Providing means for a better understanding of biodiversity

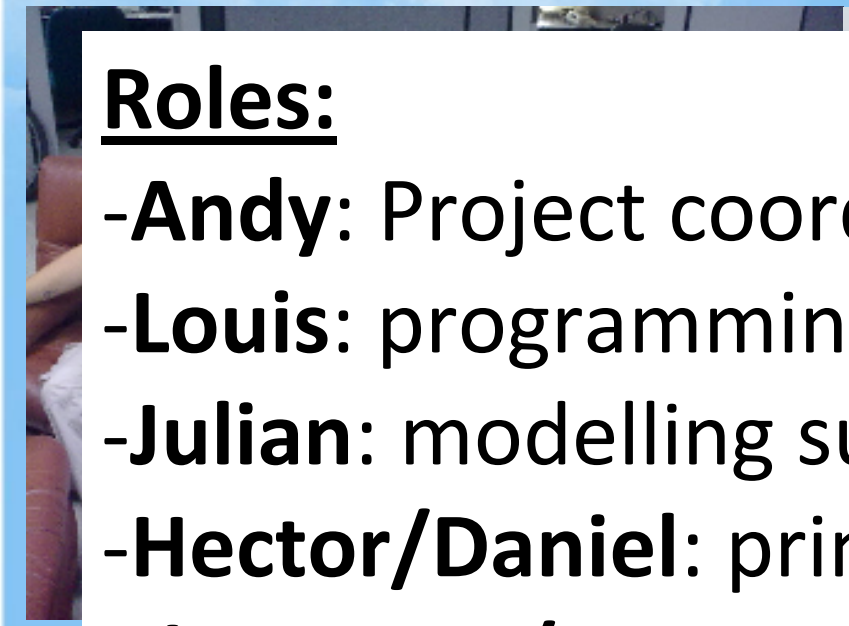
improving primary data and using it for threat assessment and in situ conservation planning in South America

Hector Tobón, Daniel Amariles, Julián Ramírez, Louis Reymondin, Jorge Camacho, Jhon Tello and Andy Jarvis

International Centre for Tropical Agriculture, CIAT
Cali, Colombia
February 16, 2011

Introducing working mates

The old crew... Louis, Hector, Julian, Daniel + Andy



Roles:

-**Andy**: Project coordination

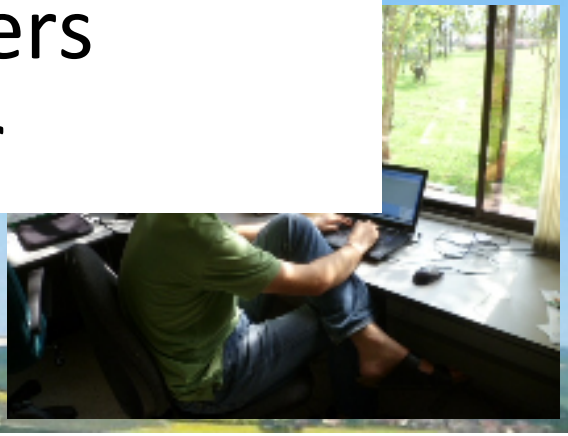
-**Louis**: programming supervision

-**Julian**: modelling supervision

-**Hector/Daniel**: primary java developers

-**Jhon Jairo/Jorge**: web developers

-**Johannes**: biodiversity modeler



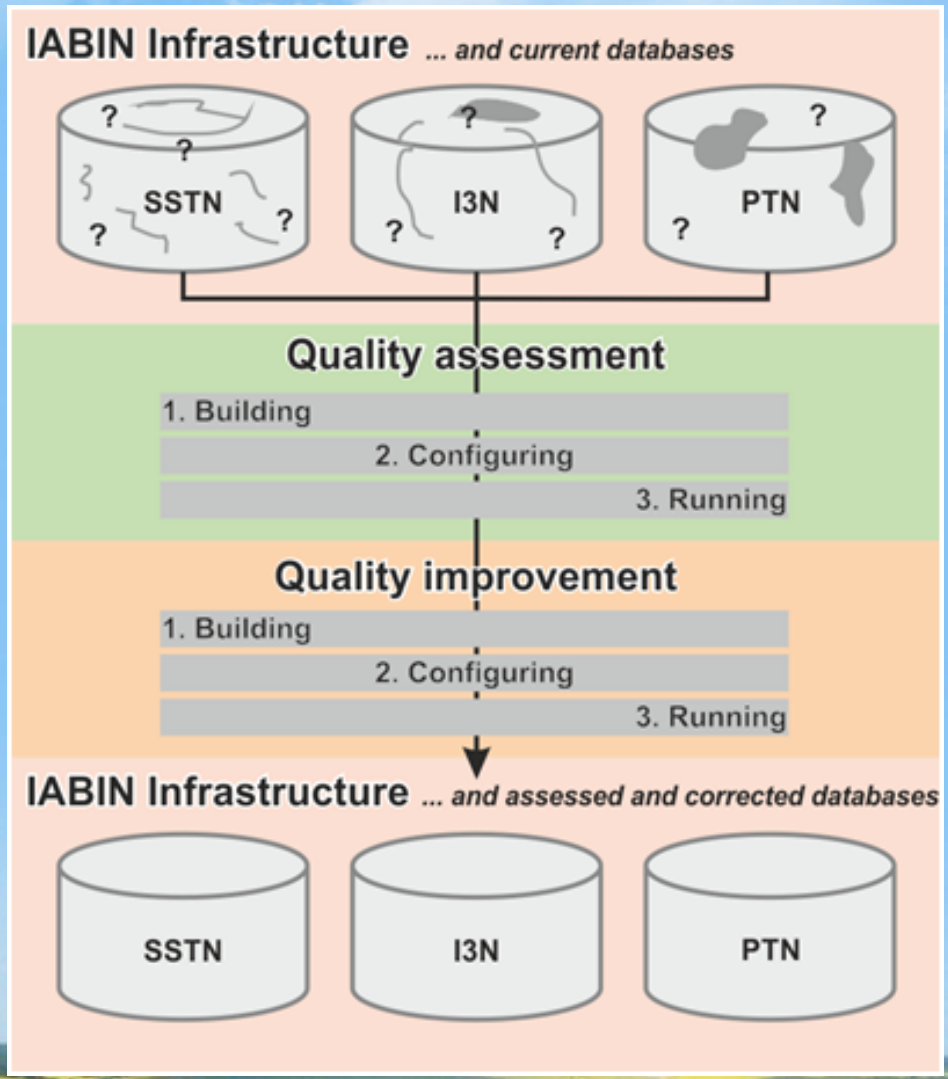
Deliverables

1. Data quality checking and improvement
    1. Cross-checking of coordinates
    2. Georreferencing
2. Niche modeling
    1. Training of niche models script
    2. Threat and conservation analysis over South America
3. Google-maps based navigation tool

Developed scripts and documentation for data cleansing

Keywords

- Automated algorithms
- Coordinate verification (error detection)
- Georeferencing process (error correction) - Biogeomancer



@Data quality: Why do we need high quality and reliable occurrence data?

