: Marcellin, why not $\mathrm{sinc}(r)$ as well? The referee is bound to ask? Also, why did you never consider a prolate spheroidal WF? I remember bringing it up a few times.
Marcellin: $\mathrm{sinc}(r)$ doesn't works well in signal recovery. I studied the prolate spheroidal wave functions(PSWF) and the oder 1 of PSWF was just slightly like the sinc but with low sidelobes. I can try and test this but for sure the computational time is extremely huge. I was thinking to talk about the Hamming*[sinc or bessel], Hinning*[sinc or bessel], blackman*[sinc or bessel] there are also performing well.

4   APPLICATION OF WFS TO VISIBILITIES

While visibilities are (usually) regularly sampled in $t\nu$-space, in uv^{m}-space this is not so. In frequency the sampling positions go as $\sim \nu^{-1}$, while in time, baselines with a longer East-West component sweep out longer tracks between successive integrations (Fig. 1). Applying a WF with a constant integration window in $t\nu$ space corresponds to different-sized windows in uv-space. In the case of normal averaging, this results in the boxcar-like window $\Pi^{[uv]}$ of eq. (23) having a baseline-dependent scale. The scale of the tapering response being inversely proportional to the scale of the WF, this

results in more smearing (i.e. a narrower FoV) on longer baselines.

By defining our alternative WFs in uv-space (in units of wavelength), we can attempt to "even out" the smearing response across baselines. For a given BDWF $X(u,v)$, we have the following recipe for computing the lo-res visibility kl (compare to eq. 28):

$$V_{pqkl}^{(\mathrm{m})} = \frac{\sum\limits_{i,j\in\mathsf{B}_{kl}} V_{pqij}^{(\mathrm{s})} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}{\sum\limits_{i,j\in\mathsf{B}_{kl}} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}, \qquad (39)$$

where $\boldsymbol{u}_{pqkl}$ is the midpoint of the resampling bin B_{kl} in uv-space. The main lobe of the WF then has the same scale across the entire uv-plane, while the resampling bins have different uv-sizes. Conversely, in $t\nu$-space the bins are regular, while the main lobe of the effective WF scales inversely with the baseline fringe rate. Furthermore, the WF is truncated at the edge of each bin; on the shortest baselines this truncation is extreme to the point of approaching the boxcar-like $\Pi^{[uv]}$ (Fig. 10).

The downside of this simple approach is twofold. Firstly, while all of the WFs above boast far lower sidelobes than the boxcar (i.e. more suppression for out-of-FoV sources), they no longer perform that well under truncation, with extremely truncated WFs at the shorter baselines becoming boxcar-like. Secondly, taking a weighted sum in eq. **??** increases the noise in comparison to normal averaging.
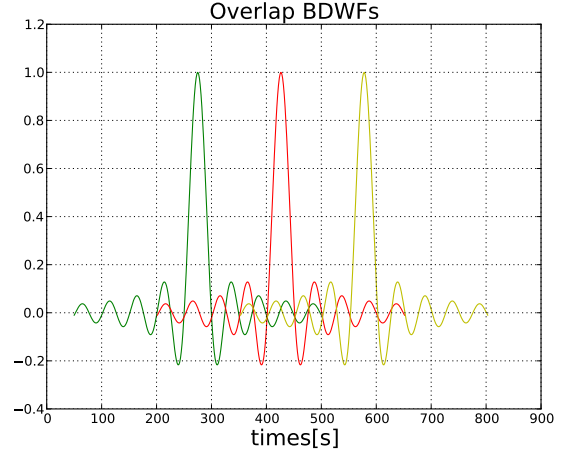
4.1 Overlapping BDWFs

A more sophisticated approach involves overlapping BD-WFs. Normal averaging implicitly assumes that the resampling bins B_{kl} in eq. (**??**) don't overlap for adjacent kl, since they represent adjacent averaging intervals. There is, however, no reason (apart from computational load) why we cannot take the sum in eq. (**??**) over larger bins. Let us define the *windowing bin* for *overlap factors* of α, β as

$$\mathsf{B}_{kl}^{[\alpha\beta]} = \{ij : t_i\nu_j \in \mathsf{B}_{kl}^{[\alpha\Delta t, \beta\Delta\nu]}\}, \qquad (40)$$

i.e. as the set of sample indices corresponding to a bin of size $\alpha\Delta t \times \beta\Delta\nu$ in $t\nu$-space. Let us then compute the sums in eq. (**??**) over the windowing bin. This becomes distinct from the resampling bin: while the latter represents the spacing of the lo-res output samples, the former represents the size of the convolution kernels. Only for $\alpha = \beta = 1$ do the two bins become the same.

