Square Kilometre Array
Expert Panel on Radio Frequency Interference

Report on the Strengths and Weaknesses of the Current Radio Frequency Interference Environment as Measured at the SKA Candidate Sites

November 10th, 2011

1. Background[1]

In accordance with its statement of work and terms of reference, the Expert Panel on Assessment on RFI[2] was charged with assessing the strengths and weaknesses of the current radio frequency environment as measured at each candidate SKA site. The panel was also to identify potential site-specific concerns that arise from these measurements that impact the ability to conduct radio astronomy as set out in the Science Requirements for the SKA.

The results of the panel's assessment are contained in this report, prepared for the SKA Science and Engineering Committee, making a direct comparison between the measured RFI environments of the two candidate sites. This report highlights any perceived shortcomings in the data taken and any measurements that suggest a limitation of the sites' suitability to host the SKA.

2. Measurement Data and their Limitations

2.a) Measurement Mode Summary

The details of the SKA site spectrum monitoring, measurement program, and data processing are described elsewhere.[3] In summary, the data were acquired in a variety of modes (see below) at the two candidate core sites and at four candidate remote sites for each of the candidate core sites. The data were acquired across nine separate bands from 70 MHz to 2 GHz, in four pointings each band (in 90° azimuthal increments directed at the horizon), and two polarizations (horizontal and vertical) each pointing. The modes and acquisition parameters for each pointing and each polarization were as follows:

- *Fast Scanning mode (FS)* Core sites, 10 second integration
- *High Sensitivity mode (HS)* Core sites, 2 hour integration
- *Rural mode (RM)* Remote sites, 10 minute integration
- *Max Hold mode (MH)* Remote sites, 90 x 0.1 s integrations, max hold

[1] Charge and deliverables are from the "SKA Statement of Work and Terms of Reference: Expert Panel on Radio Frequency Interference (RFI)"
[2] Hereafter referred to as "the panel."
[3] "SKA site spectrum monitoring, measurement program and data processing," SKA report.

Additional data were acquired in a transient mode (TR) at the core sites, which provided high time resolution (down to 1 microsecond). However, there were technical and operational difficulties in acquiring TR mode data, which made the usability of some of the data unlikely. In addition, the panel felt that the total amount of time acquired in transient mode (0.5 s), and the fraction of time the data were acquired (0.5 s out of 9 hours = 0.0015% fractional coverage), made their usefulness for RFI studies very limited.

The modes and location of the data acquisition are summarized in the following table:

Mode	**Abbrev.**	**Core Sites**	**Remote Sites**
Fast Scanning	FS	10 s integration	-
High Sensitivity	HS	2 h integration	-
Rural	RM	-	10 m integration
Max Hold	MH	-	90 x 0.1 s integration, max hold
Transient	TR	Data Not Used	-

Table 1: Summary of the data acquisition modes used at the core and remote sites.

2.b) Antenna Coverage

The data were obtained with a Rohde & Schwarz HL033 log-periodic antenna[4] with a specified frequency range (<2:1 SWR) of 80 MHz to 2 GHz. The nominal antenna gain varies between roughly 6 and 7 dBi across the band. The E- and H-plane beamwidths (3 dB down, FWHM) of the antenna across the band are approximately 100° and 180°, respectively, and the response is down 10 dB at 80° off the pointing direction (E-plane). When mounted for vertical polarization, the antenna response is therefore below 4 dBi above 50° elevation, and below -3 dBi overhead. When mounted for horizontal polarization, the overhead response is good. The antenna provided good coverage of the four 90° pointing sectors (along the azimuthal direction) in both horizontal and vertical mode. When mounted for vertical polarization, there is substantial overlap with adjacent pointings (approximately 50% overlap with both adjacent directions).

The general response of the antenna after accounting for all four azimuthal pointing directions is summarized in the following table:

[4] http://www2.rohde-schwarz.com/file_12435/HL033_brief_e.pdf

Antenna Mounting Position	Horizon Coverage (to -3 dB point)	Elevation Coverage (to -3 dB point)
Vertical Polarization	Excellent 50% overlap between pointings	To 50° elevation (~20% of sky not covered)
Horizontal Polarization	Excellent 5% overlap between pointings	Very Good (zenith coverage; no elev. overlap)

Table 2: Azimuth and elevation coverage of antenna assuming four azimuthal pointings at 90° increments.

2.c) Assessment of the Measurement Data

2.c.i) Cyclic Coverage

RFI is often cyclic on a variety of time scales. Over the course of the day, RFI from business sources may peak during the work day, while residential RFI increases in the mornings and evenings. Interference from aviations sources will ebb and flow with daily flight patterns. Anomalous propagation due to meteor scatter will peak between midnight and noon, as the local zenith points toward the direction of the Earth's revolution about the Sun. Over a week's cycle, commercial RFI may peak on weekdays and residential RFI may peak on weekends. Over seasonal cycles, the amount of interference coming from the horizon direction may vary as deciduous foliage cover comes and goes, and seasonal weather patterns may impact foliage water content, which can have a substantial impact on RFI absorption at high VHF (300 MHz) and above. The amount of outdoor activity as a function of season may also have an impact on received RFI. Major meteor showers also occur at specific times of the year, which will temporarily enhance meteor scatter propagation.

