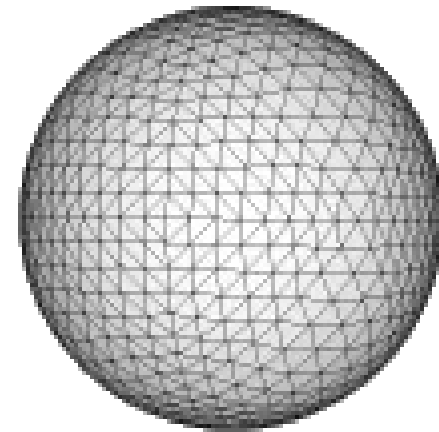
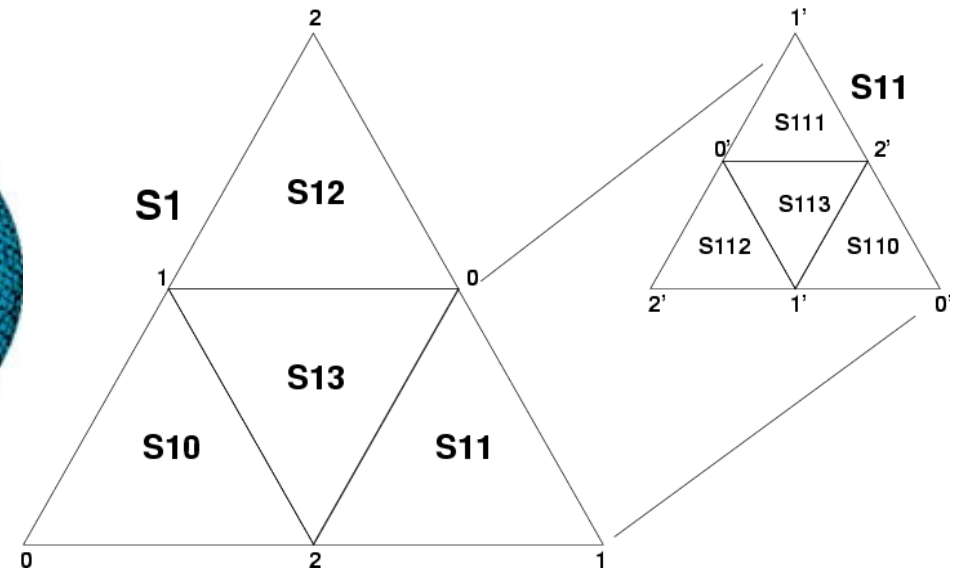
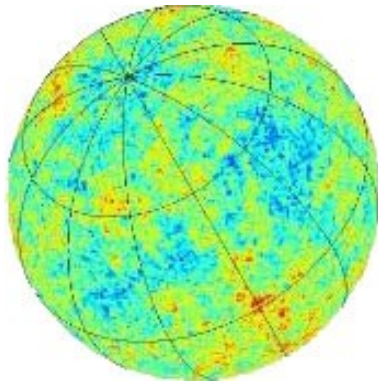
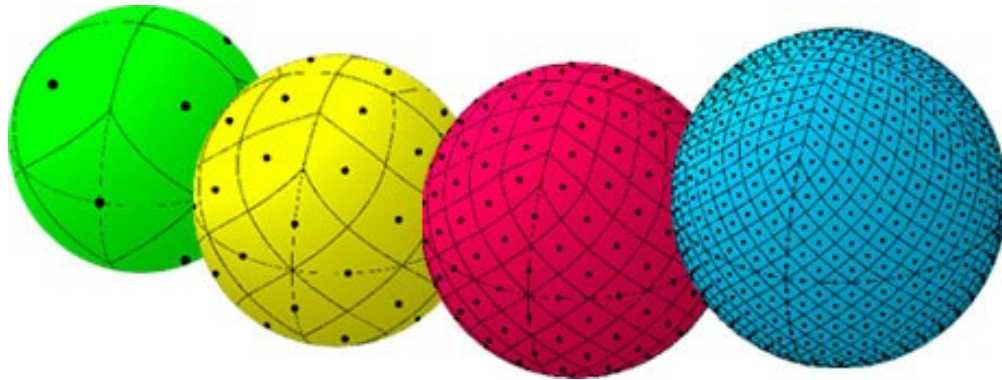


Gestione automatica della pixelizzazione del cielo per DB MySQL: HEALPix e HTM

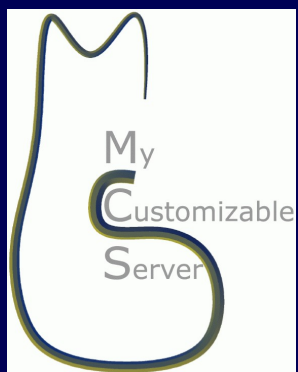


Sommario

- MCS
- Tassellizzazione della sfera e indicizzazione HTM e HEALPix
- Database relazionale: MySQL
- DIF – MCS:
 - Situazione attuale
 - Cosa c'è da fare nel futuro prossimo e non
- Demo in ambiente IDL
- (Le esigenze degli utenti)

Si rimanda ai seminari di G. Calderone per gli aspetti riguardanti l'uso dei DB.

ross.iasfbo.inaf.it/seminariDB/



MCS: My Customizable Server

ross.iasfbo.inaf.it/MCS/

A flexible resource for astronomical projects

A cosa serve:

- Applicazioni multi-thread
- Applicazioni di rete (via TCP)
- Applicazioni per database (MySQL)
- Information servers
- Accesso a tabelle VOTable e FITS

Inoltre:

- Interfacce client verso (quasi) tutti i linguaggi
- Sistema di privilegi a livello di record per tabelle
MySQL $\Rightarrow$ **MyRO**

- Sistema di indicizzazione automatico di tabelle con coordinate sferiche $\Rightarrow$ **DIF**

DBMS: Data Base Management System

I DBMS rendono più semplice:

- Archiviazione
- Accesso
- Condivisione
- Protezione dei dati (di qualsiasi tipo!)

Ma richiedono:

- Imparare ad usare un nuovo linguaggio (SQL)
 - Va seguita la logica dei database
-
-



MySQL

www.mysql.com

The world's most popular open source database

- Open Source
 - Alte performance, affidabilità e facilità di installazione e uso
 - Funziona su più di 20 piattaforme incluso Linux, Windows, OS/X, HP-UX, AIX, Netware ...
 - Flessibile e configurabile grazie a Views, Triggers, Plugins, etc.
 - Database scelto per la nuova generazione di applicazioni costruita sulla cosiddetta *LAMP stack* (Linux, Apache, MySQL, PHP / Perl / Python)
-
-

Indicizzare tabelle di DB

1-d				2-d					
Col ₁	Col ₂	...	Col _N	Col ₁	Col ₂	...	Col _N	RA	Dec
V ₁	...	...	...	V ₁	...	...	...	α ₁	β ₁
V ₂	...	...	...	V ₂	...	...	...	α ₂	β ₂
...	...	...	...	...	...	...	...	...	...
V _N	...	...	...	V _N	...	...	...	α _N	β _N

Catalogo		Righe	Colonne
Ottico:	USNO-B	1,045,913,669	30
Infrarosso:	2MASS	470,992,970	60
Raggi X:	1XMM	56,711	379

USNO-B and 2MASS distribuiti come file separati, ognuno copre 0.1° in declinazione, ordinati in RA nella fetta (*indici non univoci*).

Indicizzare tabelle di DB



B-tree:

In computer science, a B-tree is a tree data structure that keeps data sorted and allows searches, insertions, and deletions in logarithmic amortized time. It is most commonly used in databases and filesystems.

R-tree:

R-trees are tree data structures that are similar to B-trees, but are used for spatial access methods i.e., for indexing multi-dimensional information; for example, the (X, Y) coordinates of geographical data. A common real-world usage for an R-tree might be: "Find all museums within 2 miles of my current location".

Indicizzare tabelle di DB

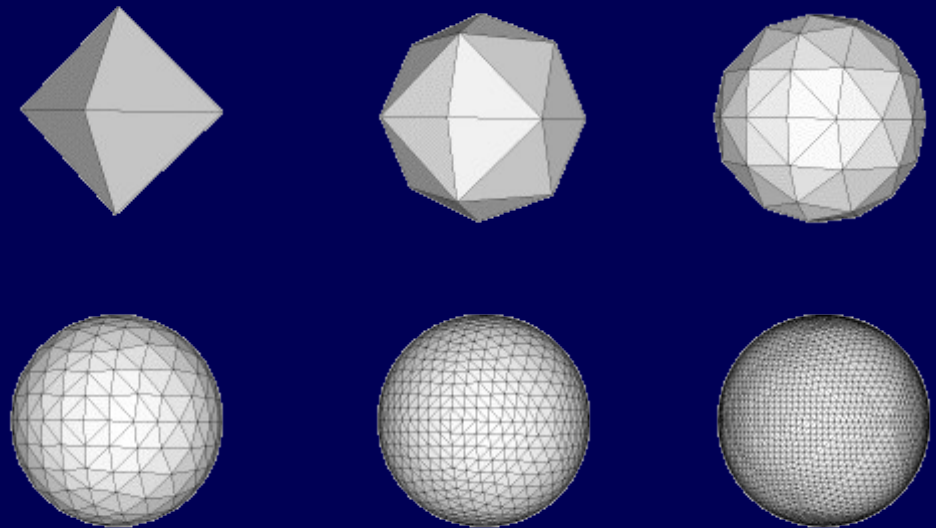
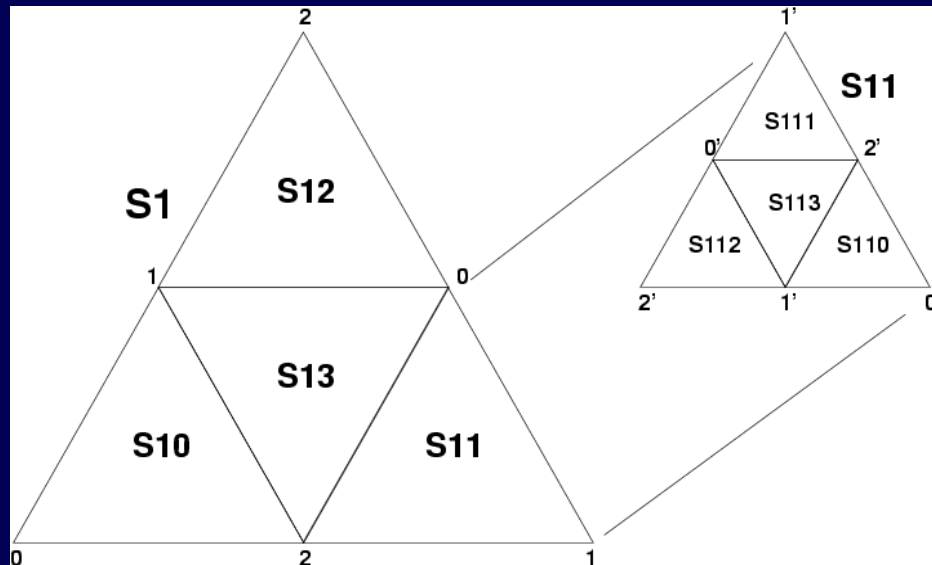
Possibili indicizzazione di una tabella 2-d (3-d, ...):