BDWFs in the overlapping regime are schematically illustrated in Fig. 11. For normal averaging overlapping offers no benefit, since it only broadens $\Pi^{[uv]}$ and therefore increases smearing. But for a well-behaved BDWF, enlarging the windowing bin (while maintaining the same WF scale) means less truncation – thus lower sidelobes – and decreased noise, as more sampled visibilities are taken into account.

To distinguish overlapping BDWFs from non-overlapping ones, in the rest of the paper we will designate the WFs employed as *WF-$\alpha\times\beta$*. For example, sinc-3×2, J_0-1×1 (i.e. no overlap), etc. If resampling is only done in one direction (only time or only frequency), we'll indicate this as e.g. J_0-3×−.

Figure 11. Overlapping BDWFs representing adjacent resampling bins. This corresponds to overlap factor $\alpha = 3$ in time.

4.2 Noise penalty estimates: analytic

OMS: Marcellin, I can't follow this, and neither will the referee. The notation is a bit of a mess. Need to straighten it out and express in terms of Sect. 2.2, i.e. lo-res (averaged) and hi-res (sampled) visibilities. Suggest also working it out for the unpolarized case instead, you can then drop the extra "4" in the dimensions and simplify the matrices. Also, you need to explain, what is N_t and n_t?

Maybe it makes sense to leave the complicated derivations for your thesis, and go with something simple here.

Let us now work out analytically the *noise penalty* associated with replacing an unweighted average by a weighted sum. For simplicity, let's assume that the noise term has constant r.m.s. σ_s across all baselines and samples. If the resampling bin B_{kl} consists of $n = n_t \times n_\nu$ samples, then the averaged visibility in eq. (25) will contain a noise term with its variance given by

$$\sigma_\mathrm{m}^2 = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_\mathrm{s}^2 = \frac{\sigma_\mathrm{s}^2}{n} \qquad (41)$$

Applying this to the imaging equation (13), we have the following expression for the variance of the noise term in each pixel of the dirty image:

$$\sigma_\mathrm{pix}^2 = \frac{(\sum_{pqkl} W_{pqkl}^2 \sigma_\mathrm{m}^2)}{(\sum_{pqkl} W_{pqkl})^2}, \qquad (42)$$

which for natural image weighting ($W \sim \sigma_\mathrm{m}^{-1}$, i.e. $W \equiv 1$ in this case) is simply

$$\sigma_\mathrm{pix}^2 = \frac{1}{N} \frac{\sigma_\mathrm{s}^2}{n}, \qquad (43)$$

where N is the total number of averaged visibilities.

To simplify notation, let's replace $pqkl$ by a single index μ, enumerating all the lo-res visibilities $V_\mu^{(\mathrm{m})}$, with $\mu = 1\ldots N$. If we now employ eq. (39) to compute the lo-res visibilities using some BDWF $X(u,v)$, the noise term
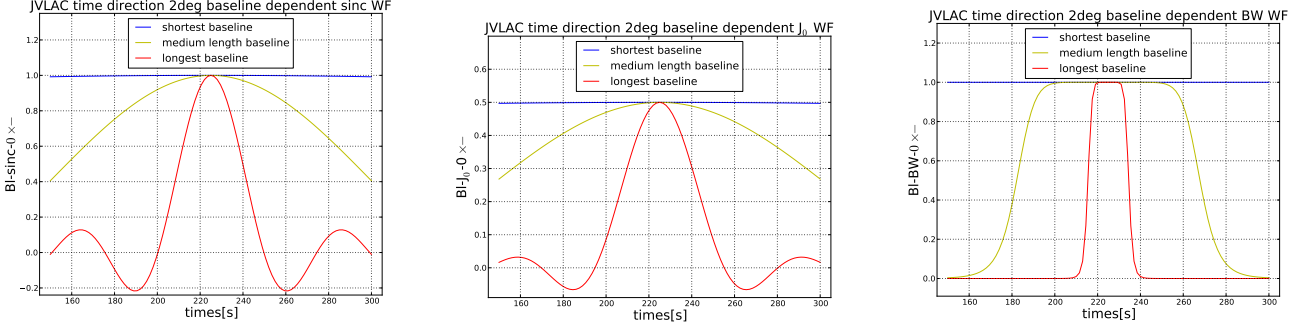
Figure 10. Cross-sections through three different BDWFs (left: sinc, centre: Bessel, right: Butterworth) defined in uv-space, plotted along the time axis. This shows that the effective WF is a scaling and truncation of the underlying WF, with the shortest baselines reducing to a boxcar-like WF.**OMS: Marcellin please make the PDF figures greyscale, and the same size. Also please leave legend on only one of them, it's redundant.**

