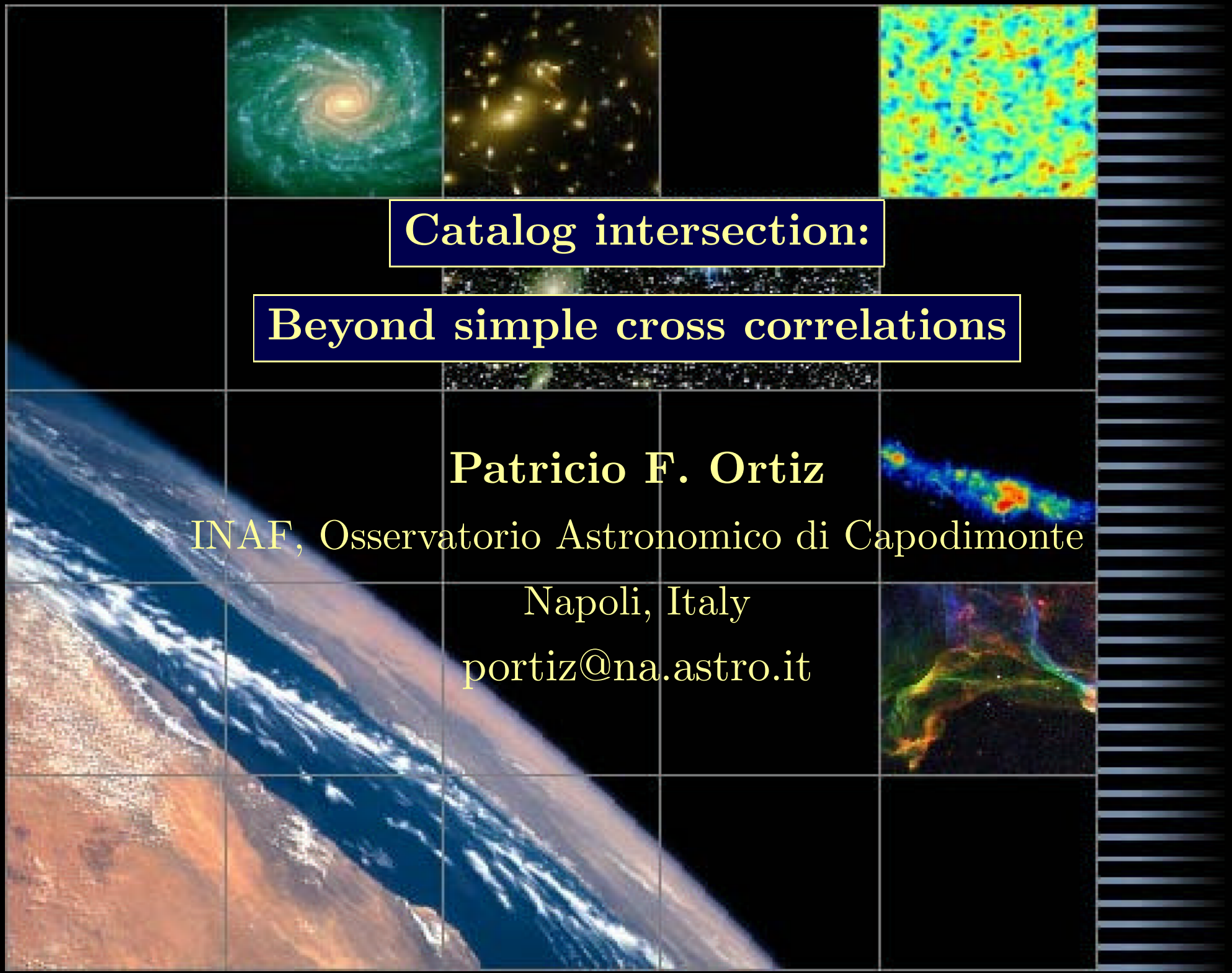


Catalog intersection:

Beyond simple cross correlations

Patricio F. Ortiz

INAF, Osservatorio Astronomico di Capodimonte

Napoli, Italy

portiz@na.astro.it

Summary

Summary

- Cross correlation: Matching amongst catalogues' elements

Summary

- Cross correlation: Matching amongst catalogues' elements
- Spliting the sky for good

Summary

- Cross correlation: Matching amongst catalogues' elements
- Spliting the sky for good
- Catalogues as individual entities