- Analyse patterns of species diversity in throughout regions
- Train and evaluate niche models
- Assess conservation issues (*in-situ*, *ex-situ*)
- Assess impacts and threats on biodiversity: habitat degradation, deforestation and… climate change
- Among others…

@Data quality: Average status of a large dataset

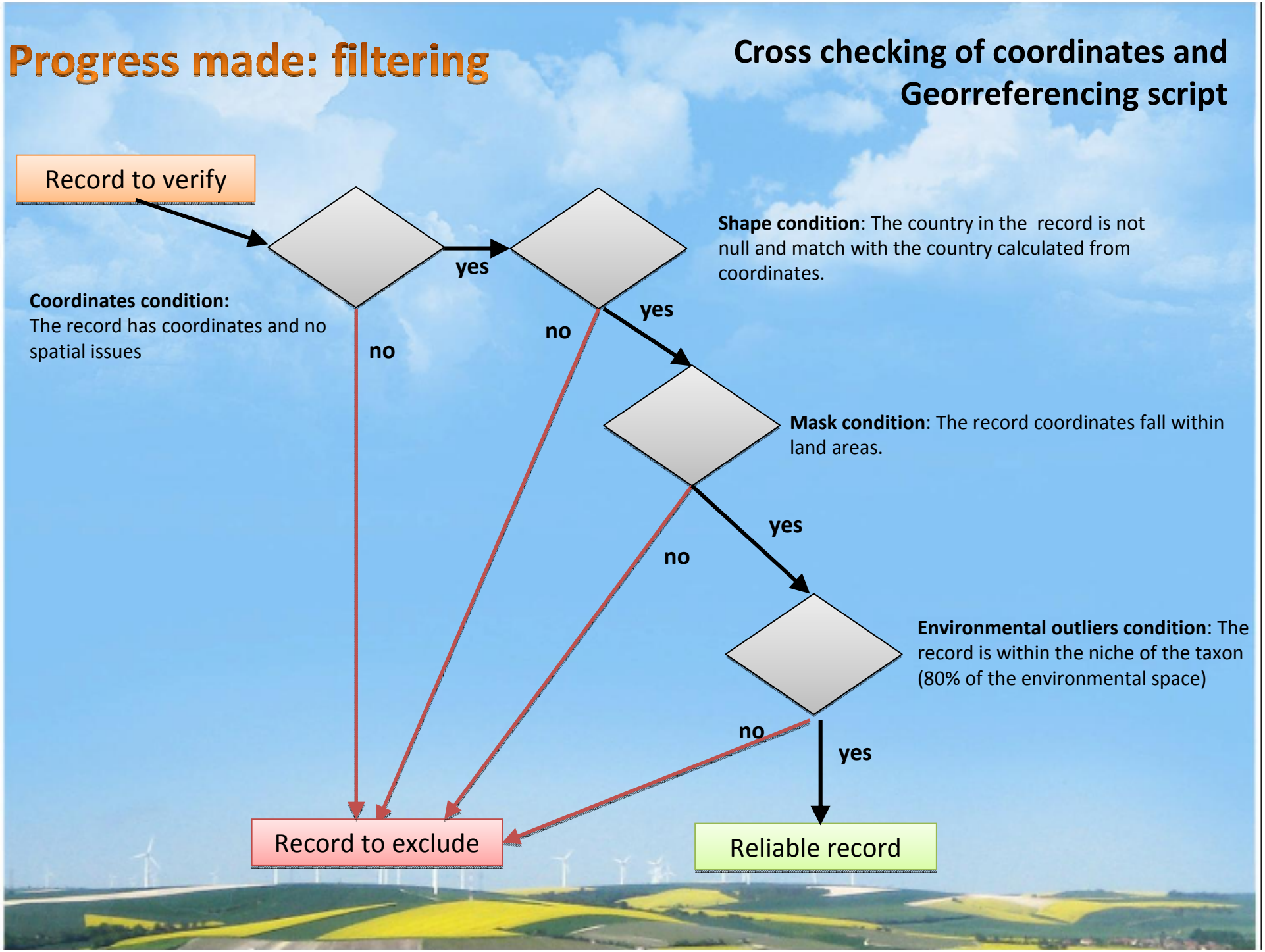
- The **IABIN TNs** databases holds a number of occurrences, but not all of them have coordinates or are free of geospatial issue
  - SSTN 3,866,145 occurrences, 3,452,938 (89.31%) with coordinates
  - PTN 1,144,678 occurrences, 583,753 (50.99%) with coordinates
  - I3N 19,663 occurrences, 2,991 (15.21%) with coordinates
- How many of them are correct, and reliable?
- How many new georreferences could we get?

@Data quality: Our approach

- How to make the data reliable enough?
  - Verify coordinates at different levels
    - Are the records where they say they are? [Country Level]
    - Are the records inside land areas (for terrestrial plant species only) [Continental Level]
    - Are all the records within the environmental niche of the taxon? [Environmental Level]
    - Sea records: not verifiable
  - Correct wrong references
  - Add references to those that do not have
  - Cross-check with curators and feedback the database

Progress made: filtering

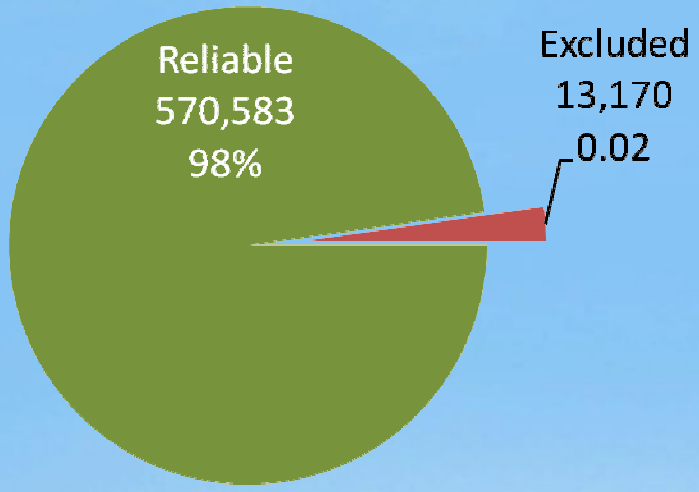
Cross checking of coordinates and Georreferencing script