becomes different per each visibility μ:

$$\sigma_{X\mu}^2 = \frac{\sum X^2(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}{\left[\sum X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})\right]^2} \, \sigma_{\mathrm{s}}^2 \qquad (44)$$

where both sums are taken over $i, j \in \mathsf{B}_{kl}$. Let us define the *visibility noise penalty* associated with BDWF X and visibility μ as the relative increase in noise over the unweighted average, i.e.

$$\Xi_{X\mu} = \frac{\sigma_{X\mu}}{\sigma_{\mathrm{m}}} = \frac{\sqrt{n \sum X^2(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}}{\sum X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}. \qquad (45)$$

It is easy to see that $\Xi_{X\mu} \geqslant 1$, and only reaches 1 when $X \equiv 1$.

We can now compute the *image noise penalty* associated with imaging weights W and BDWF X as

$$\Xi_W^X = \frac{\sigma_{\mathrm{pix},X}^2}{\sigma_{\mathrm{pix}}} = \frac{n}{\sigma_{\mathrm{s}}} \frac{(\sum_\mu W_\mu^2 \sigma_\mu^2)}{(\sum_\mu W_\mu)^2} = \frac{(\sum_\mu W_\mu^2 \Xi_{X\mu}^2)}{(\sum_\mu W_\mu)^2} \qquad (46)$$

In the case of natural weighting ($W_\mu = \sigma_\mu^{-1}$) this reduces to:

$$\Xi_{\mathrm{nat}}^X = \frac{N}{\sum_\mu \Xi_{X\mu}^{-1}}. \qquad (47)$$

4.3 Noise penalty estimates: empirical

We show in this section that on shorter baselines, the sinc, the J_0 and the BW reduce to the Boxcar window. Therefore, we effectively applying a hybrid window in the Fourier plane; a Boxcar on the shorter baselines and the WF under consideration on the longer baselines (see Figure **??**, **??** and **??**). We furthermore evaluated the array theoretical noise ratio, $R_\sigma = \epsilon_{ar,oo}^{bd}/\epsilon_{ar,oo}^{bx}$ predicted in Eq.**??** of a 2° BDWF averaging by the one of boxcar averaging and compared the ratio to the simulation one. For the analysis we used a JVLAC measurement set (MS) of 7min30s synthesis with a 1.5s integration time at 1.4Ghz, with 150 channels of width 125kHz. The MS is filled with 1Jy thermal noise. We then processed boxcar averaging and BDWF averaging over 1min30s (width of the time integration) and 12500kHz (channel with). Two overlaps BDWF are presented, the first overlap over 2min60s across time and 3125kHz across frequency; the second overlap over 3min45s across time and 6250kHz across frequency.

• In Figure **??**, **??** and **??** each *dot mark* is the ratio [2] of per baseline centre pixel theoretical rms noise ($\epsilon_{pq,oo}^{bd}$) of a 2° BDWF averaging by the one of boxcar averaging ($\epsilon_{pq,oo}^{bx}$). The ratio is plotted as a function of $\overline{\mathbf{d}}_t$ (the mean of the baseline (u, v) distance). The figures show that, the noise increases with baseline length.

• On shorter baselines, the noise estimate of BDWF averaging is approximately the same with the one of boxcar averaging. Also, on shorter baselines, the noise drops significantly with the number of overlap time and frequency bins. As mention above, this is because on shorter baselines, the BDWF are closer to the boxcar window, and when the number of overlap time and frequency bins increases, the ratio: $\epsilon_{pq,oo}^{bd}/\epsilon_{pq,oo}^{bx} \approx \sqrt{n_t n_\nu / ((n_t + 2n_{ovlpt})(n_\nu + 2n_{ovlp\nu}))}$.