- Affettamento a piacere della sfera – no indice
- Indicizzare (B-tree) su un solo asse
- Usare una funzione che mappa 2-d $\Leftrightarrow$ 1-d quindi B-tree
- Usare un vero indice spaziale come l' R-tree

La scelta più efficace è portare una tabella con coordinate sferiche (2-d) ad essere trattabile dal DB server come una tabella ad una sola dimensione così da poter essere indicizzabile con un normale schema B-tree.

Varie funzioni di riempimento (**tassellizzazione**) della sfera sono state proposte (e usate) nel passato, ma per l'uso Astronomico vengono usate HTM ed HEALPix.

HTM: Hierarchical Triangular Mesh



HTM: www.sdss.jhu.edu/htm/

Inventato alla Johns Hopkins University – ora usato da vari progetti.
Il numero totale di pixel (in questo caso *trixels*) in una mappa è governata dal parametro *depth*:

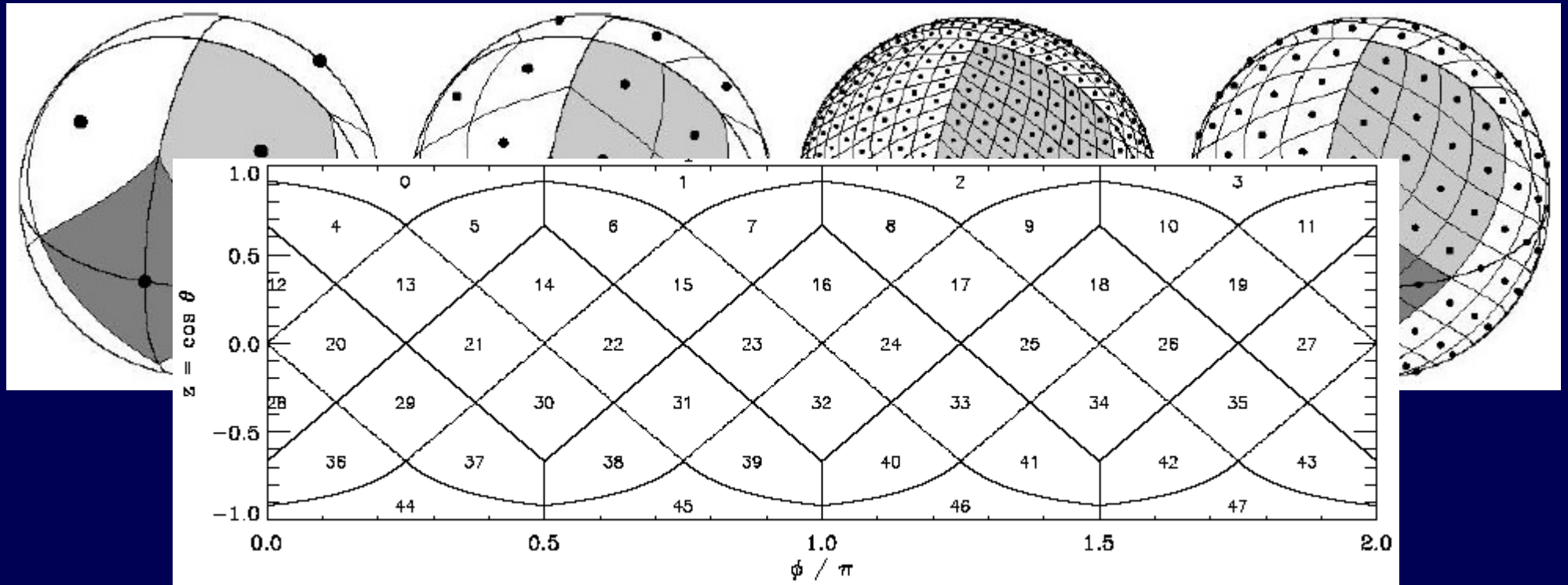
d (depth): [0 , 25]

$$N_{\text{pix}} = 8 \times 4^d$$

$$\text{ID}_{\text{range}}: [N_{\text{pix}} , 2 \times N_{\text{pix}} - 1]$$

Res. Max: 9.0×10^{15} pix $\rightarrow 7.7'' \times 10^{-3}$ (24 cm on Earth)

HEALPix: Hierarchical Equal Area isoLatitude Pixelization



HEALPix: healpix.jpl.nasa.gov

Inventato ad ESO per COBE. Ora usato da WMAP, Planck, etc.

Il numero totale di pixel in una mappa è governata dal parametro *order*:

k (order): [0 , 29]

$$N_{\text{side}} = 2^k$$

$$N_{\text{pix}} = 12 \times N_{\text{side}}^2$$

ID_{range}: [0 , $N_{\text{pix}} - 1$]

$$\Omega_{\text{pix}} = \pi / (3 \times N_{\text{side}}^2)$$

Res. Max: 3.46×10^{18} pix $\rightarrow 3.93'' \times 10^{-4}$ (1.2 cm on Earth)

DIF: Dynamyc Indexing Facility

DIF è un set di tools all'interno di MCS che permette di indicizzare automaticamente una tabella con coordinate sferiche in modo trasparente all'utente usando sia HTM che HEALPix.

Requisiti di sistema:

- MCS installato con opzione `–enable-dif`
- MySQL versione `>= 5.1.20`
- Perl con `DBD::mysql`

Per i nostri demo inoltre:

- IDL versione `>= 5.6` con libreria HEALPix (`>= 2.0`)

DIF: installazione

Installando MCS dare **--enable-dif** (oppure **--enable-all**) allo script **configure**. Dare **configure --help** per tutte le opzioni (è richiesto il modulo Perl DBD::mysql). Dopo il **make** e **make install** dare il comando **dif -install**.

Con **dif --help** si possono avere tutte le opzioni.

Per indicizzare una tabella eseguire un comando tipo:

```
dif --index-htm MyDBname MyCatname 6  
      "RAcs/3.6E5" "DECcs/3.6E5"
```

6 è il parametro di profondità ($N_{\text{pix}} = 8 \times 4^6 = 32768$). Le due espressioni seguenti servono a passare a gradi.

Per indicizzare una tabella sia con HTM che con HEALPix eseguire un comando tipo:

```
dif --index-both MyDBname MyCatname 6 0 8  
      "RAcs/3.6E5" "DECcs/3.6E5"
```

8 è l'*order* HEALPix ($N_{\text{pix}} = 12 \times (2^8)^2 = 786432$) e 0 indica schema RING (1 per NESTED).

DIF: installazione

Notare:

- La profondità della pixelizzazione (numero di pixel in cui viene divisa la sfera) dipende dal contenuto della tabella. Per cataloghi di oggetti la nostra esperienza suggerisce che si possono avere (in media) alcune migliaia di oggetti per pixel con performance ottime.
- Nel caso di cross-match tra cataloghi questo numero andrebbe ridotto di almeno un'ordine di grandezza. Prevediamo di effettuare dei benchmark in futuro.