In general, there may be a wide range of cyclic effects on the amount of RFI received at a specific site. However, the measurement data for the core and remote sites were obtained at times and durations that were convenient to the project, and are not designed to assess any cyclic variability in the RFI data. The panel's assessment of the RFI levels at the sites is therefore based on a set of data that does not sample all cyclical facets of the RFI environment, which may be a significant limiting factor in drawing conclusions from this report.

2.c.ii) Anomalous Propagation

Anomalous propagation effects – principally, meteor scatter and sporadic E propagation – can create temporary but substantial enhancements in signal levels from distant sources of RFI. These effects are typically most prevalent in the low- to mid-VHF range. Anecdotal data (unrelated to the SKA sites) demonstrate that sporadic E propagation in particular can create substantial RFI especially in TV and FM broadcast bands, with strong signals being observed

from very powerful broadcast stations hundreds of km away. Sporadic E propagation can occur on a regular basis (often during summer months), and last for an hour or more at a time. While meteor burst propagation is shorter-lived (typically a few seconds to approximately one minute), the constant bombardment of the Earth by meteors makes this mode occur very frequently, and in fact some commercial communications systems rely upon it.[5]

Anomalous propagation is especially important in determining the level of RFI as sites that are located far from population centers. Since anomalous propagation may create the reception of signals from hundreds of km away, it may be the ultimate limiting factor in some bands. However, as pointed out in the previous section, the limited time intervals over which measurement data were obtained do not allow an assessment of the frequency and strength of anomalous propagation at either of the candidate sites.

2.c.iii) Aeronautical and Satellite RFI

Regardless of how remote a site is, it will still be subject to RFI from aeronautical and satellite sources.

A flight at cruising altitude (10 km) will have a radio horizon approaching 400 km, and by volumetric arguments, assuming random flight patterns, a greater number of flights will be at farther distances from the site. Therefore, even regulatory restrictions on overflights of the site will have relatively little effect, unless the diameter of the restricted zone approaches 800 km, which seems unlikely. Aeronautical interference (air-to-ground) is particularly prevalent in the 118 – 137 and 960 – 1215 MHz ranges, with the former band having potentially great impact on EoR observations.

Overflights by satellites cannot be restricted. There are several satellite downlink bands within the 80 – 2,000 MHz range. For example, Orbcomm uses 137-138 MHz, various weather satellites also use 137-138 MHz, some Russian navigation satellites use 149.9-150.05 MHz, GPS uses 1260MHz and 1575.42 MHz, in addition Galileo will use 1164-1215 MHz and 1278.75 MHz, amateur radio satellites use 144 and 430 MHz, Iridium uses 1618.25-1626.5 MHz[6] , the 2.5 kW U.S. military satellites operate at 243-260 MHz. There are also WorldSpace (AFRISTAR or ASIASTAR at 1467-1492 MHz) lines visible on some of the data. INMARSAT operates on 1525 – 1559 MHz and GLONASS on 1242.94-1248.19 MHz and 1598.06 – 1604.81MHz.

While the RFI measurement data have fairly good sensitivity to satellite signals when the antenna is mounted in horizontal polarization mode, in vertical polarization mode, about 20%

[5] Schanker, Jacob, Meteor Burst Communications, (Boston: Artech House), 1990.

[6] IRIDIUM downlinks are also known to cause radio interference at 1610.6-1613.8 MHz and on frequencies up to 1630 MHz. The sensitivity of the surveys was however not sufficient to detect the out-of-band IRIDIUM interference, but it will affect future SKA operations.

of the sky is not covered. In addition, since most of the satellite RFI in the bands above is from LEO satellites (GPS and Galileo are the exception; they are MEOs), the overflights last at most about 10 minutes. This is too long to be adequately sampled in FS mode. In addition, satellites will come and go in directions that the antenna was not pointing at that particular moment. Even with a direct overflight aligned exactly in the cardinal direction the antenna is pointing, the total duration of time in the beam is ~5 minutes, which is only about 4% of the HS integration period, creating a dilution of its impact on the RFI measurements.

2.c.iv) Transients

As mentioned previously, transient measurements (TR mode) with 1 microsecond resolution over 0.5 second burst-mode acquisition were obtained at the core sites. However, data acquisition and reduction were apparently somewhat problematic. In addition, the panel believes that the low duty cycle (0.0015%) would render analysis of the TR data somewhat meaningless. Ultimately, however, the project should continue its attempt to obtain good TR mode data, since they are relevant to pulsar mode observing, and may provide insight into bursty types of RFI (spark plugs and electrical motors) that is not available from other measurement modes.