The table below summarized the array theoretical rms noise ratio and the simulation one. One do not need to worry about a perfect agreement of the theoretical noise results with the simulation one. Unfortunately, the theoretical noise results are the rms noise of the sky map centre pixel while the simulation results are an average of the sky map overall pixels rms noise.

BDWF	Theoretical	Simulation
Bl-sinc-0×0	1.17	1.23
Bl-sinc-100×25	1.02	1.18
Bl-sinc-150×50	1.03	1.27
Bl-J_0-0×0	1.11	1.14
Bl-J_0-100×25	0.91	1.08
Bl-J_0-150×50	0.92	1.13
Bl-BW-0×0	1.16	1.55
Bl-BW-100×25	1.44	1.49
Bl-BW-150×50	1.42	1.44

5   SIMULATIONS AND RESULTS

In order to test the approaches described in sections **??** and 4.1, multiple tests are performed on JVLAC simulated MS.

---

[2] $\epsilon_{pq,oo}^{bd}/\epsilon_{pq,oo}^{bx}$, see Eq.**??** for a recall of $\epsilon_{pq,oo}^{bd}$ and $\epsilon_{pq,oo}^{bx}$.
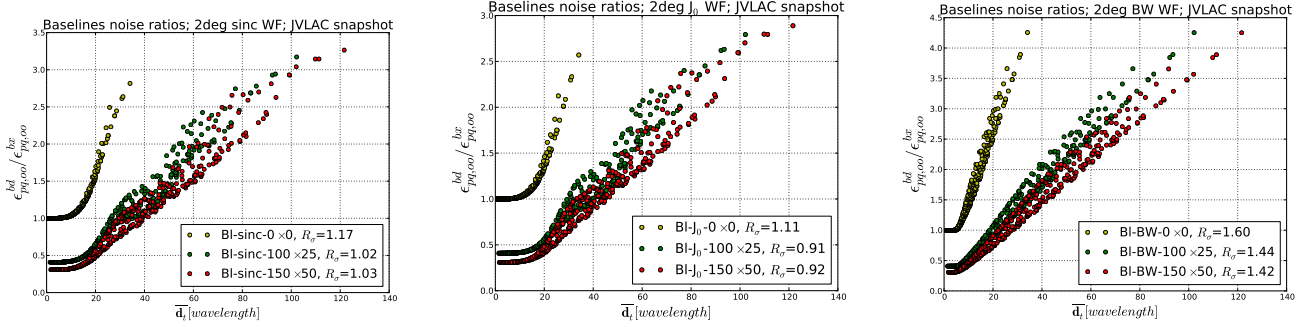
Figure 12. Noise penalty w.r.t. normal averaging as a function of baseline length for three different filters (sinc, Bessel and Butterworth), with and without overlap. JVLA-C, 1.4 GHz, sampling intervals of 150s and 6.25 MHz. **OMS: Marcellin please check the numbers I gave are correct. Also please make the PDF figures the same size .**

This section summarized and discussed those results. Two MS are used in the simulation, a high resolution MS (HR-MS) that contains the observed JVLAC data of short integration time and frequency and a low resolution MS (LR-MS) where the results of boxcar averaging or BDWF averaging are saved. In order to apply an overlap BDWF the following conditions have to be satisfied:

(i) If t_{start}^{hrms} and t_{start}^{lrms} are the starting time of the HR-MS and LR-MS respectively and Δt^{hrms} the HR-MS integration time, then $t_{start}^{lrms} \geqslant t_{start}^{hrms} + 2n_{ovlpt} \times \Delta t^{hrms}$.

(ii) If t_{end}^{hrms} and t_{end}^{lrms} are the ending time of the HR-MS and LR-MS respectively, then $t_{end}^{lrms} \leqslant t_{end}^{hrms} + 2n_{ovlpt} \times \Delta t^{hrms}$.

(iii) If ν_{start}^{hrms} and ν_{start}^{lrms} are the starting frequency of the HR-MS and LR-MS respectively and $\Delta\nu^{hrms}$ the HR-MS channel width, then $\nu_{start}^{lrms} \geqslant \nu_{start}^{hrms} + 2n_{ovlp\nu} \times \Delta\nu^{hrms}$.