Funzioni MySQL create:

```
DIF.getHTMDepth(db, tab)
DIF.getHEALPOrder(db, tab)
DIF.getHEALPNested(db, tab)
DIF.getRa(db, tab)
DIF.getDec(db, tab)
... (TBC)...
```

```
DIF_HTMLLookup(d, Ra, Dec)
DIF_HEALPLookup(s, k, Ra, Dec)
DIF_HTMCircle(Ra, Dec, R)
DIF_HTMRect(Ra, Dec, S1, S2)
DIF_HTMRectV(Ra1, Dec1, ...)
DIF_HEALPCircle(Ra, Dec, R)
DIF_Sphedist(Ra1, Dec1, Ra2, Dec2)
```

DIF: sessione MySQL interattiva

```
% mysql --local-infile=1 -u root -p DIF
```

```
Enter password:
```

```
mysql> show tables;
```

```
+-----+  
| Tables_in_DIF |  
+-----+  
| dif            |  
| tbl            |  
+-----+
```

```
mysql> describe dif;
```

```
+-----+-----+-----+-----+-----+-----+  
| Field | Type          | Null | Key | Default | Extra |  
+-----+-----+-----+-----+-----+-----+  
| id    | bigint(20)    | YES  |     | NULL    |       |  
| full  | tinyint(1)    | YES  |     | NULL    |       |  
+-----+-----+-----+-----+-----+-----+
```

```
mysql> select count(*) from dif;
```

```
+-----+  
| count(*) |  
+-----+  
| 0        |  
+-----+
```

DIF: sessione MySQL interattiva

```
mysql> describe tbl;
```

Field	Type	Null	Key	Default	Extra
db	char(50)	NO	PRI		
name	char(50)	NO	PRI		
HTMDepth	int(11)	YES		NULL	
HEALPNested	int(11)	YES		NULL	
HEALPOrder	int(11)	YES		NULL	
Ra_field	char(100)	YES		NULL	
Dec_field	char(100)	YES		NULL	

```
mysql> select * from tbl;
```

db	name	HTM Depth	HEALP Nested	HEALP Order	Ra_field	Dec_field
RossCats	GSC_22	6	0	8	RAcs/3.6e5	DECcs/3.6e5
RossCats	UCAC_2	6	0	6	RAcs/3.6e5	DECcs/3.6e5
RossCats	ASCC_25	6	0	6	RAcs/3.6e5	DECcs/3.6e5

DIF: demo::Messier

```
mysql> DROP TABLE IF EXISTS Messier;
```

```
mysql> CREATE TABLE Messier (  
  M          INT      NOT NULL,  
  Type       CHAR(2)  DEFAULT '**',  
  Const      CHAR(3)  DEFAULT '***',  
  Mag        FLOAT,  
  Ra         FLOAT,  
  Decl       FLOAT,  
  Dist       CHAR(20),  
  App_size   CHAR(20) DEFAULT 'unknown',  
  PRIMARY KEY (M));
```

```
mysql> LOAD DATA LOCAL INFILE './messier' INTO TABLE Messier;
```

```
mysql> select count(*) from Messier;
```

```
+-----+  
| count(*) |  
+-----+  
|      110 |  
+-----+
```

```
mysql> SELECT * FROM Messier;
```

```
...
```


DIF: demo::Messier

```
% dif -index-htm test Messier 6 "RA * 15E0" Decl
```

```
mysql> use test;
```

```
mysql> describe Messier;
```

Field	Type	Null	Key	Default	Extra
M	int(11)	NO	PRI		
Type	char(2)	YES	MUL	**	
Const	char(3)	YES		***	
Mag	float	YES		NULL	
Ra	float	YES		NULL	
Decl	float	YES		NULL	
Dist	char(20)	YES		NULL	
App_size	char(20)	YES		unknown	
htmID	smallint(5) unsigned	YES	MUL	NULL	

DIF: demo::Messier

Selezione circolare intorno a M31 e rettangolo in Leo:

```
mysql> SELECT * FROM Messier_htm WHERE  
        DIF_HTMCircle(0.7*15, 41.3, 30);
```

M	Type	Const	Mag	Ra	Decl	Dist	App_size	htmID
31	GX	And	4.8	0.7116	41.268	2.2 Mly	192.4'x62.2'	64538
32	GX	And	8.7	0.7116	40.860	2.2 Mly	8.7'x6.4'	64539
110	GX	And	9.4	0.6733	41.686	2.2 Mly	21.9'x10.9'	64571

```
mysql> SELECT * FROM Messier_htm WHERE  
        DIF_HTMRect(0.7*15, 41.3, 60);
```

...same result...

```
mysql> SELECT * FROM Messier_htm WHERE  
        DIF_HTMRectV(180.,28, 142.5,32, 180.,10, 142.5,6);
```

M	Type	Const	Mag	Ra	Decl	Dist	App_size	htmID
66	GX	Leo	8.2	11.3367	12.9917	35 Mly	9.1'x4.1'	57426
95	GX	Leo	10.4	10.7333	11.7033	38 Mly	7.5'x5.0'	58216
96	GX	Leo	9.1	10.78	11.8217	38 Mly	7.6'x5.2'	58216
65	GX	Leo	9.3	11.315	13.0933	35 Mly	8'x1.5'	58273
105	GC	Leo	9.2	10.7967	12.5817	38 Mly	5.4'x4.8'	58356

DIF: IDL_demo::ASCC / UCAC / GSC

Cataloghi *custom* utilizzati:

- ASCC-2.5:
 - catalogo astro-fotometrico (B, V) di 2.5 M-oggetti
- UCAC-2:
 - catalogo astrometrico (mag A in [V,R]) di 48.3 M-oggetti
- GSC-2.2:
 - catalogo fotometrico (I, R, B, V) di 455 M-oggetti



DIF: IDL_demo::ASCC / UCAC / GSC

Uso delle librerie utente `mcs_usrlib.pro` e `mcs_healplib.pro` (libreria HEALPix installata):

Librerie e programmi scaricabili da ross.iasfbo.inaf.it/MCS/

Demo con query che legge sempre (e solo) la colonna con i pixel ID:

```
mcshpdemo1.pro
```

Keywords:

Order = ordine della mappa (in [0,29])

TABName = nome base della tabella

...

Demo con query che produce sempre 2 colonne, quella con i pixel ID + una qualsiasi altra (anche combinazione):

```
mcshpdemo2.pro
```

Keywords:

Order = ordine della mappa (in [0,29])

TABName = nome base della tabella

...

DIF: IDL_demo::ASCC / UCAC / GSC

```
; UCAC2 objects density 4 deg around (0, 0)
IDL> mcshpdemo1, order=6, $
  Query = 'SELECT healpID FROM UCAC_2_healp WHERE
DIF_HEALPCircle(0, 0, 240)'
...
; UCAC2 Average A color 4 deg around (0, 0)
IDL> mcshpdemo2, order=6, $
  Query = 'SELECT healpID,Amm FROM UCAC_2_healp WHERE
DIF_HEALPCircle(0, 0, 240)'
...
; ASCC2.5 full-sky objects density
IDL> mcshpdemo1, tabname='ASCC_25', $
  Query='select healpID from ASCC_25', col=5
; ASCC2.5 average B-V color per pixel
IDL> mcshpdemo2, order=6, $
  Query='select healpID, (Bcm-Vcm)/1e2 AS BmV from ASCC_25 where
Bcm>0 and Vcm>0', outfile='ascc_b-v.fits'
; Replot in Galactic coords
IDL> mollview, 'ascc_b-v.fits', coord=['c','g'], grat=[30,30], $
  title='ASCC B-V'
```

Plot 1

Plot 2

Plot 3

Plot 4

DIF: IDL_demo::ASCC / UCAC / GSC

```
; Downgrade from Nside 64 to 16 and plot the boundaries of
Healpix pixels
IDL> loaddata_healpix, 'ascc_b-v.fits', dummy, map64
IDL> ud_grade, map64, map16, nside=16, order_in='RING'
IDL> mollview, map16, /ONLINE, coord=['c','g'], $
    title='ASCC B-V', hbound=16
; All the objects falling into a circle of radius 4 deg
; (use HTMCircle)
IDL> mshpdemo1, tabname='UCAC_2', /gnomic, $
    Query='SELECT healpID, Amm FROM UCAC_2_htm WHERE
DIF HTMCircle(0,0,240)', $
    OUTLINE={coord:'c', ra:[-4,4,4,-4,-4], dec:[-4,-4,4,4,-4],
line:2}
; All the objects falling into a square of side 8 deg
; (use HTMRect)
IDL> mshpdemo1, tabname='UCAC_2', /gnomic, $
    Query='SELECT healpID FROM UCAC_2_htm WHERE
DIF HTMRect(0,0,480,480)', $
    OUTLINE={coord:'c', ra:[-4,4,4,-4,-4], dec:[-4,-4,4,4,-4],
line:2}
```

Plot 1

Plot 2

DIF: IDL_demo::ASCC / UCAC / GSC

```
; UCAC 2 full sky density
IDL> mcshpdemo1, tabname='UCAC_2',
outfile='ucac2_cover_k8.fits', col=5, $ query='select healpID
from UCAC_2`

; PNG for UCAC 2 objects density - galactic, mollview
IDL> mollview, 'ucac2_cover_k8.fits', coord=['C','G'], /grat, $
col=5, title='UCAC 2 - objects density', $
png='ucac2_density_gal.png'
```

DIF: IDL_demo::ASCC / UCAC / GSC

```
; SMC
IDL> hprectsel, tabname='GSC_22',rect=[13.175,-72.833,500,300],$
  out='gsc22_smc_k8.fits'
IDL> loaddata_healpix, 'gsc22_smc_k8.fits', dummy, map
IDL> map2 = map*1.
IDL> map2[where(map eq 0)] = !values.f_nan & map = 0
IDL> gnomview, map2, /ONLINE, rot=[13.175,-72.833], /grat, $
  coord=['c','c'], col=5, title='GSC 2.2 - SMC region', $
  png='gsc22_smc_k8.png', unit='Nr of entries'
IDL> meshpdemo1, order=10, out='gsc22_smc_k10.fits', $
  query='SELECT DIF_HEALPLookup(0,10,RACs/3.6e5,DECcs/3.6e5) FROM
GSC_22_htm WHERE DIF_HTMRect(13.175,-72.833,500,300)'
IDL> loaddata_healpix, 'gsc22_smc_k10.fits', dummy, map
IDL> map2 = map*1.
IDL> map2[where(map eq 0)] = !values.f_nan & map = 0
IDL> gnomview, map2, /ONLINE, rot=[13.175,-72.833], /grat, $
  coord=['c','c'], col=5, title='GSC 2.2 - SMC region', $
  png='gsc22_smc_k10.png', unit='Nr of entries'
IDL> meshpdemo2, order=10, out='gsc22_smc_k10_b-r.fits', $
  query='SELECT DIF_HEALPLookup(0,10,RACs/3.6e5,DECcs/3.6e5),
(Bcm-Rcm)/1e2 FROM GSC_22_htm WHERE
DIF_HTMRect(13.175,72.833,500,300)and Rcm>0 and Bcm>0'
```