Summary

- Cross correlation: Matching amongst catalogues' elements
- Spliting the sky for good
- Catalogues as individual entities
- Sky-coverage based Operations amongst catalogues

Summary

- Cross correlation: Matching amongst catalogues' elements
- Spliting the sky for good
- Catalogues as individual entities
- Sky-coverage based Operations amongst catalogues
- Applications to archives and VO

Summary

- Cross correlation: Matching amongst catalogues' elements
- Spliting the sky for good
- Catalogues as individual entities
- Sky-coverage based Operations amongst catalogues
- Applications to archives and VO
- Conclusions

Cross correlation: matching amongst catalogues' elements

Cross correlation: matching amongst catalogues' elements

The most basic operation we would like to have access to is:

Cross correlation: matching amongst catalogues' elements

The most basic operation we would like to have access to is:

Find the elements in catalogue A which exhibit similar characteristics to elements listed in catalogue B

Cross correlation: matching amongst catalogues' elements

The most basic operation we would like to have access to is:

Find the elements in catalogue A which exhibit similar characteristics to elements listed in catalogue B

Matching elements which are located close in the sky is a common first step.

Cross correlation: matching amongst catalogues' elements

The most basic operation we would like to have access to is:

Find the elements in catalogue A which exhibit similar characteristics to elements listed in catalogue B

Matching elements which are located close in the sky is a common first step.

In order to provide this functionality efficiently it is common that the database system handling the catalogue supports some sort of double parameter indexation scheme, eg, in (α, δ)

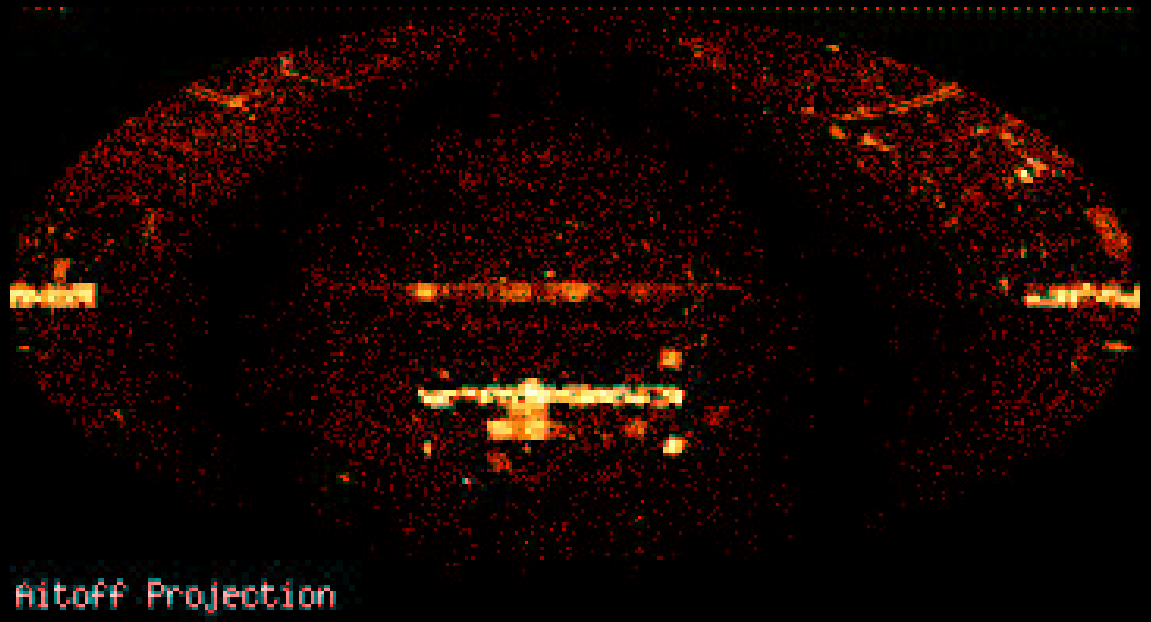
Spliting the sky for good

Spliting the sky for good

The very existence of an indexation scheme prompts for a way of counting the number of sources located in a certain region of the sky listed in any given catalogue.

Spliting the sky for good

The very existence of an indexation scheme prompts for a way of counting the number of sources located in a certain region of the sky listed in any given catalogue.