(iv) If ν_{end}^{hrms} and ν_{end}^{lrms} are the ending frequency of the HR-MS and LR-MS respectively, then $\nu_{end}^{lrms} \leqslant \nu_{end}^{hrms} + n_{ovlp\nu} \times \Delta\nu^{hrms}$.

5.1 Smearing elimination and out FoV suppression

The study considered 40 different sky models each containing a 1Jy source; the sky models differ from each other by the source coordinates. The reason for having only one source in the sky model is to avoid standard imaging artefacts in the sky maps which can affect the brightness of sources. Another obvious reason is the CLEAN algorithm (Cornwell et al. 1999), which does not create perfect clean maps. Each sky model is simulated with a JVLAC HR-MS of 7min30s synthesis, with a $\Delta t^{hrms} = 1.5s$ integration time at 1.4GHz, with 150 channels width $\Delta\nu^{hrms}$=125kHz. Each HR-MS is then processed with boxcar averaging and BDWF averaging the results are saved into a JVLAC LR-MS of one timeslot of with 150s, with 1 channel of width 6.25MHz. The study also considered, $\{t_{start}^{hrms}, \nu_{start}^{hrms}\} = \{0s, 125kHz\}$, $\{t_{start}^{lrms}, \nu_{start}^{lrms}\} = \{150s, 6250kHz\}$, $\{t_{end}^{hrms}, \nu_{end}^{hrms}\} = \{7min30s, 18750kHz\}$, $\{t_{end}^{lrms}, \nu_{end}^{lrms}\} = \{6min, 12500kHz\}$, $n_{ovlpt} \in \{0, 100, 150\}$ and $n_{ovlp\nu} \in \{0, 25, 50\}$.

The source flux density as a function of the source distance from the phase centre is depicted in Figures **??** to **??**. The study in Figure **??**, **??** and **??** compared boxcar aver-

aging with 2° FoV of Bl-sinc-$n_{ovlpt} \times n_{ovlp\nu}$, Bl-J$_0$-$n_{ovlpt} \times n_{ovlp\nu}$ and Bl-BW-$n_{ovlpt} \times n_{ovlp\nu}$ respectively. A study with a 2° FoV suggest that: "*Regime 1*" is defined from 0° to 1°; "*Regime 2*" from 1° to 1.5° and "*Regime 3*" from 1.5° to infinity. The study in Figure **??**, **??** and **??** compared boxcar averaging with 4° FoV of Bl-sinc-$n_{ovlpt} \times n_{ovlp\nu}$, Bl-J$_0$-$n_{ovlpt} \times n_{ovlp\nu}$ and Bl-BW-$n_{ovlpt} \times n_{ovlp\nu}$ respectively. A study with a 4° FoV suggest that: "*Regime 1*" is defined from 0° to 2°; "*Regime 2*" from 2° to 3° and "*Regime 3*" from 3° to infinity. The results show that:

• "*Regime 1*" (FoV recovery): The source flux density is recovered with the BDWF averaging compared to the boxcar averaging. Also, as the overlap time and frequency bins increase, the source flux density becomes optimal with the overlap sinc and J$_0$ (see the green and the red curve from Figures **??** to **??**) which is not the case with the overlap BW as shown in Figures **??** and **??** (the green and the red curve).

• "*Regime 2*" (nearby sources suppression): The source flux density is suppressed using boxcar averaging compared to BDWF averaging without overlap bins (see Bl-sinc-0 × 0, Bl-J$_0$-0 × 0 and Bl-BW-0 × 0 shown from Figure **??** to **??**). It is also seen from Figure **??** to **??** (the green and the red curve) that, the source flux density is suppressed within ≈ 25% of the regime using boxcar averaging than the sinc and J$_0$ overlap BDWF averaging while within ≈ 75% of the regime the sinc and J$_0$ overlap BDWF averaging suppressed the source compared to the boxcar averaging. The performance of the overlap BW BDWF averaging is not accurate in this regime.

• "*Regime 3*" (far away sources attenuation): It appears on Figures **??** to **??** that the curves of boxcar averaging (see Bx-avg-0 × 0) mid that of Bl-sinc-0 × 0 and Bl-J$_0$-0 × 0 which implied an equivalent flux density attenuation while the curve of Bl-sinc-100 × 25, Bl-sinc-150 × 50, Bl-J$_0$-100 × 25 and Bl-J$_0$-150 × 50 are below the one of Bx-avg-0 × 0 which implied that the overlap sinc and J$_0$ BDWF attenuated the source in the regime more than boxcar averaging. Also, as the overlap time and frequency bins increase, the attenuation becomes significant. The similar results occurs 1° away from the regime starting interval with BW BDWF averaging.