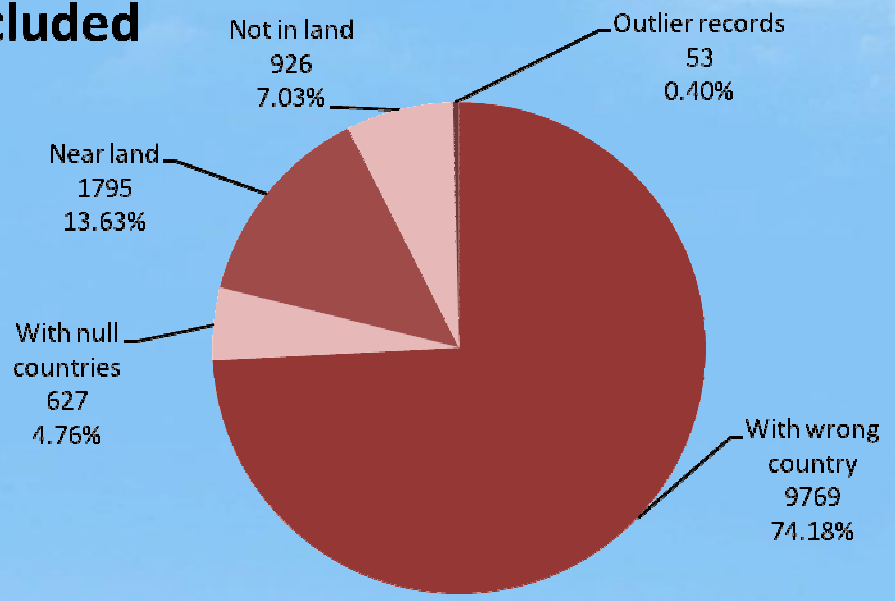
PTN Evaluation Results

Evaluated records: 583,753

Reliable vs Excluded



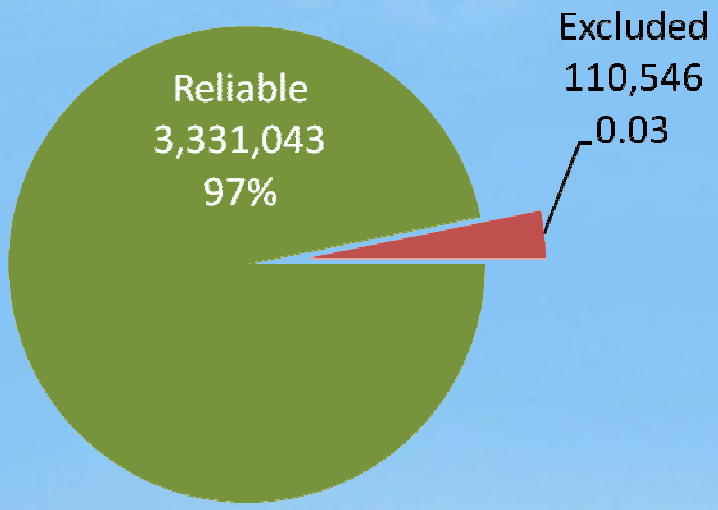
Excluded

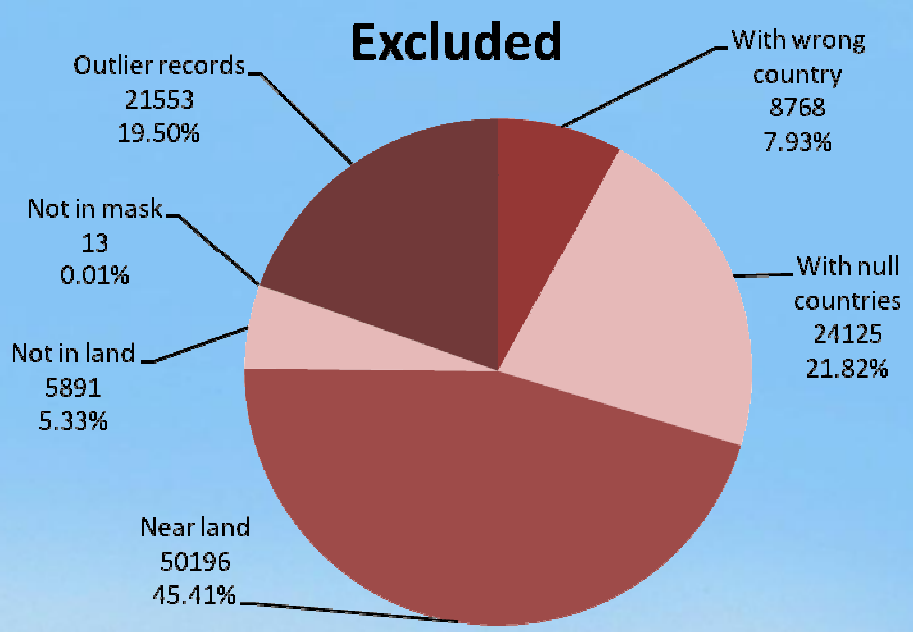


SSTN Evaluation Results

Evaluated records: 3,441,589