2.c.v) Snapshot in Time

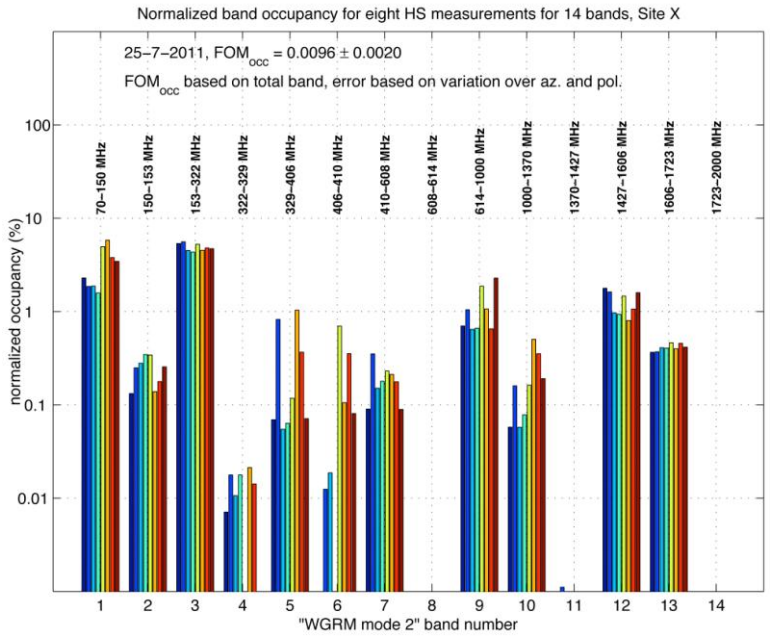
The RFI measurements which were used to compare and contrast the sites have been obtained at one epoch (the present). The data are therefore not necessarily representative of the relative RFI environments at the sites in the coming years, when the SKA would be built, or 10 – 50 years in the future, when the SKA may be operating. Many factors over the coming decades could create an imbalance in the relative merits of the sites, and for the purpose of this report, these factors are completely unpredictable.

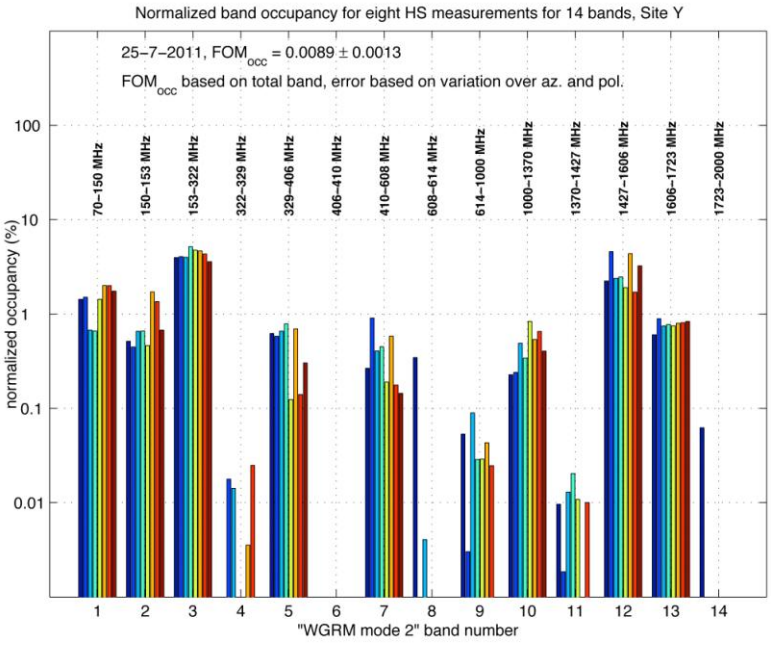
2.c.vi) Remote Site Coverage

The RM and MH data were obtained at four out of 25 candidate remote sites for each candidate core site. While the panel can compare and contrast the remote site data that were supplied, we have no ability to compare how the four remote sites compare with the other 21 remote sites (for each candidate core site) for which we have no data. We understand that the SKA project office based their choice of each of the sets of four sites in a manner that should make the sites representative and intercomparable between the two sets, but the panel has no independent way to validate the choices.

3. Figures of Merit

There are many ways to analyze the RFI data, but the most basic criterion is the amount of interference-free spectrum that is available over the measurement range. The SKA project office analyzed the core site data and derived statistics regarding channels in which interference was flagged. The results are shown in the following figures. The panel could not discern any significant difference between the two sites.





3.1 Alternative Figures of Merit

3.1.1 Basic Considerations

The information about the radio environment of the core and remote sites was provided in the form of radio spectra having 70438 spectral channels covering the range of 70 MHz to 2000 MHz. Typical spectra showed more than 2000 interference lines of varying strength, location and origin. The large amount of detailed information suggest that a method of deriving an objective quality criterion should be found and applied to all data, preferably as a programmable algorithm that takes the spectral data as numerical input and provides a unique scalar value (=FOM) as the result. The FOM should be sensitive to variations in the data, take into account all available information and reflect the actual radio astronomical data acquisition process as much as possible.

The radio astronomical radio flux sensitivity ΔS depends on the spectral noise power P_{sys} of the receiver, which contains the sum of all unwanted intrinsic and extrinsic (including RFI) signal power, the receiver bandwidth $\Delta\nu$, and the integration time Δt of the measurements. For white noise signals it is given by the radiometer formula[1,2]:

$$\Delta S\left(P_{sys},\Delta\nu,\Delta t\right)=\frac{P_{sys}}{A_{eff}}\cdot\Delta\nu^{\frac{-1}{2}}\cdot\Delta t^{\frac{-1}{2}}$$

with A_{eff} being the effective collecting area of the receiver. Hence the strength, bandwidth and duration of interference together determine the achievable sensitivity in a given environment. In an ideal situation one would have no interference.