5.2 Maximum integration

In this section, the maximum frequency and time integration intervals that can be considered in the frequency and the time directions without loss of signal amplitude is evaluated. The study considered 6 sky models each containing 1Jy source; the sky models differ by the source coordinates with radius $r \in \{0.09°, 0.25°, 0.5°, 1°, 1.5°, 2°\}$. The study compared the results of boxcar averaging to a 2° BDWF averaging. As mention above, a 2° FoV suggest that: "*Regime 1*" is defined from 0° to 1°; "*Regime 2*" from 1° to 1.5° and "*Regime 3*" from 1.5° to infinity. Therefore, a sky model with a source radius $r \in \{0.09°, 0.25°, 0.5°, 1°\}$ belongs to "*Regime 1*", $r = 1.5°$ belongs to "*Regime 2*" and $r = 2°$ to "*Regime 3*".

5.2.1 Frequency direction

Boxcar and BDWF averaging are processed across the frequency direction. Each of the sky models is simulated with a JVLAC HR-MS of 1 timeslot of width $\Delta t^{hrms} = 0.1s$ at 1.4GHz. The raison of a short HR-MS integration time is to avoid time direction smearing. The simulation considered 150 channels and varied the channels width $\Delta \nu^{hrms}$ in the range $[125, 1187.5]$kHz. The result of the process was saved into a LR-MS of 1 timeslot of width $\Delta t^{lrms} = 0.1s$ at 1.4GHz with 1 channel of width $\Delta \nu^{hrms} \times 50$kHz. We considered $\nu_{start}^{hrms} = \Delta \nu^{hrms} kHz$, $\nu_{start}^{lrms} = \nu_{start}^{hrms} \times 50kHz$, $\nu_{end}^{hrms} = \nu_{start}^{hrms} \times 150kHz$ and $\nu_{end}^{lrms} = \nu_{start}^{lrms} \times 100kHz$.

The source flux density as a function of the channel width is depicted in Figures ?? to ??. The study compared the maximal channel width of the boxcar frequency averaging (Bx-avg-− × 0) with that of 2° FoV BDWF frequency averaging (Bl-sinc-− × 0, Bl-sinc-− × 50, Bl-J$_0$-− × 0, Bl-J$_0$-− × 50, Bl-BW-− × 0 and Bl-BW-− × 50). The results show that:

• "*Regime 1*": The BDWF frequency averaging maintained the flux density of the source for wider frequency channels. However, it is seen on the Figures that the curve of sources in the regime are closer to the affine line y=1 when the channel width increases.

• "*Regime 2*" and "*Regime 3*": It appears on the Figures that Boxcar frequency averaging attenuated the sources in these regime compare to BDWF frequency averaging without overlap while overlap sinc and J$_0$ frequency averaging attenuated the source for channels widths between [6.2, 12.6] compare to the boxcar.

5.2.2 Time direction

Boxcar and BDWF averaging are processed across the time direction. Each of the sky models is simulated with a JVLAC HR-MS of 300 timeslots of width $\Delta t^{hrms}s$ at 1.4GHz. The value of Δt^{hrms} is chosen within the range $[0.5, 5.5]$. The simulation considered 1 channel of width $\Delta \nu^{hrms} = 125kHz$ and the results of the process was saved into a LR-MS of $100 \times \Delta t^{hrms}s$ synthesis, with a $100 \times \Delta t^{hrms}s$ integration time at 1.4GHz, with 1 channel of width $125kHz$. We considered $t_{start}^{hrms} = 0s$, $t_{start}^{lrms} = 100 \times \Delta t^{hrms}s$, $t_{end}^{hrms} = 300 \times \Delta t^{hrms}s$ and $t_{end}^{lrms} = 200 \times \Delta t^{hrms}s$.

The source flux density as a function of the integration time width is depicted in Figures ?? to ??. The study compared the maximal integration time width of the boxcar time averaging (Bx-avg-0 × −) with that of 2° FoV BDWF time averaging (Bl-sinc-100 × −, Bl-sinc-100 × −, Bl-J$_0$-0 × −, Bl-J$_0$-100 × −, Bl-BW-0 × − and Bl-BW-100 × −). The results show that:

• "*Regime 1*": The BDWF time averaging maintained the flux density of the source for large integration time. However, it is seen on the Figures that the curve of sources in the regime are closer to the affine line y=1 when the integration time increases.