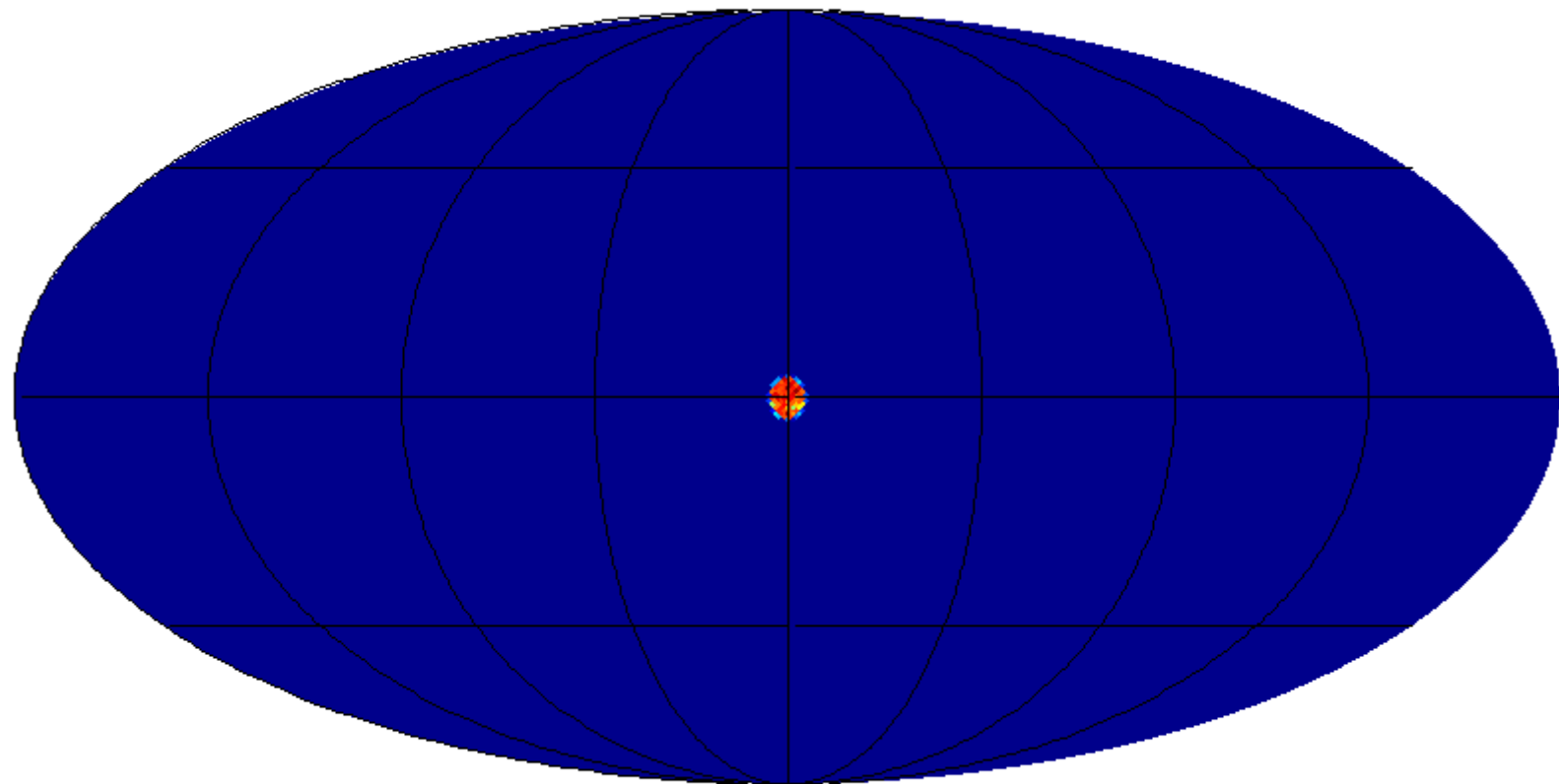
Plot 1

Plot 2

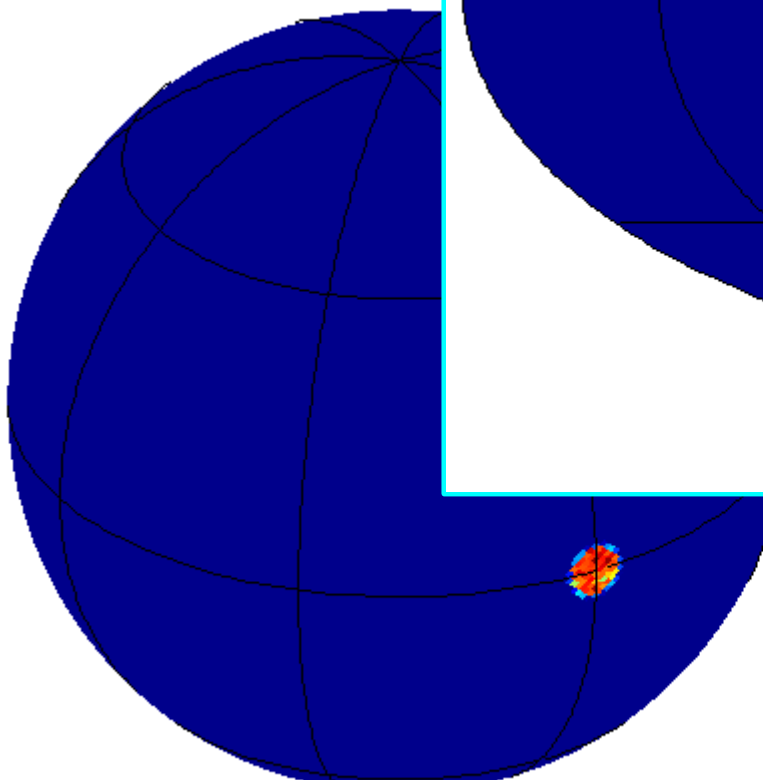
*Possiamo passare a Linux per i demo
in tempo reale*