Radio astronomy, like any other empirical science aims to obtain information by measurements. In the case of radio astronomy one measures the noise level of radio sources, which can be seen as an information transfer from source to receiver (information channel). Information flow can be quantified in bits/second and the channel capacity C (in Bit/s) as a function of modulation bandwidth B_m , signal power S and noise power N is given by [3]

$$C=B_m\cdot\log_2\left(1+\frac{S}{N}\right)$$

with the modulation bandwidth B_m being equal to Δt^{-1} for radiometer measurements. For a radio astronomical source with spectral power P_{src} at a distance d, we obtain

$$I_s=\log_2\left(1+\frac{A_{eff}}{4\cdot\pi\cdot d^2}\cdot\frac{P_{src}}{P_{sys}}\cdot\Delta\nu^{\frac{1}{2}}\cdot\Delta t^{\frac{1}{2}}\right)$$

bits of information from one measurement having a bandwidth $\Delta\nu$ lasting a time Δt. Under these conditions, the detecting range d_{lim} for sources is limited and proportional to $\sqrt{\frac{\Delta\nu^{\frac{1}{2}}\Delta t^{\frac{1}{2}}}{P_{sys}}}$.

For an idealised sky populated with similar sources of uniform space density, the number of detectable sources will be proportional to $P_{sys}^{\frac{-3}{2}}\Delta\nu^{\frac{3}{4}}\Delta t^{\frac{3}{4}}$ and so will be the total amount of information obtainable from the sky by adding the contributions of all detectable sources[7]:

$$I_{sky} = \eta P_{sys}^{\frac{-3}{2}}\Delta\nu^{\frac{3}{4}}\Delta t^{\frac{3}{4}}$$

with η relating to the source properties and the antenna size. So far we modelled only the maximum bandwidth continuum measurement. If one were to split the total available bandwidth into **n** channels, then the amount of information grows proportional to $\mathbf{n^{1/4}}$. Although spectroscopy will detect fewer flat spectrum sources, the amount of information in a spectrum is always greater[8]

3.1.2 First Order FOM

Hence an appropriate figure of merit **q** would be the ratio of information obtainable from the sky with interference I_{rfi} to I_{sky} of the undisturbed sky.
Let the rfi spectral power density be P_{rfi}, the duration of rfi be $\Delta t_{rfi} \leq \Delta t$ and the occupied bandwidth be $\Delta\nu_{rfi} \leq \Delta\nu$ then the noise background P_S of a receiver at a chosen site S is given by $P_S = P_{sys} + P_{rfi}\cdot\frac{\Delta\nu_{rfi}}{\Delta\nu}\cdot\frac{\Delta t_{rfi}}{\Delta t}$. Introducing three variables that characterise the relative duration of rfi $\tau = \frac{\Delta t_{rfi}}{\Delta t}$, the relative spectral coverage $\sigma = \frac{\Delta\nu_{rfi}}{\Delta\nu}$ and the relative rfi signal level $\beta = \frac{P_{rfi}}{P_{sys}}$ results in the figure of merit q in a particular band $\Delta\nu$ of a given site s being expressed as

$$q(\beta,\sigma,\tau) = \left(1+\beta\ \sigma\tau\right)^{-\frac{3}{2}}$$

Here we use only the dependence of the radiometric channel information on the average unwanted noise power.
The variables β,τ,σ will depend on frequency and one way of getting a figure of merit for a site would be to calculate the mean value of q over all nine sub-bands in the spectrum covering the whole frequency range:

---

[7] The integration of $4\pi r^2 I_s(r)$ up to the detection limit yields a constant factor and results in an average contribution of 1.62 bits per source.

[8] Orders of magnitude: For an assumed sky with a density of 10^{-9} sources per pc^3, each emitting 10^{26} W over 10 GHz the undisturbed sky may yield 440 MBit for 2 GHz bandwidth, 2000 s integration time and a system temperature of 300 K. Splitting the band into 70438 channels yields up to 7.17 GBit of information under the same conditions.

$$Q_1 = \frac{1}{N}\sum_{i=1}^{N}\left(1+\beta_i\ \sigma_i\tau_i\right)^{-\frac{3}{2}}$$

An interference strength of β=20 (13 dB) , lasting the full time and spanning the whole channel (τ,σ=1) will reduce Q_1 to 0.01. Hence any interference with stronger levels will be treated the same: as a total loss of the channel.

3.1.3 Additional Losses by Second-Order Intermodulation and Spillover; Second Order FOM

Real correlators or spectrometers will always suffer from non-linearities and exhibit a limited suppression of signals from adjacent spectral channels. The power P_n of n-th order intermodulation (IM) products as a function of input power P of two signals is given as