• "*Regime 2*" and "*Regime 3*": It appears on the Figures that Boxcar time averaging attenuated the sources in these regime compare to BDWF time averaging without overlap while overlap sinc and J$_0$ time averaging attenuated the source in a variable integration time range compare to the boxcar (see black and red curve in Figure ?? and ??). The integration time range where the source is attenuated depends on the WF (sinc or J$_0$) and the overlap time bins which cause its to variate.

5.2.3 Time and frequency direction

Boxcar and BDWF averaging are processed across the time and the frequency direction. Each sky model is simulated with a JVLAC HR-MS of 300 timeslots and the integration time Δt^{hrms} is chosen within the range $[0.5, 5.5]s$ at 1.4GHz. The simulation considered 150 channels of width $\Delta \nu^{hrms}$ chosen within the range $[125., 750.]kHz$. The result of the process was saved into a LR-MS of $100 \times \Delta t^{hrms}s$ synthesis, with $100 \times \Delta t^{hrms}$ integration time and 1 channel of width $50 \times \Delta \nu^{hrms} kHz$. We considered, $t_{start}^{hrms} = 0s$, $t_{start}^{lrms} = 100 \times \Delta t^{hrms}s$, $t_{end}^{hrms} = 300 \times \Delta t^{hrms}s$ and $t_{end}^{lrms} = 200 \times \Delta t^{hrms}s$, $\nu_{start}^{hrms} = \Delta \nu^{hrms} kHz$, $\nu_{start}^{lrms} = \nu_{start}^{hrms} \times 50kHz$, $\nu_{end}^{hrms} = \nu_{start}^{hrms} \times 150kHz$ and $\nu_{end}^{lrms} = \nu_{start}^{hrms} \times 100kHz$.

The source flux density as a function of the integration time and channels width is depicted in Figures ?? to ??. The study compared the maximal integration time and channels width of two dimensional boxcar averaging (Bx-avg-0 × 0) with that of 2° FoV of two BDWF averaging (Bl-sinc-100 × 50, Bl-sinc-100 × 50, Bl-J$_0$-0 × 50, Bl-J$_0$-100 × 50, Bl-BW-0 × 50 and Bl-BW-100 × 50). The results show that:

• "*Regime 1*": BDWF averaging maintained the flux density of the source for large integration time and wide channels widths. However, it is seen on the Figures that the curve of sources in the regime are closer to the affine line y=1 when the integration time and channels width increases.

• "*Regime 2*" and "*Regime 3*": It appear that the curves of the overlaps sinc and J$_0$ are below the one of boxcar averaging for all integrations and channels widths. The results suggest that one do not need to worry about sidelobes confusion for large integration time and wide channels widths.

Discussion: The study describes the performances of the WF under consideration. As presented, the sinc performed well in *Regime 1* compared to the J$_0$ while the J$_0$ performed well in *Regime 2* and *Regime 3* compared to the sinc. The great potential of the two dimensional overlap sinc and the J$_0$ is the accurate performance in the three regimes. An interesting improvement of this study is the possibility of using WF capable of providing higher dynamic range images. This

advanced techniques is required by the radio telescopes of the future.

6 REALISTIC FIELD SYNTHESIS: 3C147

OMS: commenting out the huge number of figures for now to save compilation time. Will uncomment them one by one as I absorb them into the main text.

7 CONCLUSIONS

The goal of this paper was threefold. The original motivation behind the work presented in this
paper was to **** windowing functions***
The second objective was to study ****first algorithm data compression***
The final objective was to ****second algorithm data compression and out field suppression***
Drawback and futures works*** drawback and futures works*****

REFERENCES

Bregman J. D., 2012, PhD thesis, Ph. D. Thesis, University of Groningen, Groningen, The Netherlands, pp81-82

Cornwell T., Braun R., Briggs D. S., 1999, in Synthesis Imaging in Radio Astronomy II Vol. 180, Deconvolution. p. 151

Delsarte P., Janssen A., Vries L., 1985, SIAM Journal on Applied Mathematics, 45, 641