ross.iasfbo.inaf.it/MCS/

MCS - DIF demo plot: UCAC_2

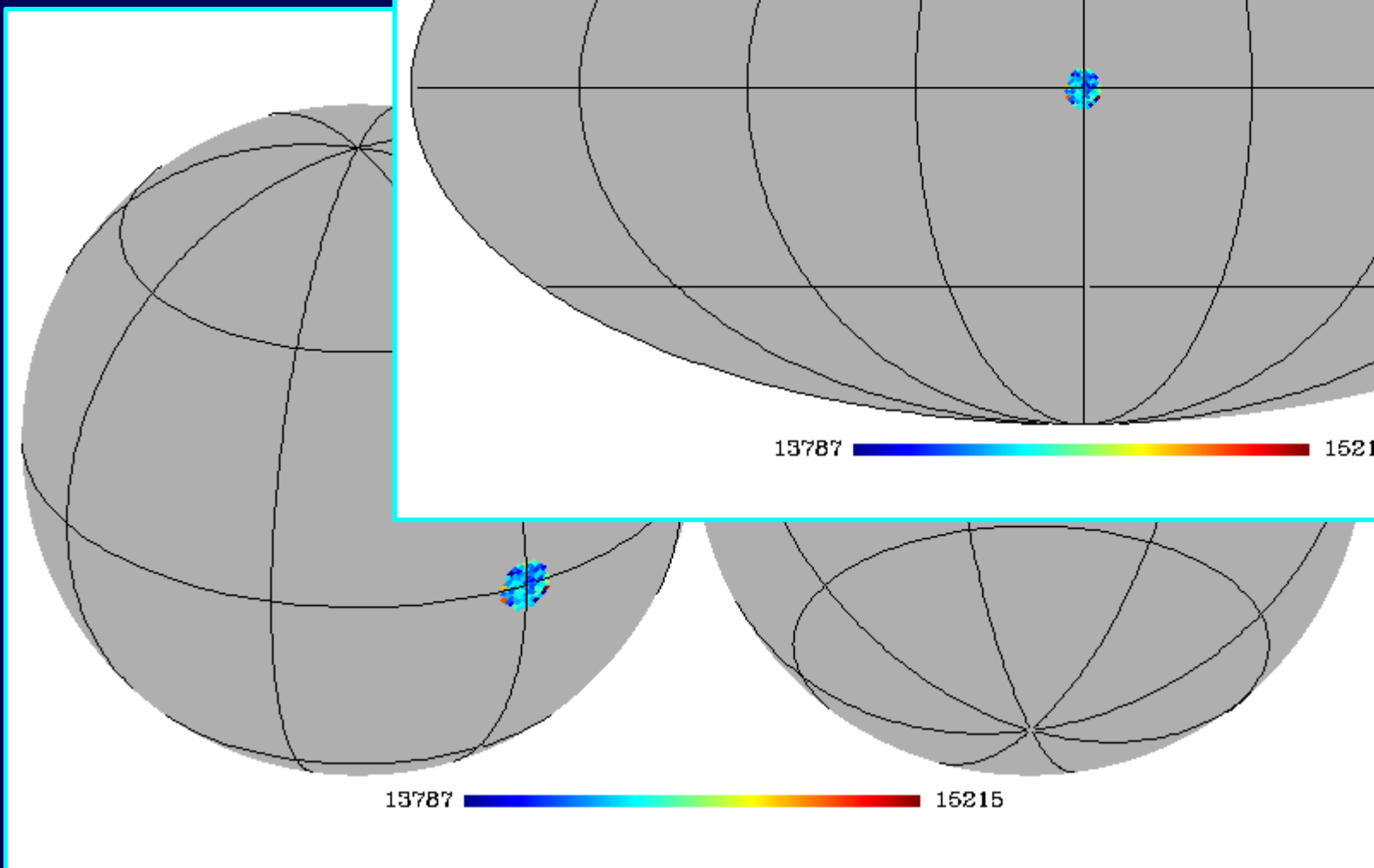
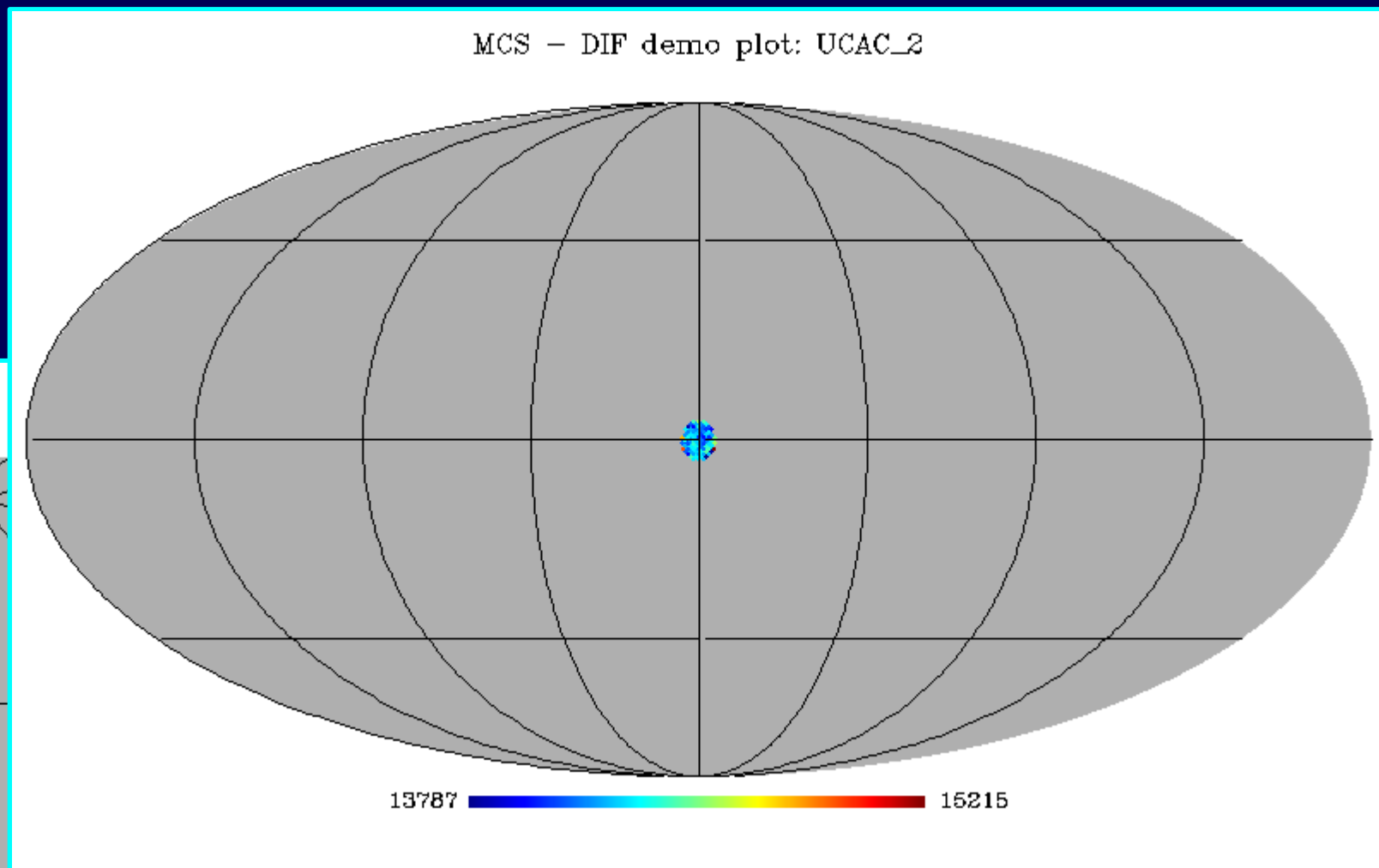