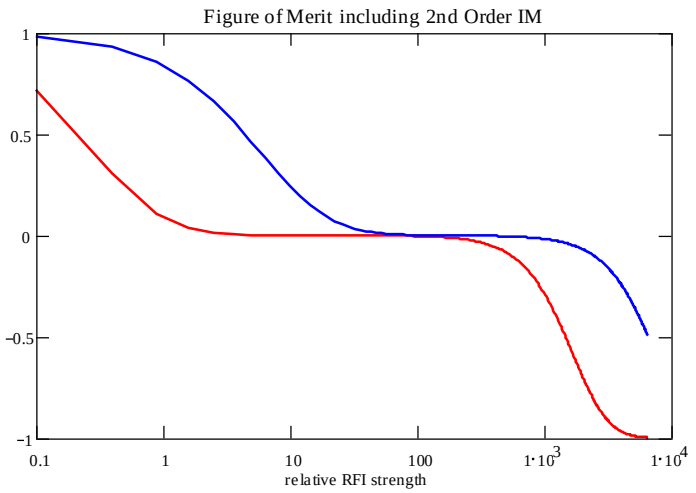
$$P_n = \frac{P^n}{IIP_n^{n-1}}$$

where IIP_n is the n-th order input intercept point.[9] The IM products can raise the background outside the actual bands where interference is present and will lead to a decrease of sensitivity there. Although the IM power increases more steeply with higher orders, the second-order intercept point is usually the lowest and most important and we therefore consider only second order IM effects. In this case, the affected bandwidth is similar to the primary interference bandwidth (two sidebands for two frequencies) . The intercept point can be expressed as a fraction $\kappa = P_{sys}/IIP_2$ of the system noise level, with $\kappa := 10^{-7}$ the IIP_2 is 70 dB above the noise level, and the IM interference power is then given by $\beta^2 \cdot \kappa$. Spillover of strong signals into adjacent spectral channels will depend on the shape of the channel spectral response and can to some degree be approximated in the same manner. The sensitivity *outside* the band may be reduced in the case of weak interference, and can drop to effectively zero for strong interference. In the strong interference case, the total amount of obtainable information over all bands is therefore further reduced compared to the low level interference case. The quality factor (figure of merit) q_2 can now be written as

$$q_2(\beta,\sigma,\tau) = \left(1+\beta\ \sigma\tau\right)^{-\frac{3}{2}} + \left(\left(1+\kappa\ \beta^2\sigma\tau\right)^{-\frac{3}{2}} - 1\right)$$

with the second term being the contribution of strong IM or spillover signals.

[9] C.f. “Intermodulation Distortion Measurements,” Anritsu, http://downloadfile.anritsu.com/RefFiles/en-US/Services-Support/Downloads/Application-Notes/Application-Note/11410-00257a.pdf



The graph shows the information ratio q_2 (β,σ,τ) as a function of β for σ=1 (red) and σ=0.05 (blue)

The value of q_2 may even be negative and approach -1 in the case of extreme and permanent (τ=1) interference. This reflects the fact that additional bandwidth (= information) is lost by strong IM signals. Here we neglect the possibility, that neighbouring bands may also be affected by their own sources of interference.

Again taking the average over all N spectral channels will provide a second order FOM:

$$Q_2 = \frac{1}{N}\sum_{i=1}^{N}\left(\left(1+\beta_i\ \sigma_i\tau_i\right)^{-\frac{3}{2}} + \left(1+\kappa\beta_i{}^2\sigma_i\tau_i\right)^{-\frac{3}{2}}\right)$$

which takes account of the additional effects of strong signals which have already led to the total information loss in their respective interference channels.

3.1.4 The Limit of $\beta \to \infty$ and $\kappa \to 0$: Occupation statistics

All site measurements were made at sensitivity levels at least 20 dB above the spectrometry interference thresholds given in ITU-R RA. 769 [4] for Δt=2000 seconds, and one can assume that for long integrations with high sensitivity, any detection of rfi in site survey spectra would lead to the total loss of the information contributed by the affected channels hence $\beta \to \infty$. Furthermore, if we assume, that measuring equipment is ideally linear and spillover is negligible ($\kappa \to 0$), then Q_2 may be replaced by the relative number of unaffected channels:

$$Q_0 = 1 - \frac{1}{N}\sum_{i=1}^{N}\Theta(\beta_i - \beta_{det})$$

Here β_{det} is the site survey detection limit and $\Theta(\beta_i - \beta_{det})$ is supplied as the 'detect' flag in the case of HS and FS mode data.

As an example we select two FS measurements from each site and compare them by band, assuming full time interference (τ=1) and an IIP_2=70 dB.

		Site X			
Measurement Number	Band	Detection % σ	Background (dB) N	typ. Power (dB) I	max. separation
	MHz				
1	80-200	4	-225	-180	27
1	200-400	10	-220	-185	120
1	400-600	2,5	-215	-200	20
1	600-800	1	-212	-185	70
1	800-1000	2,5	-212	-160	140
1	1000-1200	4	-210	-180	50
1	1200-1400	0	-209	-195	150
1	1400-1600	10	-205	-185	230
1	1600-1800	5	-205	-180	90
1	1800-2000	0	-205	-205	290
3	80-200	3	-225	-175	30
3	200-400	10	-220	-185	120
3	400-600	5	-215	-180	42
3	600-800	1	-212	-180	70
3	800-1000	3	-212	-160	170
3	1000-1200	4,5	-210	-180	30
3	1200-1400	0,5	-209	-195	130
3	1400-1600	6	-205	-185	100
3	1600-1800	4	-205	-180	70
3	1800-2000	0	-205	-205	290

mean value of detections: 3,8

			Site Y			
Measurement Number	Band	clustering width	Detection % σ	background N	typ. Power I	max. separation
	MHz					
1	80-200	20	2,5	-225	-205	20
1	200-400	25	10	-220	-185	60
1	400-600	20	1,5	-215	-190	40
1	600-800	15	2	-212	-190	70
1	800-1000	30	1	-212	-190	130
1	1000-1200	30	6	-210	-180	50
1	1200-1400	20	0,5	-209	-195	130
1	1400-1600	40	15	-205	-180	180
1	1600-1800	10	5	-205	-180	80