Landau H., Pollak H., 1978, Bell Syst. Tech. J, 57, 1371

Smirnov O. M., 2011, arXiv preprint arXiv:1101.1764

Thompson A. R., 1999, in Synthesis Imaging in Radio Astronomy II Vol. 180, Fundamentals of radio interferometry. p. 11

Thompson A. R., Moran J. M., Swenson, Jr. G. W., 2001, Interferometry and Synthesis in Radio Astronomy, 2 edn. Wiley, New York

APPENDIX A: DERIVATION OF COMPLEX MATRICES

The complex matrices used in section 4.2 are explicitly derived in this appendix. In Eq.??, the matrices $\mathbf{C}_{(t,\nu)}^{block}$ and

$\mathbf{W}_{pq,(t,\nu)}^{block}$ are blocks diagonals both of size $(4 n_t n_\nu) \times (4 n_t n_\nu)$ explicitly expressed as follow:

$$\mathbf{C}_{(t,\nu)}^{block} = \begin{bmatrix} \mathbf{c}_{(t,\nu)} & 0 & 0 & 0 \\ 0 & \mathbf{c}_{(t,\nu)} & 0 & 0 \\ 0 & 0 & \mathbf{c}_{(t,\nu)} & 0 \\ 0 & 0 & 0 & \mathbf{c}_{(t,\nu)} \end{bmatrix}$$

$$\mathbf{W}_{pq,(t,\nu)}^{block} = \begin{bmatrix} \mathbf{W}_{pq,(t,\nu)} & 0 & 0 & 0 \\ 0 & \mathbf{W}_{pq,(t,\nu)} & 0 & 0 \\ 0 & 0 & \mathbf{W}_{pq,(t,\nu)} & 0 \\ 0 & 0 & 0 & \mathbf{W}_{pq,(t,\nu)} \end{bmatrix}$$

In Eq.??, the matrices $\mathbf{C}_{pq,(t,\nu)}^{block,n_{block}}$ and $\mathbf{W}_{pq,(t,\nu)}^{block,n_{block}}$ are blocks diagonals both of size $(4 N_v^{pq} n_t n_\nu) \times (4 N_v^{pq} n_t n_\nu)$, and the sampled visibilities $\mathbf{V}_{pq,(t,\nu)}^{samp,nblock}$ is a one row matrix of size $(N_v^{pq} 4 n_t n_\nu) \times (4 n_t n_\nu)$ made of $\mathbf{V}_{pq,(t,\nu)}^{samp}$. These matrices are explicitly expressed as follow:

$$\mathbf{C}_{pq,(t,\nu)}^{block,n_{block}} = \begin{bmatrix} \mathbf{C}_{pq,(t,\nu)}^{block} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \mathbf{C}_{pq,(t,\nu)}^{block} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & \mathbf{C}_{pq,(t,\nu)}^{block} \end{bmatrix}$$

$$\mathbf{W}_{pq,(t,\nu)}^{block,n_{block}} = \begin{bmatrix} \mathbf{W}_{pq,(t,\nu)}^{block} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \mathbf{W}_{pq,(t,\nu)}^{block} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & \mathbf{W}_{pq,(t,\nu)}^{block} \end{bmatrix}$$

$$\mathbf{V}_{pq,(t,\nu)}^{samp,n_{block}} = \left(\mathbf{V}_{pq,(t,\nu)}^{samp,1}, \ldots, \mathbf{V}_{pq,(t,\nu)}^{samp,k}, \ldots, \mathbf{V}_{pq,(t,\nu)}^{samp,N_v^{pq}} \right)^\dagger .$$

In Eq.??, the matrix $\mathbf{B}$ of size $(N_v 4 n_t n_\nu) \times (4 n_t n_\nu)$ is defined as follow:

$$\mathbf{B} = \begin{bmatrix} \mathbf{C}_{01,(t,\nu)}^{block,n_{block}} \cdot \mathbf{W}_{01,(t,\nu)}^{block,n_{block}} \\ \vdots \\ \mathbf{C}_{ik,(t,\nu)}^{block,n_{block}} \cdot \mathbf{W}_{ik,(t,\nu)}^{block,n_{block}} \\ \vdots \\ \mathbf{C}_{jl,(t,\nu)}^{block,n_{block}} \cdot \mathbf{W}_{jl,(t,\nu)}^{block,n_{block}} \end{bmatrix}$$

This paper has been typeset from a TeX/ LaTeX file prepared by the author.