Reliable vs Excluded

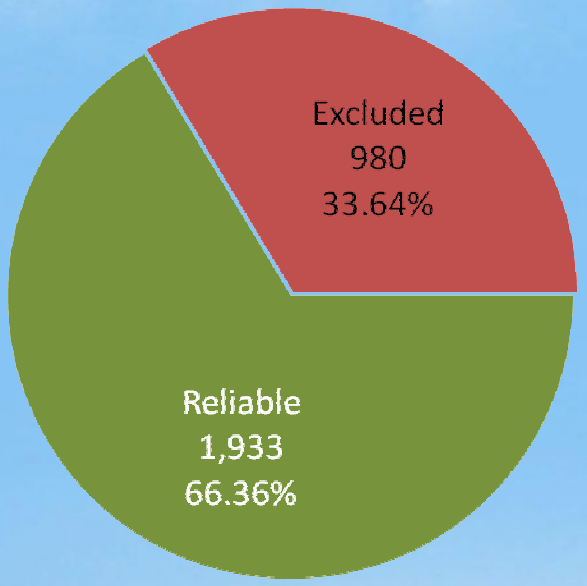




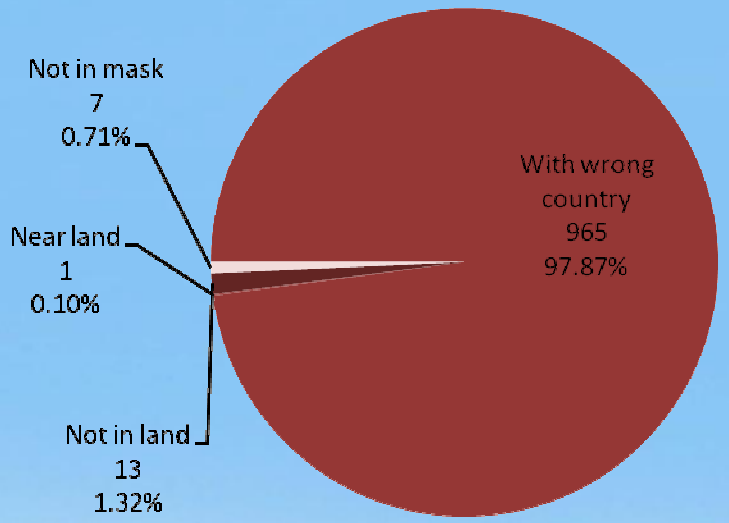
I3N Evaluation Results

Evaluated records: 2.913

Reliable vs Excluded

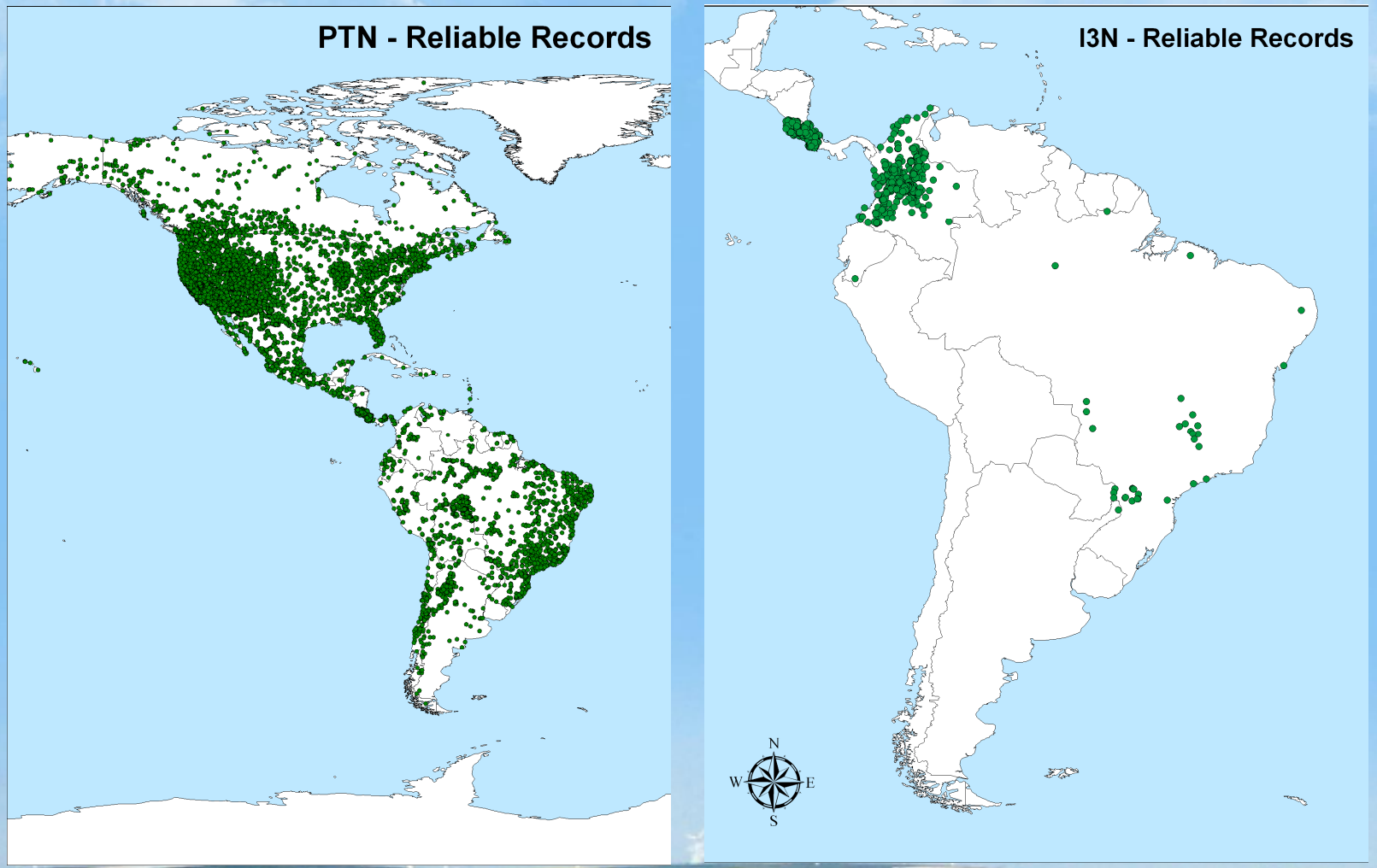


Excluded

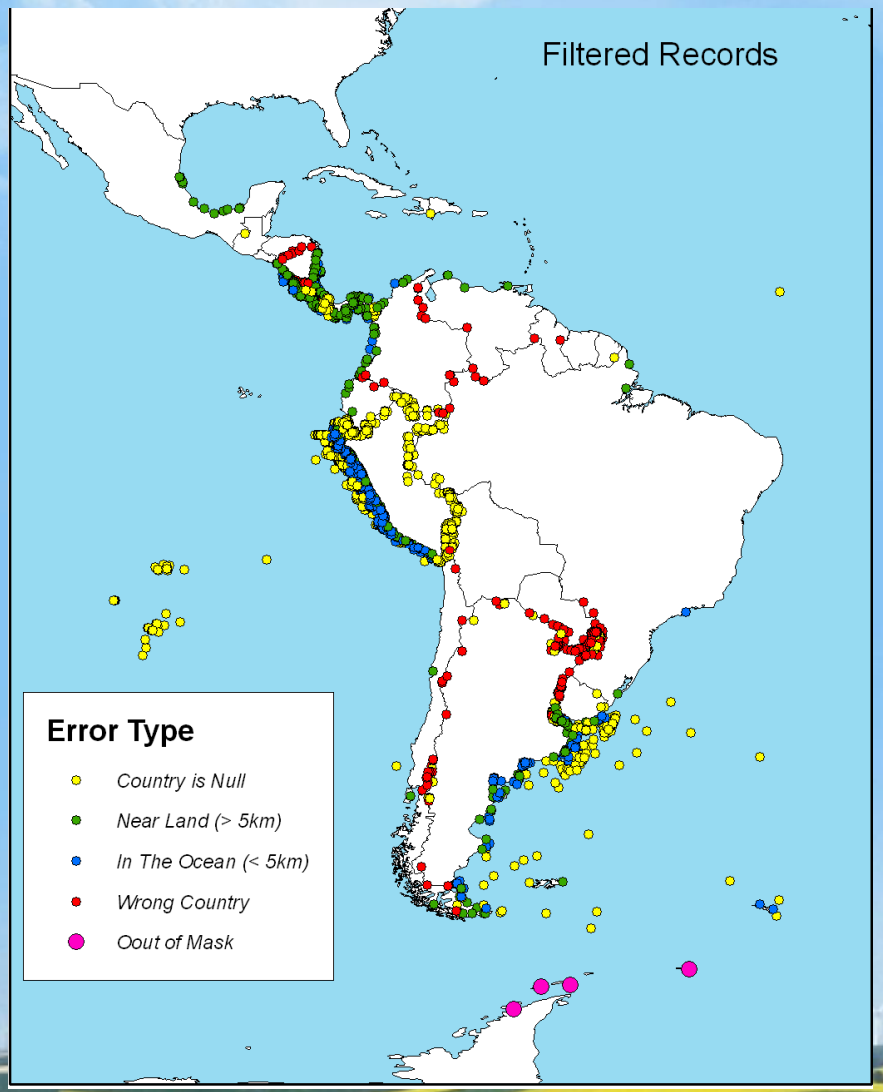