1	1800-2000	0	0	-205	-205	280
3	80-200	20	2,5	-225	-185	20
3	200-400	30	10	-220	-185	60
3	400-600	20	8	-215	-180	70
3	600-800	60	2	-212	-205	90
3	800-1000	30	1	-212	-190	150
3	1000-1200	30	9	-210	-185	30
3	1200-1400	10	2,5	-209	-190	30
3	1400-1600	40	10	-205	-185	180
3	1600-1800	10	5	-205	-180	60
3	1800-2000	0	0	-205	-205	290

Mean value of detections 4,675

One can now calculate the FOMs:

Site X

$\beta = 10^{\frac{I - N}{10}}$ q_1 q_2

31622,7766	2,22022E-05	-9,11E-01
3162,27766	0,000176988	-1,33E-01
31,6227766	0,417362338	4,17E-01
501,187234	0,067839928	6,75E-02
158489,319	4,00798E-06	-9,98E-01
1000	0,003809116	-2,16E-03
25,1188643	1	1,00E+00
100	0,027410122	2,73E-02
316,227766	0,014507569	1,38E-02
1	1	1,00E+00
100000	6,08276E-06	-9,94E-01
3162,27766	0,000176988	-1,33E-01
3162,27766	0,000498239	-7,01E-02
1584,89319	0,014459106	1,07E-02
158489,319	3,04917E-06	-9,98E-01
1000	0,00320526	-3,51E-03
25,1188643	0,837388823	8,37E-01
100	0,053994925	5,39E-02
316,227766	0,019830948	1,92E-02
1	1	1,00E+00

	Q_1	Q_2
Mean value	0,223034785	0,01019909

Standard deviation	0,389981645	0,63796983

Site Y

$\beta = 10^{\frac{I-N}{10}}$	q_1	q_2
100	0,15272071	1,53E-01
3162,27766	0,00017699	-1,33E-01
316,227766	0,07265149	7,24E-02
158,489319	0,11744358	1,17E-01
158,489319	0,24062242	2,41E-01
1000	0,00209897	-6,83E-03
25,1188643	0,83738882	8,37E-01
316,227766	0,00296669	7,21E-04
316,227766	0,01450757	1,38E-02
1	1	1,00E+00
10000	0,00025147	-2,84E-01
3162,27766	0,00017699	-1,33E-01
3162,27766	0,00024706	-1,09E-01
5,01187234	0,86650359	8,67E-01
158,489319	0,24062242	2,41E-01
316,227766	0,00625374	4,91E-03
79,4328235	0,19382261	1,94E-01
100	0,02741012	2,73E-02
316,227766	0,01450757	1,38E-02
1	1	1,00E+00

	Q_1	Q_2
mean	0,23951864	0,20579209
Stddev.	0,36285781	0,3926892

Visual inspection of the measurements shows, that the sites are similar in respect of of Q_1, but site X has a slightly higher Q_0=1-3.68 compared to site Y with Q_0=1-4.65. On the other hand, The greater strength of interference lines at the site X causes Q_2 to be smaller than for site Y. However the visual estimates are not accurate themselves and the different bands show great variation, expressed by large standard deviations:

	combined Standarddeviations
Q_1	0,53268328
Q_2	0,74913971

Site averages of Q_1 and Q_2 agree within one standard deviation and the same is expected for the detection estimates Q_0. Visual inspections of complex spectra are by nature inaccurate, and

we find the example of the FS data inconclusive with regard to solving the question of which site is more suitable for the SKA.

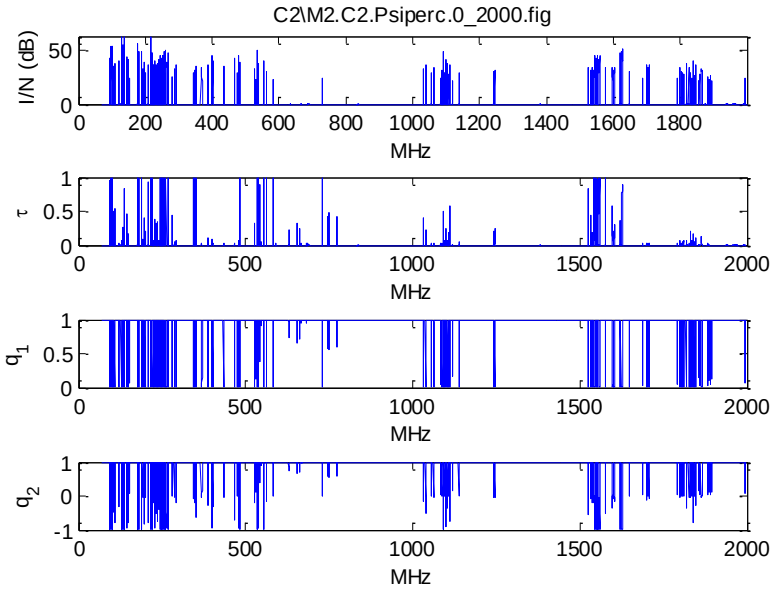
3.2 Analysis of HS Data

HS data was provided not only as image data like the FS data, but also in the form of MATLAB fig files which can be used not only to display and zoom into the graphs, but also to re-extract the spectra for additional processing and conversion into an ascii format.