From Veron et al. 2001, List of known QSO's

Spliting the sky for good (cnt'd)

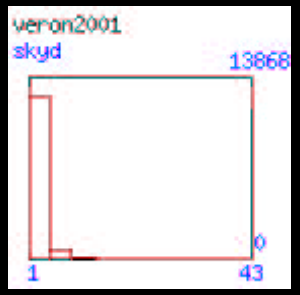
Spliting the sky for good (cnt'd)

This opens up the door to answer questions like:

Spliting the sky for good (cnt'd)

This opens up the door to answer questions like:

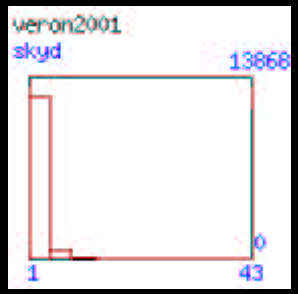
Find and list the areas of the sky where the source density falls in a given range.



Spliting the sky for good (cnt'd)

This opens up the door to answer questions like:

Find and list the areas of the sky where the source density falls in a given range.

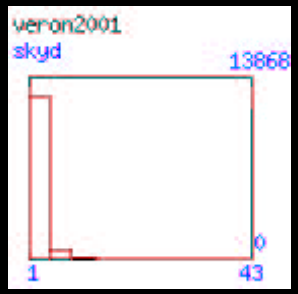


With a good sky partition method, accessing the whole catalogue is not necessary to find these areas: a great benefit in the case of XXL catalogues.

Spliting the sky for good (cnt'd)

This opens up the door to answer questions like:

Find and list the areas of the sky where the source density falls in a given range.



With a good sky partition method, accessing the whole catalogue is not necessary to find these areas: a great benefit in the case of XXL catalogues.

The type of scientific questions to answer depends on the nature of the catalogue itself.

Catalogues as individual entities

Datamining requires the handling of operation at the level of catalogues' entries as well as at the catalogues' properties level.

Catalogues as individual entities

Datamining requires the handling of operation at the level of catalogues' entries as well as at the catalogues' properties level.

Global properties should make catalogues retrievable individually, eg:

Catalogues as individual entities

Datamining requires the handling of operation at the level of catalogues' entries as well as at the catalogues' properties level.

Global properties should make catalogues retrievable individually, eg:

- Sky coverage (limits, source distribution)

Catalogues as individual entities

Datamining requires the handling of operation at the level of catalogues' entries as well as at the catalogues' properties level.

Global properties should make catalogues retrievable individually, eg:

- Sky coverage (limits, source distribution)
- nature

Catalogues as individual entities

Datamining requires the handling of operation at the level of catalogues' entries as well as at the catalogues' properties level.

Global properties should make catalogues retrievable individually, eg:

- Sky coverage (limits, source distribution)
- nature
- variables' content

Catalogues as individual entities

Datamining requires the handling of operation at the level of catalogues' entries as well as at the catalogues' properties level.

Global properties should make catalogues retrievable individually, eg:

- Sky coverage (limits, source distribution)
- nature
- variables' content
- variables' ranges

Catalogues as individual entities

Datamining requires the handling of operation at the level of catalogues' entries as well as at the catalogues' properties level.

Global properties should make catalogues retrievable individually, eg:

- Sky coverage (limits, source distribution)
- nature
- variables' content
- variables' ranges

In addition, comparison of these properties should be available.

Sky-coverage based Operations amongst catalogues

Sky-coverage based Operations amongst catalogues

Assuming the existence of an (α, δ) indexation scheme including index-cell counters, comparison amongst catalogues becomes possible.

Sky-coverage based Operations amongst catalogues

Assuming the existence of an (α, δ) indexation scheme including index-cell counters, comparison amongst catalogues becomes possible.

Questions like the following can be easily addressed:

Sky-coverage based Operations amongst catalogues

Assuming the existence of an (α, δ) indexation scheme including index-cell counters, comparison amongst catalogues becomes possible.

Questions like the following can be easily addressed:

- *Do catalogues A, and B intersect at all?*

Sky-coverage based Operations amongst catalogues

Assuming the existence of an (α, δ) indexation scheme including index-cell counters, comparison amongst catalogues becomes possible.

Questions like the following can be easily addressed:

- *Do catalogues A, and B intersect at all?*
- *Find the lowest density areas amongst cats A, B, C, and D* Looking for a deep field pointing?

Sky-coverage based Operations amongst catalogues

Assuming the existence of an (α, δ) indexation scheme including index-cell counters, comparison amongst catalogues becomes possible.