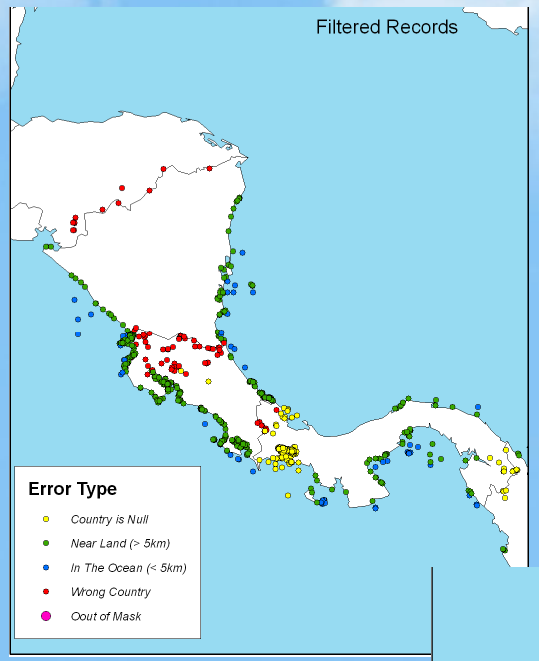
SSTN Reliable Records

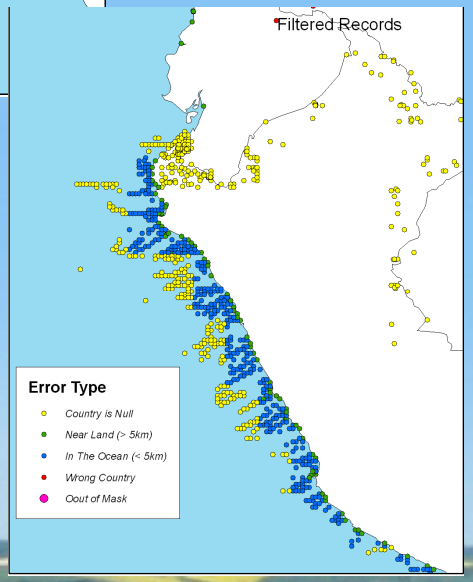
PTN I3N Reliable records



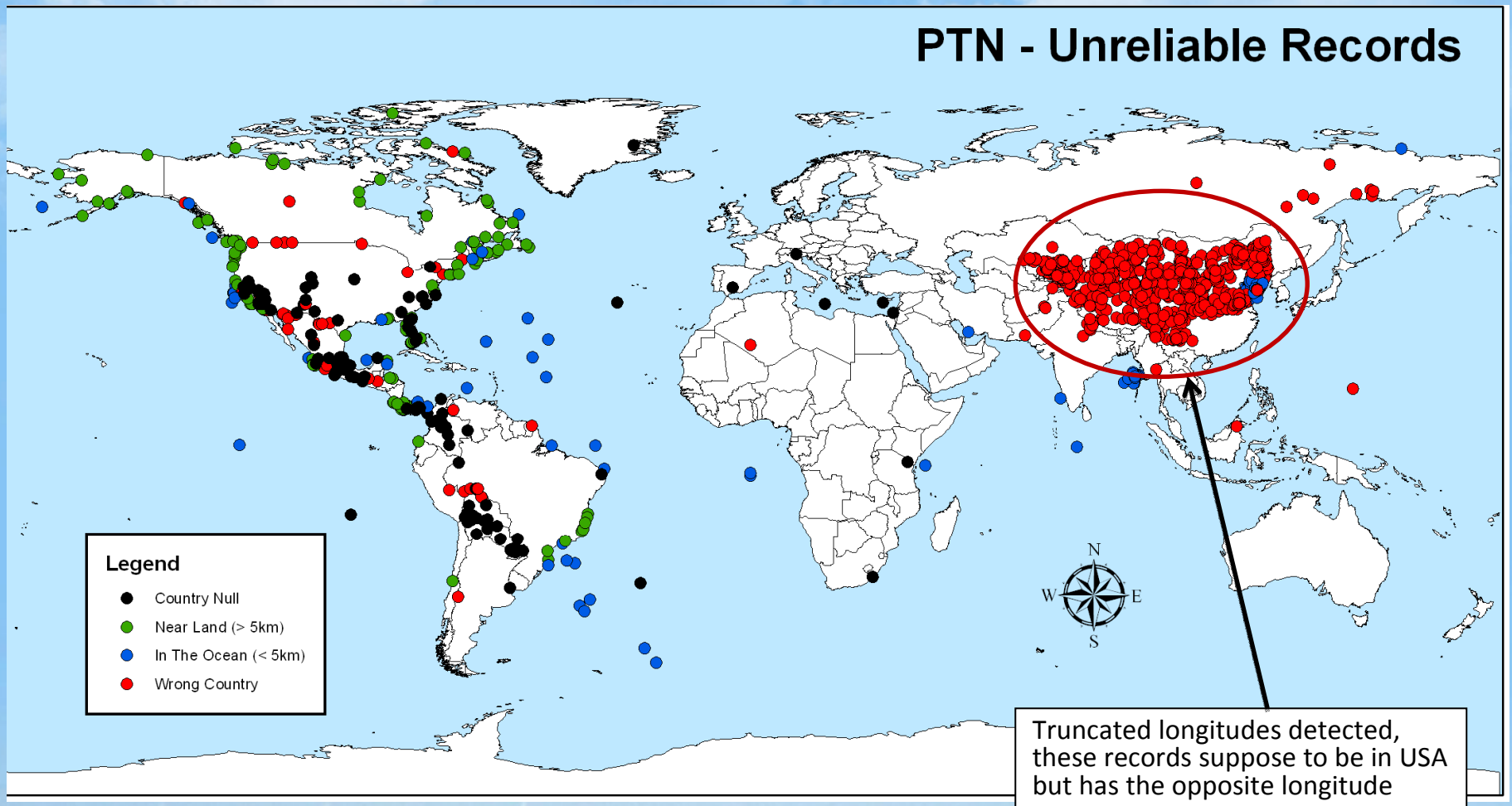
SSTN Filtered Records





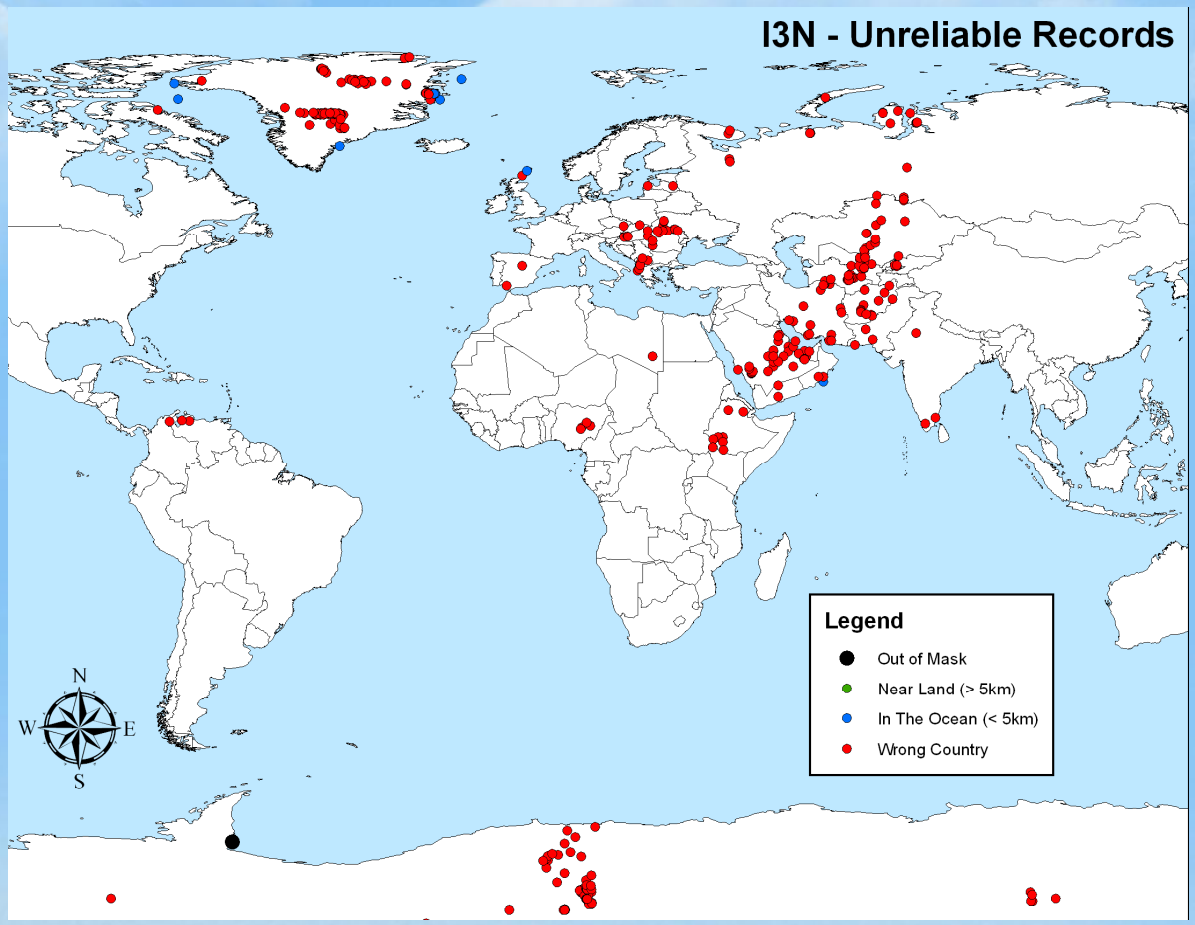


PTN Filtered Records



I3N Filtered Records

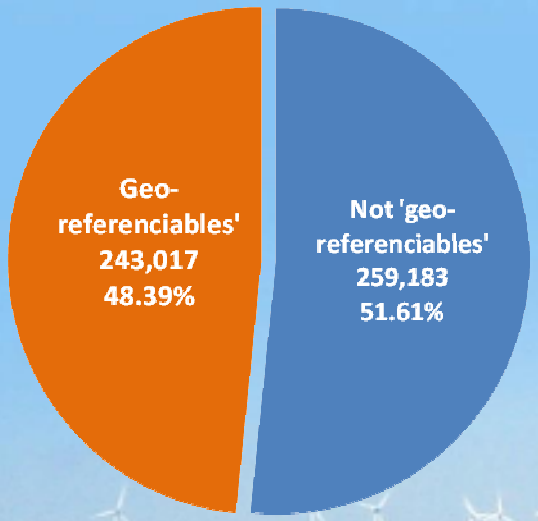
Unfortunately coordinates of I3N records didn't have the same format, indeed most information in this fields were not coordinates or were null. That make it hard to interpret geospatial information