3.2.1 FOM Analysis

In the first step, the fig files were opened and the data extracted. The occupancy files contained only two arrays one for frequency (x) and one for occupancy (y). We divided the occupancy by 100 and called the variable τ. The spectra were actually split into 9 fragments which had to be re-concatenated. The detection flag, 99 Percentile flux, 90 Perc. flux and median flux arrays were retrieved. The concatenated spectrum arrays had the same dimension and scale as the contiguous occupancy arrays from the occupancy fig files, an indication of the success of the extraction process and consistency of the input data.

The detection flag was rescaled to a logical value 0 or 1 and its complement used as the weights for the calculation of the noise background (=baseline), using the MATLAB smoothing spline function 'csaps' with the smooth parameter set to 1. As a result the background was retrieved without the rfi and with spline interpolation where rfi was detected. Then the background was subtracted from the signal giving I/N in dB. It worked very well as one can see from the top panel on the supplied example plots for the two sites. The I/N was converted to linear scale and the channel specific q_1 and q_2 values were computed using σ=1 as the I/N values were channel averages. An IIP2 of 70 dB above noise was arbitrarily chosen for q_2 calculations.



The graph shows an example of processing results for HS observation 2 on site Y. The interference strength $I/N=10\log_{10}(\beta)$, the channel occupancy τ and the q_1 and q_2 FOM are shown in successive panels.

As expected, the 99 percentile spectrum showed the greatest impact and so it was used for all further calculations.

The following table gives the mean values from all 70438 channels of detections, the occupancy (τ), and the FOM q_1 and q_2 per spectrum with sequence numbers increasing downwards.

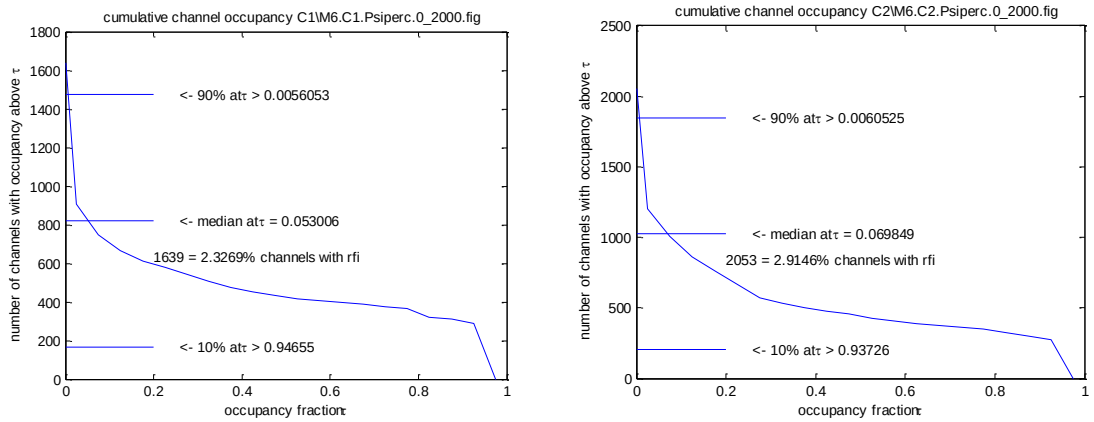
C1 M2-M16 Site X

	Detections	$<\tau>$	Q_1	Q_2
	0,026	0,0088	0,9783	0,9709
	0,0283	0,0102	0,9758	0,9669
	0,0233	0,072	0,9808	0,9749
	0,0212	0,007	0,9822	0,9769
	0,0377	0,0123	0,9693	0,9569
	0,0348	0,0106	0,9703	0,9597
	0,0294	0,009	0,9751	0,9666
	0,0362	0,0121	0,969	0,9552
Mean	0,0296125	0,01775	0,9751	0,966
stddev	0,00610794	0,02199039	0,00517577	0,00812703

C2 M2-M16 Site Y

	Detections	$<\tau>$	Q_1	Q_2
	0,0347	0,0075	0,9759	0,9703
	0,0314	0,0106	0,9731	0,9638
	0,0291	0,0081	0,9759	0,9697
	0,0291	0,0088	0,9758	0,9683
	0,0282	0,0086	0,9761	0,968
	0,0386	0,0113	0,9674	0,958
	0,0306	0,0079	0,9754	0,9681
	0,0296	0,0083	0,9749	0,9684
Mean	0,0314125	0,0088875	0,9743125	0,966825
Stddev	0,00353409	0,00134742	0,00295753	0,00405383

3.2.2 Comparison of Cumulative Channel Occupancies



The two graphs show typical channel occupancy distributions for the site X (left) and Y (right), with the ~68000 unoccupied channels not shown. The distributions were obtained from the occupancy data by computation of a 20 bin histogram for the occupancy fraction τ and the subsequent calculation of the cumulated distribution from the histogram. Outputs of Median (50%), 10% and 90% distribution values were also provided.

3.2.3 Results

None of the evaluated FOM quantities (detections, $<\tau>$, Q_1, Q_2) differed significantly between the two sites and their variation between individual measurements was also very small. There was also no discernible difference between the statistical distribution of channel occupancies.