0 263 Nr of entries



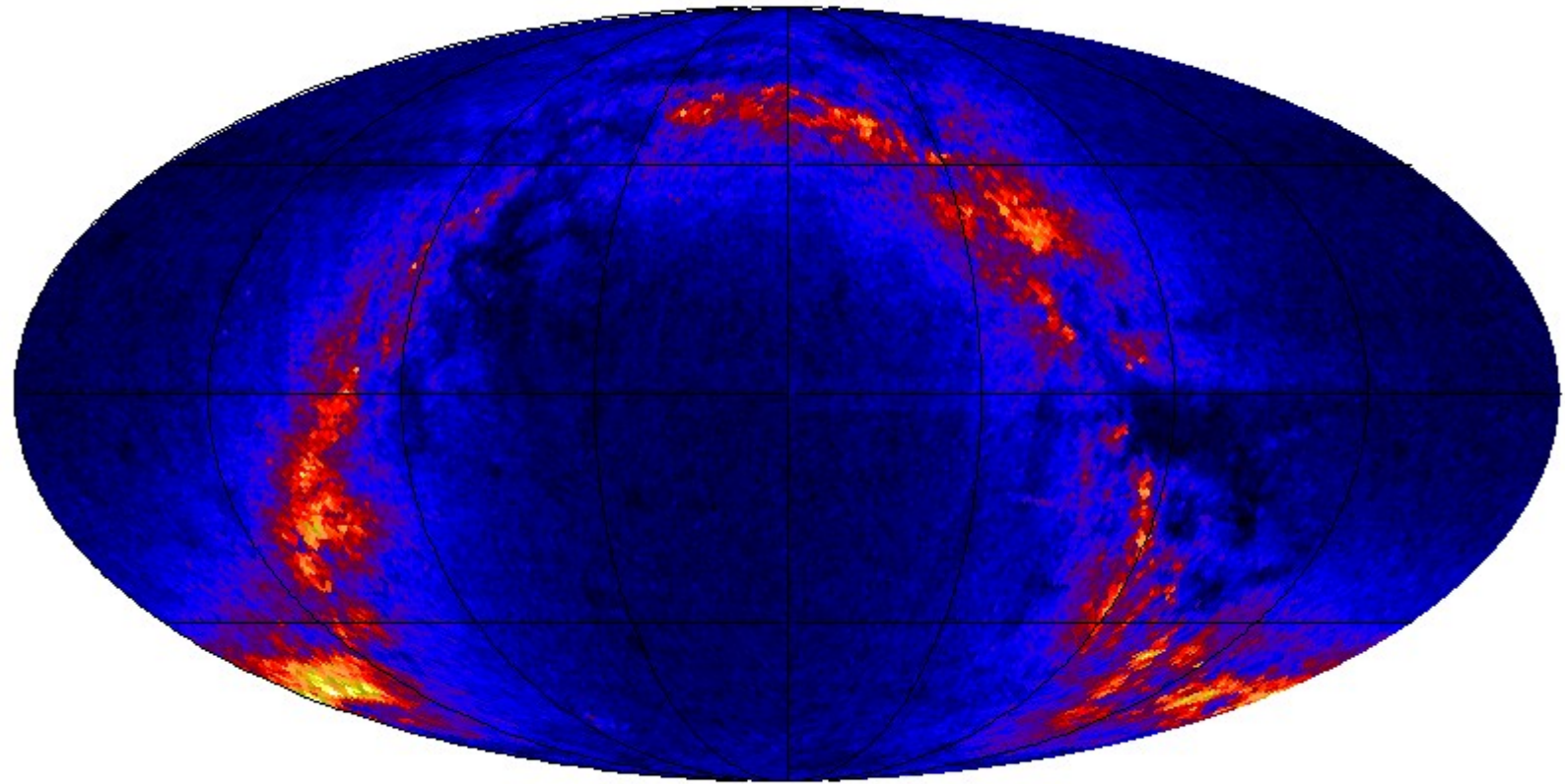
0 263 Nr of entries



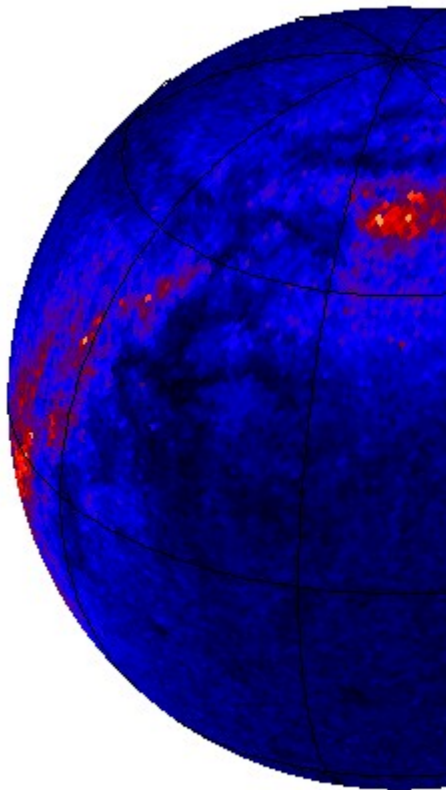


mcshpdemo1: ASCC2.5 density

MCS - DIF demo plot: ASCC_25



1  581 Nr of entries



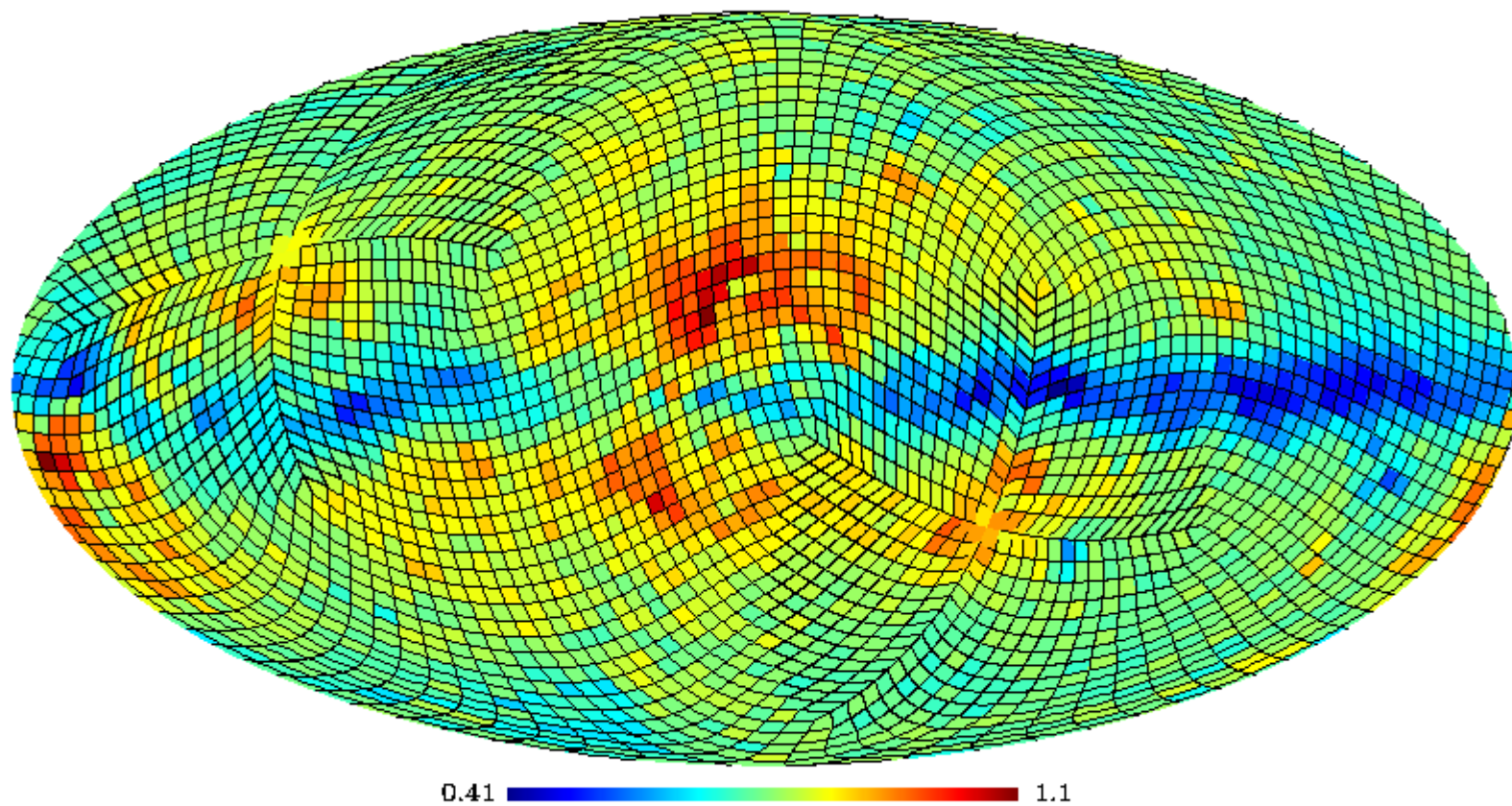
1  581 Nr of entries



mcshpdemo2: ASCC2.5 B-V

$K = 4$

ASCC B-V



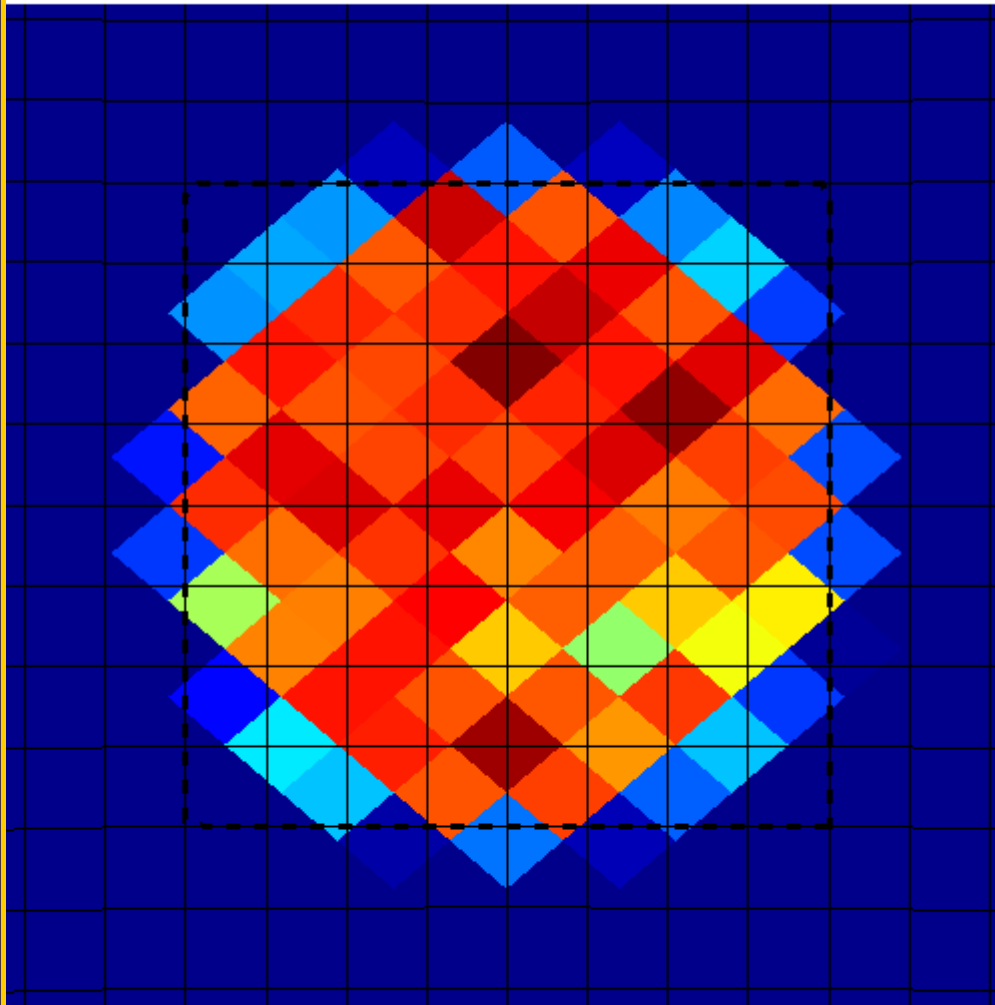
0.18 2.0 (B-V)