Questions like the following can be easily addressed:

- *Do catalogues A, and B intersect at all?*
- *Find the lowest density areas amongst cats A, B, C, and D* Looking for a deep field pointing?
- *Look for the densest areas of Cat A and the lowest density areas of Cat B.* Looking for good extragalactic fields?

How to solve this problem?

12	8	67	99	14	43
28	79	135	170	159	82
100	125	75	10	5	5
54	15	0	150	27	89

1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	0	1	1	1

12	8	67	99	14	43
28	79	135	170	159	82
100	125	75	10	5	5
54	15	0	150	27	89

0	0	0	0	0	0
0	0	1	1	1	0
1	1	0	0	0	0
0	0	0	1	0	0

12	8	67	99	14	43
28	79	135	170	159	82
100	125	75	10	5	5
54	15	0	150	27	89

1	1	0	0	1	1
1	0	0	0	0	0
0	0	0	1	1	1
0	1	1	0	1	0

The indexation scheme issue

The capacity to address the previous questions heavily relies on choosing a method to split the sky in fixed cells in order to compare the same areas.

The indexation scheme issue

The capacity to address the previous questions heavily relies on choosing a method to split the sky in fixed cells in order to compare the same areas.

An Earth-bound system (α, δ) may not be the most adequate over the long run. Using an extragalactic based system (based on QSO's for instance) may prove more adequate.

The indexation scheme issue

The capacity to address the previous questions heavily relies on choosing a method to split the sky in fixed cells in order to compare the same areas.

An Earth-bound system (α, δ) may not be the most adequate over the long run. Using an extragalactic based system (based on QSO's for instance) may prove more adequate.

In the experiments I've run, a system of nearly equal area tiles starting in the SCP and wrapping up to the NCP is used.

The indexation scheme issue

The capacity to address the previous questions heavily relies on choosing a method to split the sky in fixed cells in order to compare the same areas.

An Earth-bound system (α, δ) may not be the most adequate over the long run. Using an extragalactic based system (based on QSO's for instance) may prove more adequate.

In the experiments I've run, a system of nearly equal area tiles starting in the SCP and wrapping up to the NCP is used.

At a 30' x 30' ~ 165200 of these tiles is needed; the database uses it to speed up positional cross correlations.

Applications to archives and VO

Initially *targetless* (yet positionally oriented) queries can be addressed and provided as a service to the community.

Applications to archives and VO

Initially *targetless* (yet positionally oriented) queries can be addressed and provided as a service to the community.

Find the most heavily observed areas could be a query which would allow a user to dig an archive: in time to search for variability or to produce very deep fields (eg, photometric calibration fields) using the images stored.

Applications to archives and VO

Initially *targetless* (yet positionally oriented) queries can be addressed and provided as a service to the community.

Find the most heavily observed areas could be a query which would allow a user to dig an archive: in time to search for variability or to produce very deep fields (eg, photometric calibration fields) using the images stored.

Global positional comparison of remotely located catalogues can be sped up by just comparing their positional indexes before touching individual elements.

Conclusions

Conclusions

- Astronomical-object oriented cross correlations are a primary way to deal with issues related to point sources.

Conclusions

- Astronomical-object oriented cross correlations are a primary way to deal with issues related to point sources.
- sky zones cross correlations and operations would allow even more general datamining operations. Correlation involving extended objects (GMC's, large galaxies, wide field images, etc) could be then handled in a very effective way.

Conclusions

- Astronomical-object oriented cross correlations are a primary way to deal with issues related to point sources.
- sky zones cross correlations and operations would allow even more general datamining operations. Correlation involving extended objects (GMC's, large galaxies, wide field images, etc) could be then handled in a very effective way.
- Exploiting sky-coverage properties of catalogues on an individual basis should provide astronomers with a very powerful discovery tool.

Conclusions

- Astronomical-object oriented cross correlations are a primary way to deal with issues related to point sources.
- sky zones cross correlations and operations would allow even more general datamining operations. Correlation involving extended objects (GMC's, large galaxies, wide field images, etc) could be then handled in a very effective way.
- Exploiting sky-coverage properties of catalogues on an individual basis should provide astronomers with a very powerful discovery tool.
- Extending these tools to combinations of catalogues should enable us to address a number of new questions and in particular, provide a better way to exploit archived data.