3.3 Analysis of the MH datasets

MH fig files held contiguous arrays of spectra with max and median flux values, taken at four sample sites on each continent. No detection flag data was supplied. It was therefore computed by first calculating a mask for the baseline computations. This was done by numerically calculating the absolute values of the differentiated spectra, then detecting excursions larger than 10 standard deviations. The resulting mask array was blurred by adding circular shifts of ±1 and ±2 positions to it. That way, the spectral lines and their environment were given zero weight in the baseline calculation which used cubic smoothing spline interpolation. The baselines were subtracted from the data yielding the interference strength $I/N=10\log_{10}(\beta)$ as well as detection flags, the latter indicating where the signal exceeded 6 standard deviations. The algorithm could not suppress artefacts resulting from the sharp signal changes at the band edges. However these were only small in number (<10) and insignificant in comparison with the thousands of detected interference lines. Only the detection numbers and the q_2 FOM were computed, as these were most sensitive to interference. The IIP_2 setting was

chosen to be 2^{16} times the standard deviation of the spectra with baseline and spectral lines removed. Typical IIP_2 values were 35-40 dB which accounted for the higher noise levels of the MH spectra.

The following table summarises the results for MH data (all values are percentages)

occupation		$1\text{-}Q_2$ FOM	
median	max	median	max
1.3672	5.2159	1.1397	5.2847
1.6965	5.7554	1.3700	5.3942
2.5696	7.0544	2.2034	6.5010
2.8820	8.2654	2.6745	8.8007
mean			
2.1288	6.5727	1.8469	6.49515
stddev			
0.7138	1.3670	0.7164	1.63226

combined spectrum occupation Sites X:
med: 4.44 max: 15.20

occupation		$1\text{-}Q_2$ FOM	
median	max	median	max
0.9881	3.8843	0.7626	3.3292
0.5480	3.0907	0.4230	2.6406
2.1551	10.6732	1.6319	11.409
2.0103	6.5476	1.5442	5.2986
mean			
1.230	5.883	0.939	5.793
stddev			
0.830	4.168	0.623	4.876

combined spectrum occupation Sites Y:
med: 3.70 max: 17.24

(The combined occupation was obtained by a logical or of all four detection masks.)

3.3.1 Results

Two of the sites from Y show much better occupation and Q_2 values than the best sites of X and the median averages differ by about 1 sigma. However, site Y shows an extreme variability with measurement 3 having an extraordinary amount of stong rfi. The equipment went out of the linear realm in some parts which may have affected the results. The reason is unclear and one cannot say if that was the result of equipment failures (self interference), an anomalous propagation event, how often it happens or if it may occur on the other site too.

3.4 Analysis of RM Data

The analysis was carried out on the calibrated RM data in the same manner as for the MH data, but this time only the 95 percentile fluxes were evaluated because they showed the most number of rfi detections. The table below shows the results:

Remote Sites X

detect	$1-Q_2$
1.1826	2.0986
1.2905	2.3167
1.5418	2.7184
1.6412	2.9724
mean	
1.414	2.527
stddev	
0.2135	0.393

combined occupation 3.36 %

Remote Sites **Y**

detect	$1-Q_2$
1.0534	1.6830
0.7524	1.1348
1.1414	1.8776
1.1698	1.8771
mean	
1.0292	1.6431
stddev	
0.1911	0.3510

combined occupation 2.47 %

(The combined occupation was obtained by a logical ‘or’ of all four detection masks.)

3.4.1 Results

In this case detections and Q_2 values differ by about two standard deviations, clearly favouring the 'Y' sites. If one were to remove bands with common satellite allocations, one may expect the difference increase even more. The differences within X and Y sets of measurements are not significant.

4. Summary & Comments

It was hard to see any decisive difference in all the data from the core sites. In fact all conceivable measures agreed within one standard deviation taking individual measurements/pointings as separate statistical samples. However this is not the case for the remote sites. Using the calibrated data given in i.e. and concentrating on the 95 percentile fluxes one finds different percentages of detections favouring remote sites Y. But one ought to bear in mind, that the data provided only a spot check on a small selection of remote sites, even though selected remote sites were better at Y, one cannot know if that will be the rule and persist in the future.

Site X showed GSM 900 signals (920-960 MHz) which were absent in site Y. However site Y showed sporadic GSM 1800 emissions while the upper end of the spectrum (>1700 MHz) was clean for site X. WorldSpace satellite lines were seen on Spectra from site X, but were absent at site Y.

The peak strength of interference below 400 MHz was about 10 dB higher at X than at Y, indicating shorter distances to the interfering sources which would be consistent with the fact that remote sites for Y appeared to show less interference compared to X. As a result, one may tentatively conclude, that Y is less densely populated with interference sources than Y. This may speak in favour of Y, but the differences are too small compared to the known uncertainties and insufficient as a basis of a clear decision in favour of one or the other site.

5. References:

[1] ITU Handbook Radio Astronomy Second Edition, Radio Communication Bureau, Geneva 2003
[2]ECC Document SE40(11)025 'Scaling of Interference limits', http://www.craf.eu/CRAF-11-02.pdf
[3]C. Shannon, PROCEEDINGS OF THE IEEE, VOL. 86, NO. 2, FEBRUARY 1998
[4]Recommendation ITU-R RA. 769-2, Radio Communication Bureau, Geneva 2003