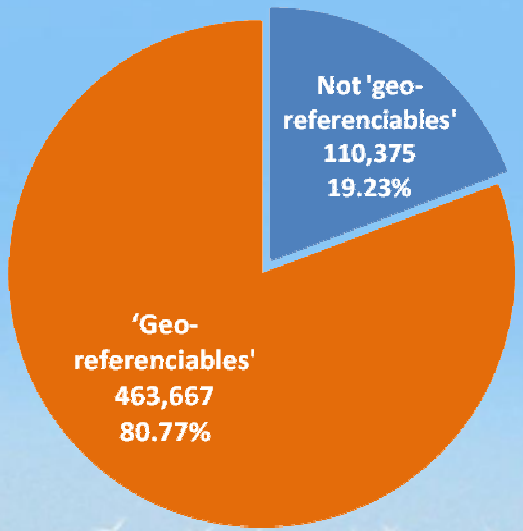
‘Georeferenciability’

The occurrences from the different datasets that have no coordinates, or have geospatial issues or any detected problem to make it excludable might be ‘georeferenciable’ if these occurrences records contain enough administrative information to find a possible geospatial location and include them for our analysis.

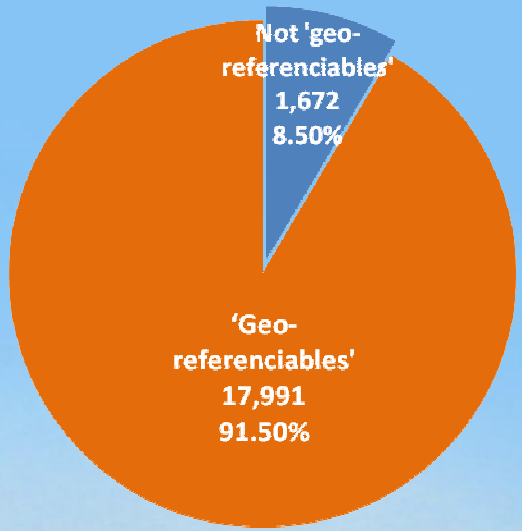
SSTN ‘Georeferenciability’



PTN ‘Georeferenciability’



I3N ‘Georeferenciability’



Georeferenciability current process

- SSTN Started on February 7 2011

SSTN		
Total	**Total Processed**	**Successfully Georefrenced**
243,017	35,289	11,440
100%	14.5%	4.70%

- SSTN Started on February 11 2011

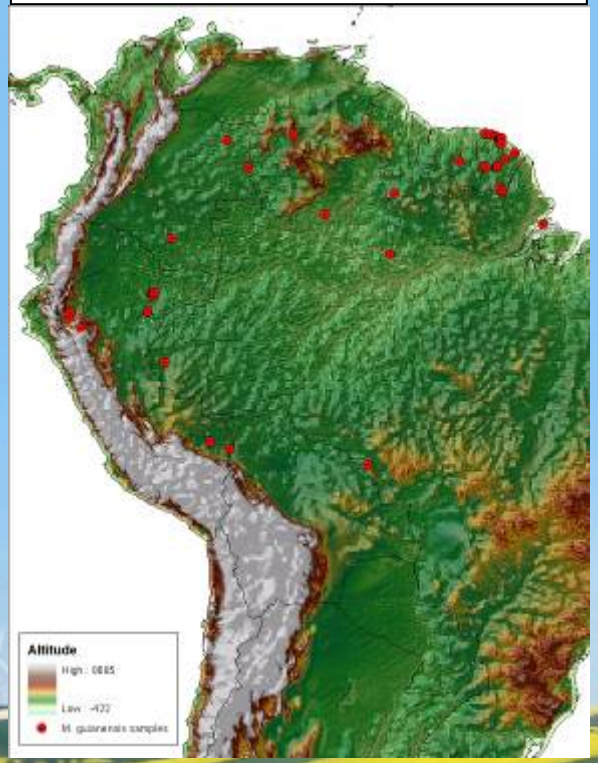
PTN		
Total	**Total Processed**	**Successfully Georefrenced**
463,667	13,613	3,819
100%	2.9%	0.8%

- I3N Not started yet

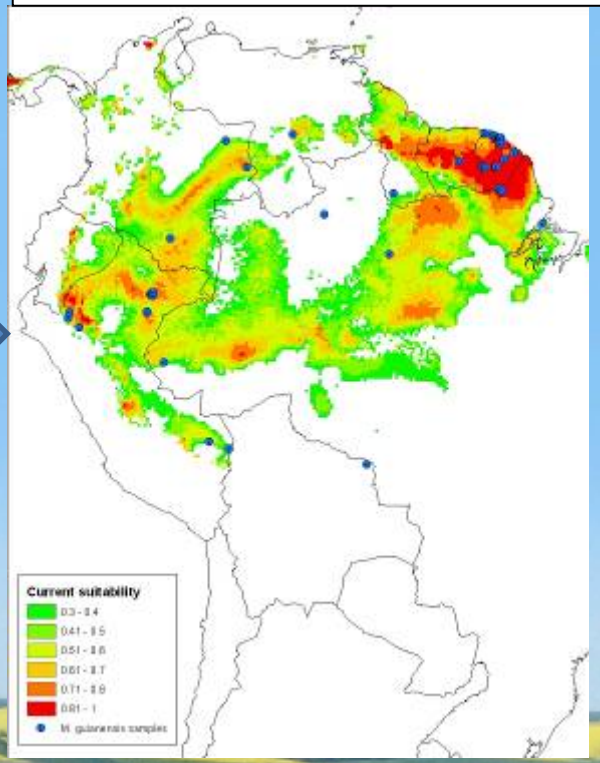
@Niche modeling: Training of niche models script

- Aplying the maximum entropy algorithm
  - *Macoubea guianensis* Aubl.: food for rural indigenous communities in the Amazon