$K = 6$



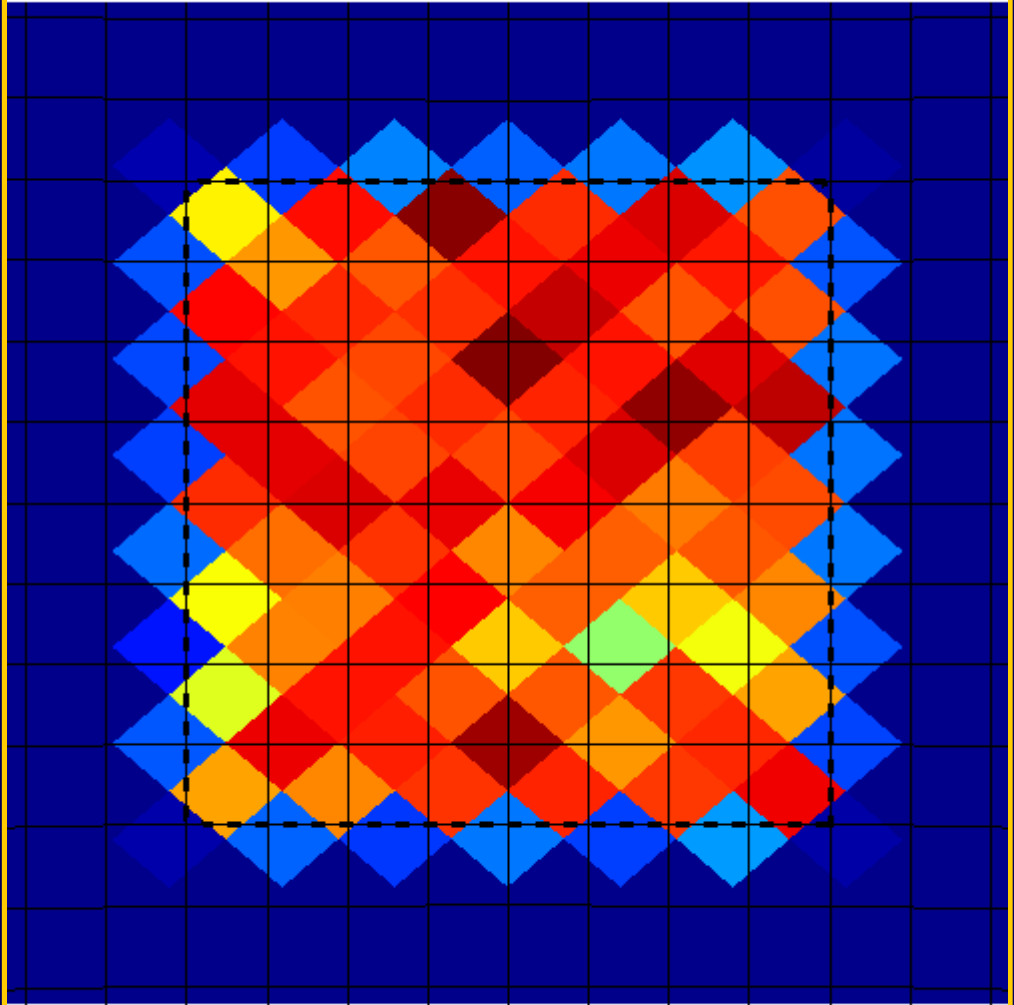
mcshpdemo1: UCAC2

MCS - DIF demo plot: UCAC_2



Circle

MCS - DIF demo plot: UCAC_2



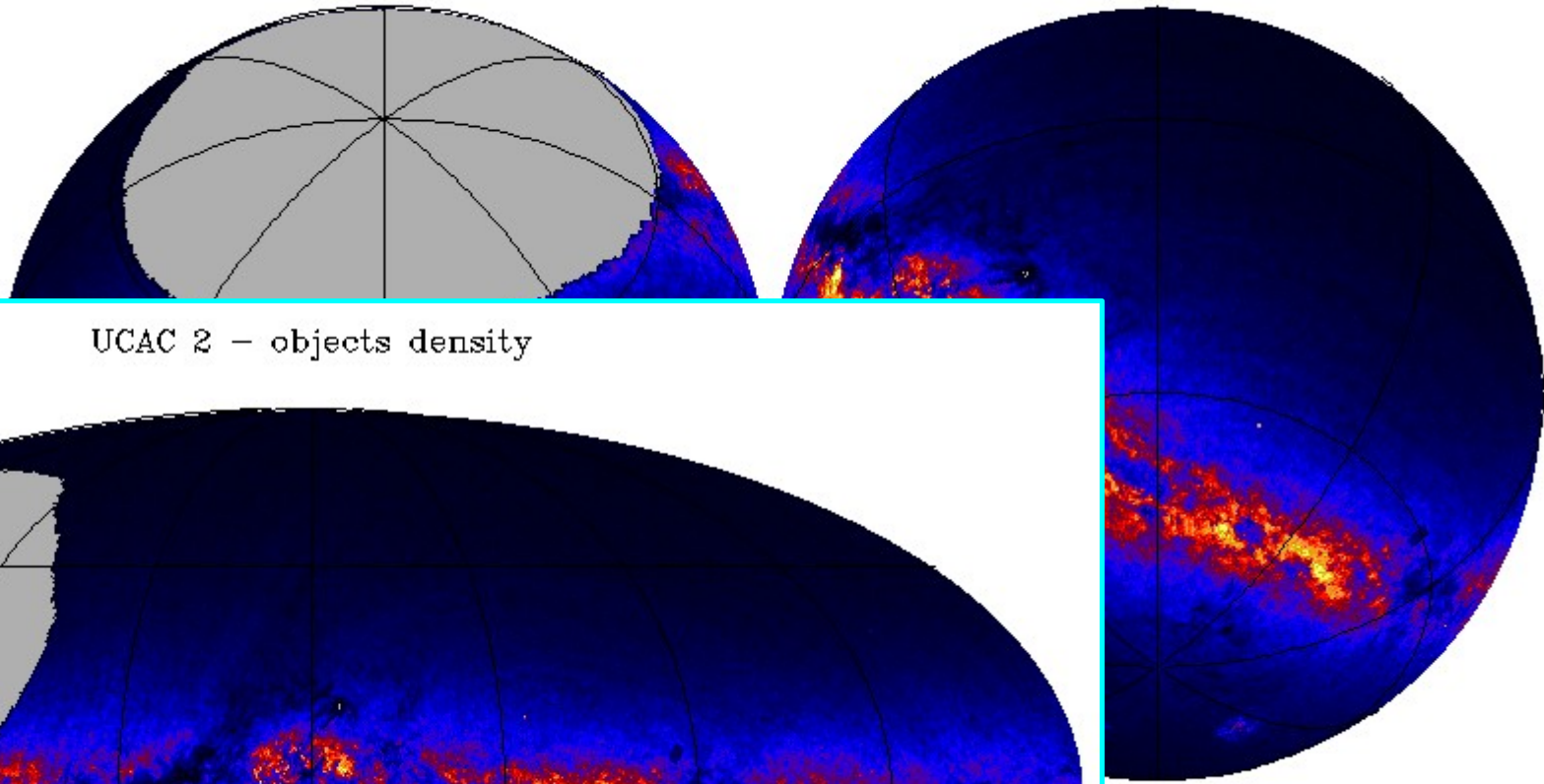
Square



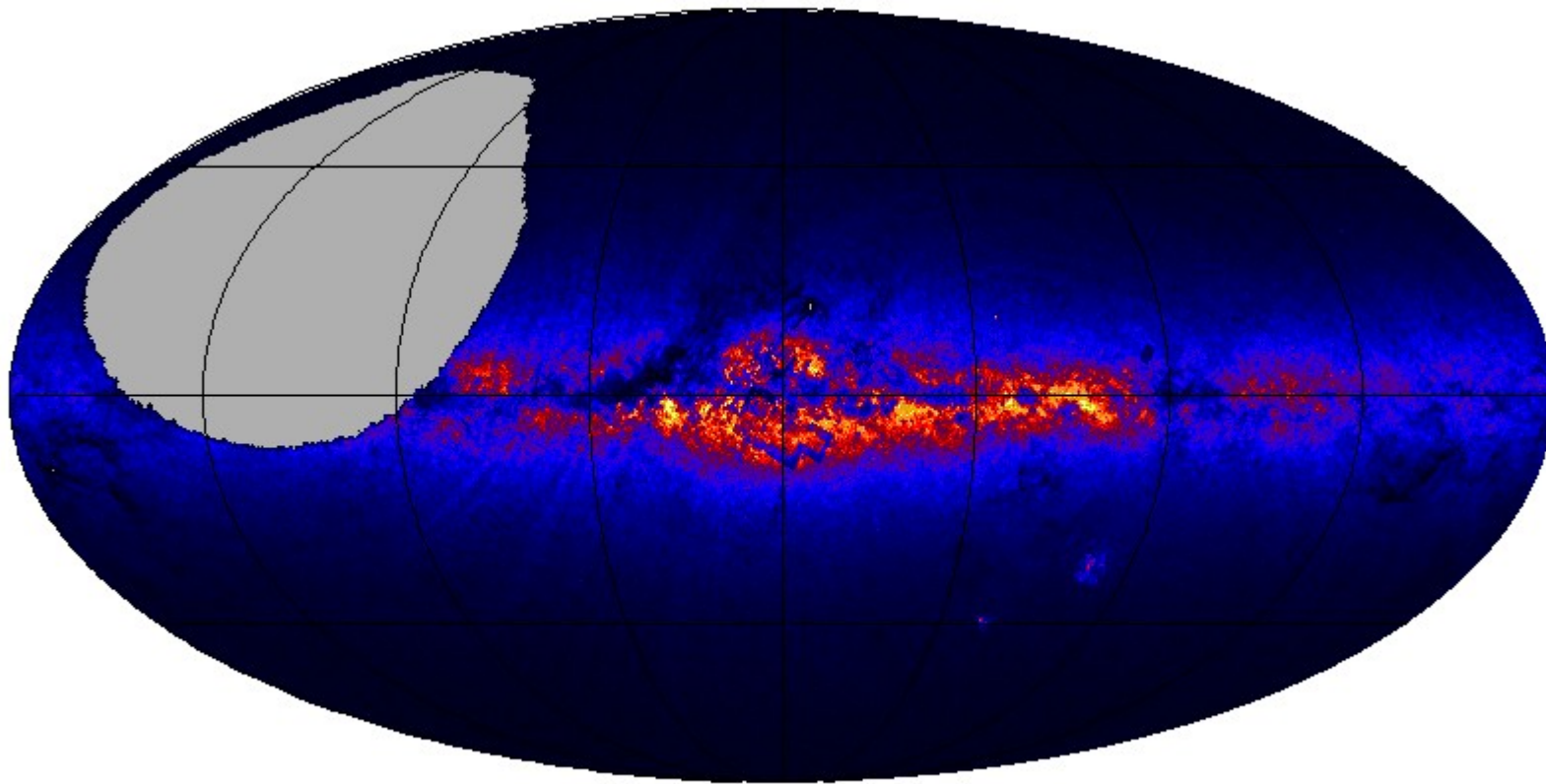
mcshpdemo1: UCAC2 density full sky

Ortographic
Equatorial

UCAC 2 - objects density



UCAC 2 - objects density



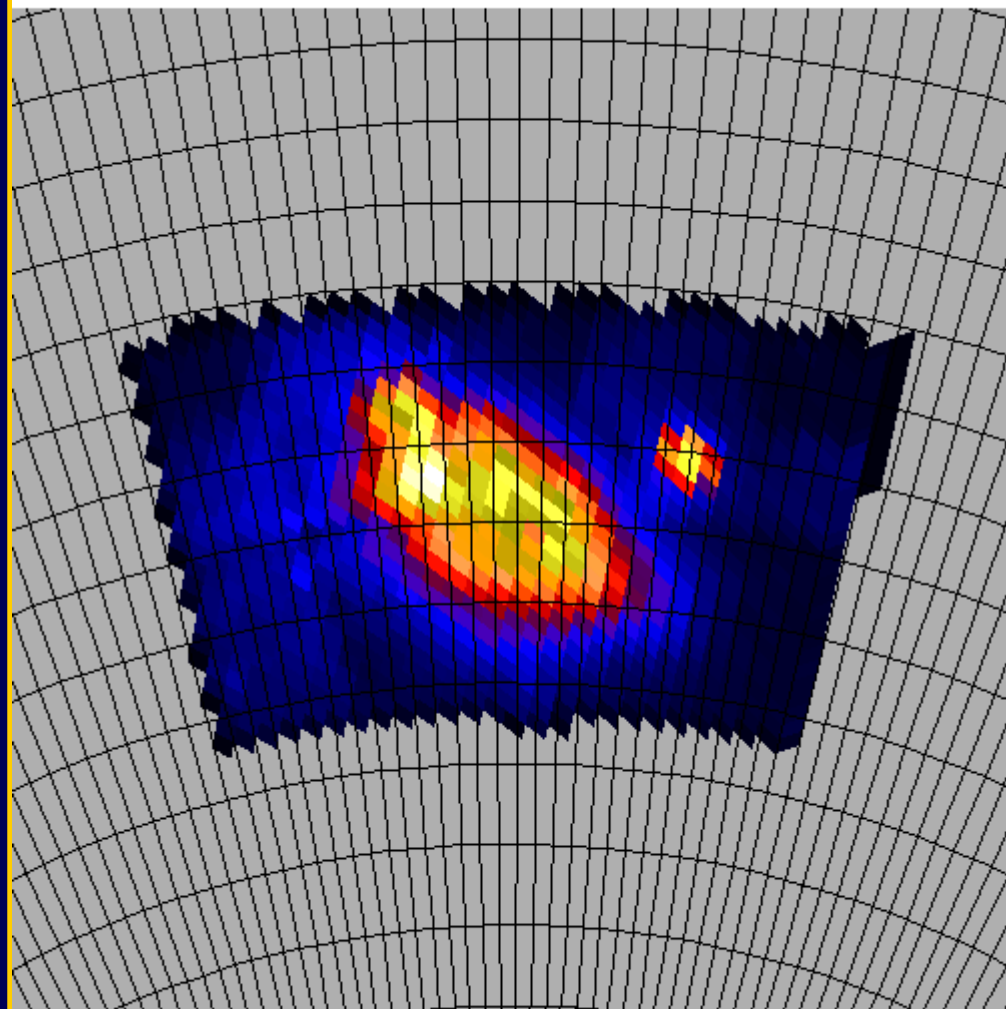
1 867

Mollview
Galactic



mcshpdemo1: GSC2.2 SMC

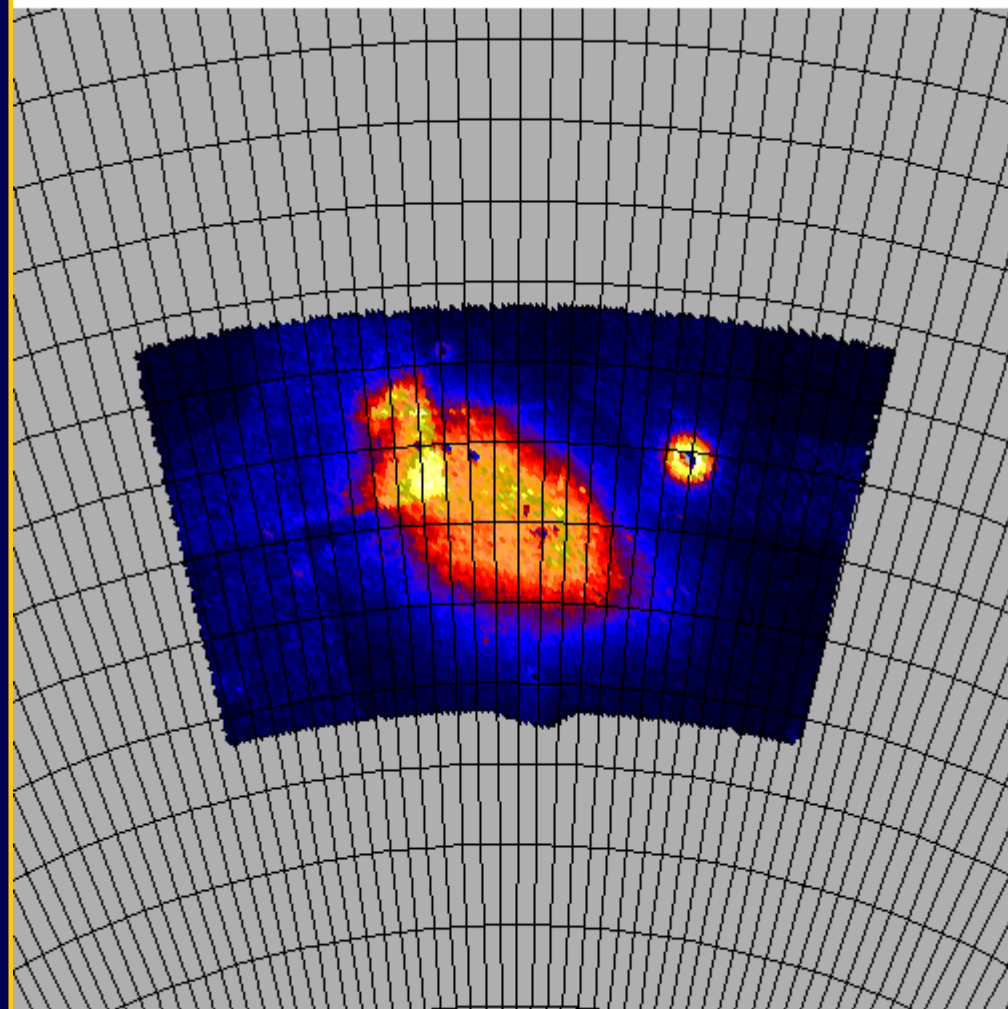
GSC 2.2 - SMC region



1 7464 Nr of entries
(13.2, -72.8) Equatorial

$K = 8$

GSC 2.2 - SMC region

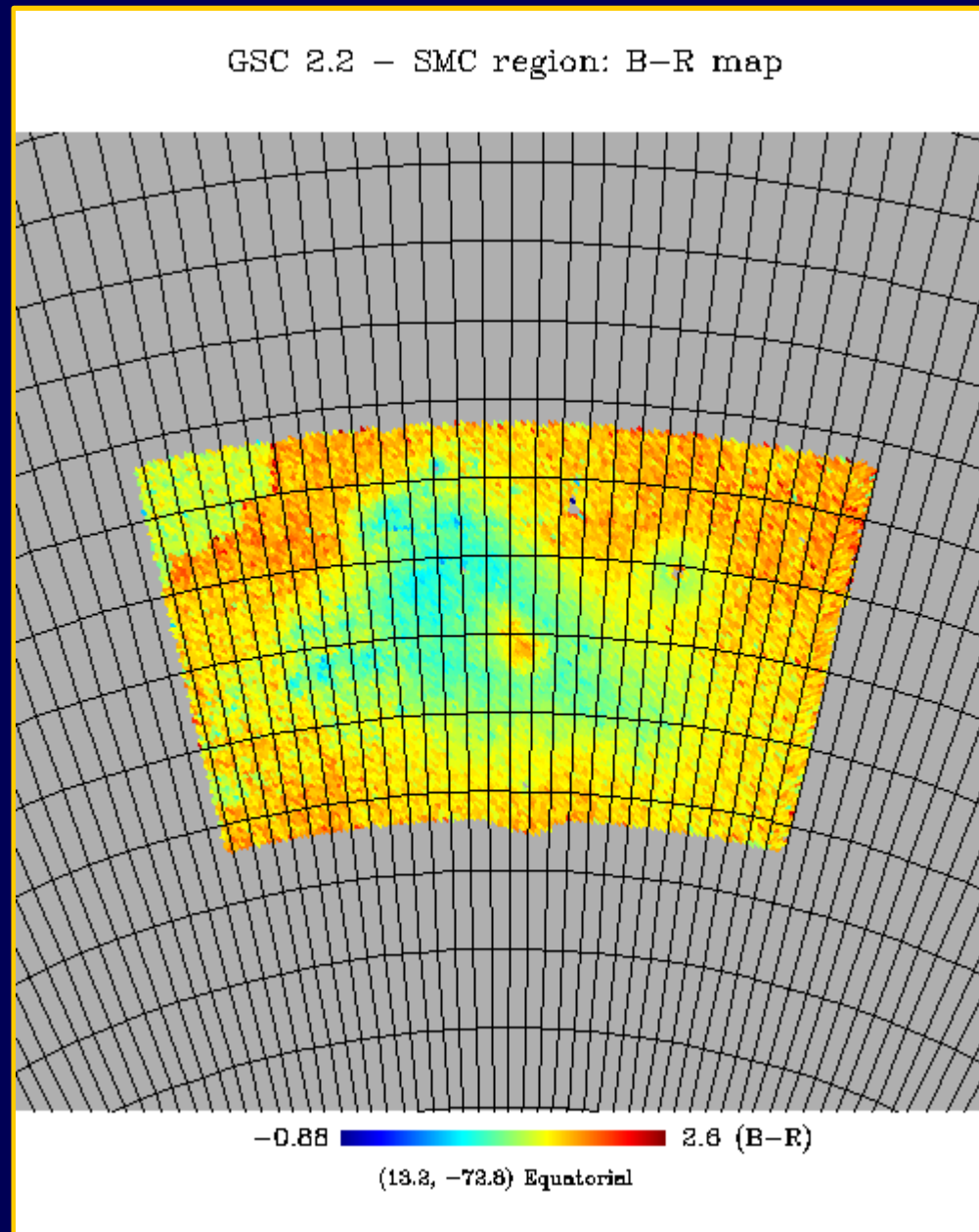


1 549 Nr of entries
(13.2, -72.8) Equatorial

$K = 10$



mcshpdemo1: GSC2.2 SMC



B - R