Data harvesting

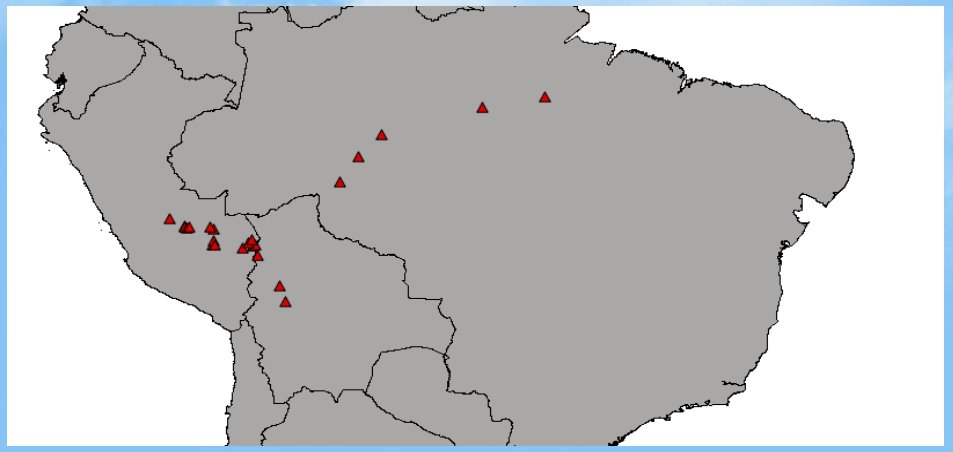


Building the presence model

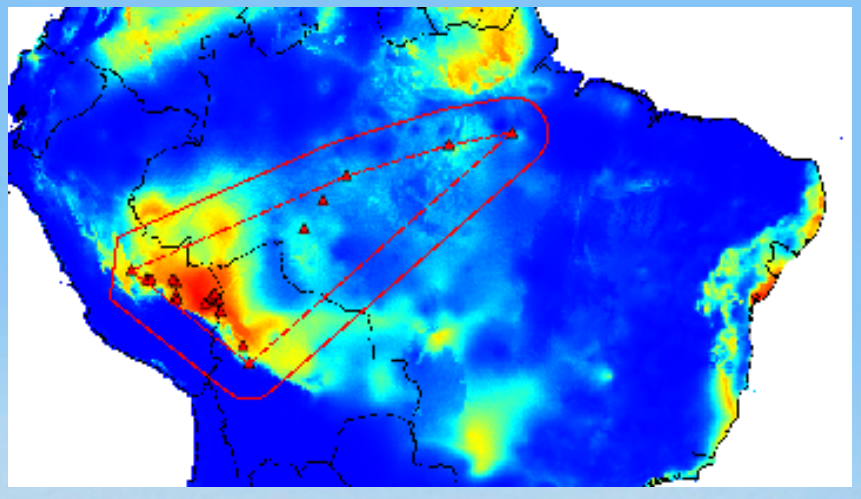


@Niche modeling: Assessing threats and current conservation actions

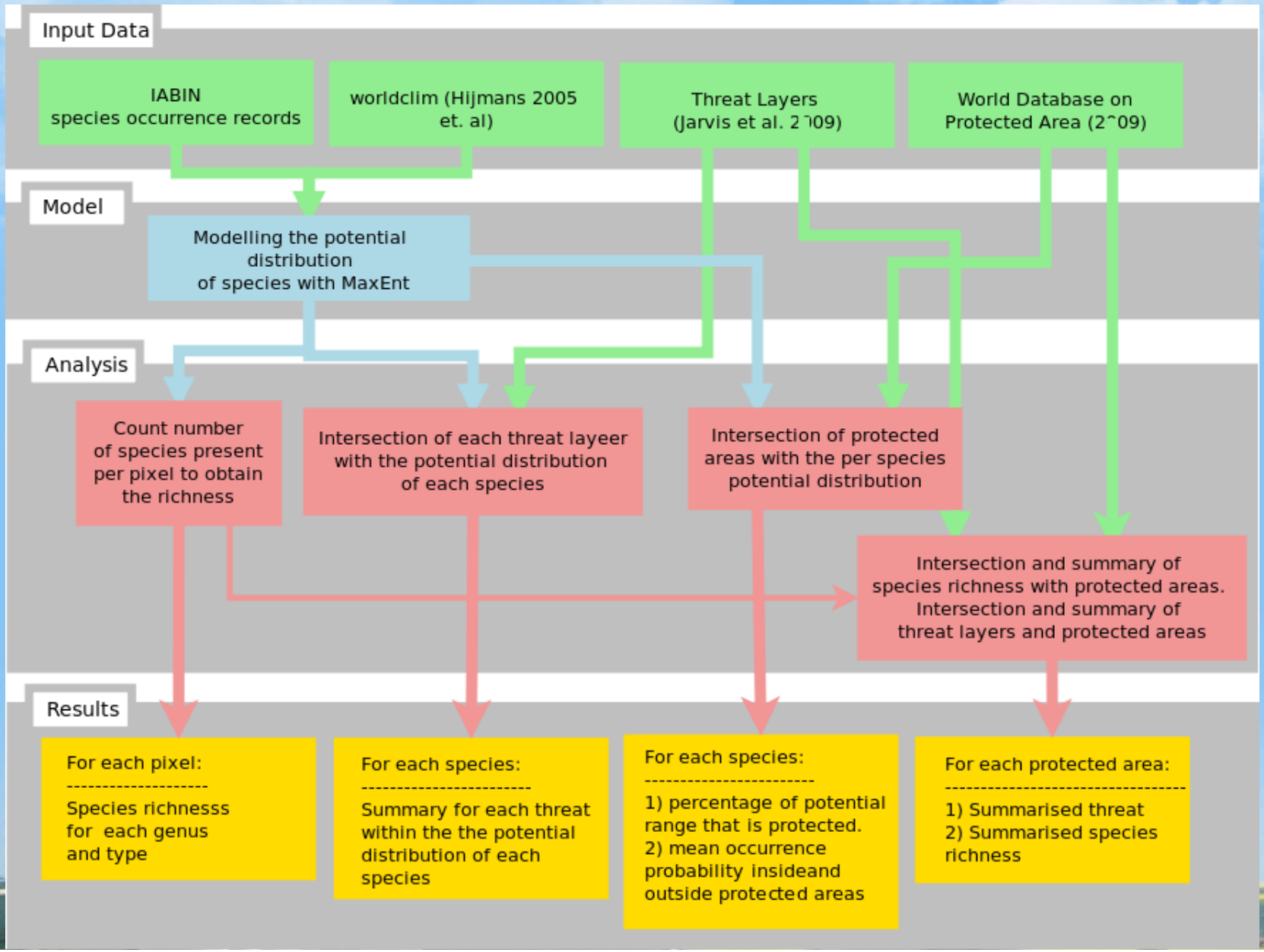
- Model the distribution of species with >= 10 records of presence in the IABIN database.



- Asses the distribution of each species in relation to:
  - Different threat scenarios.
  - Existing protected areas.



@Niche modelling: Workflow



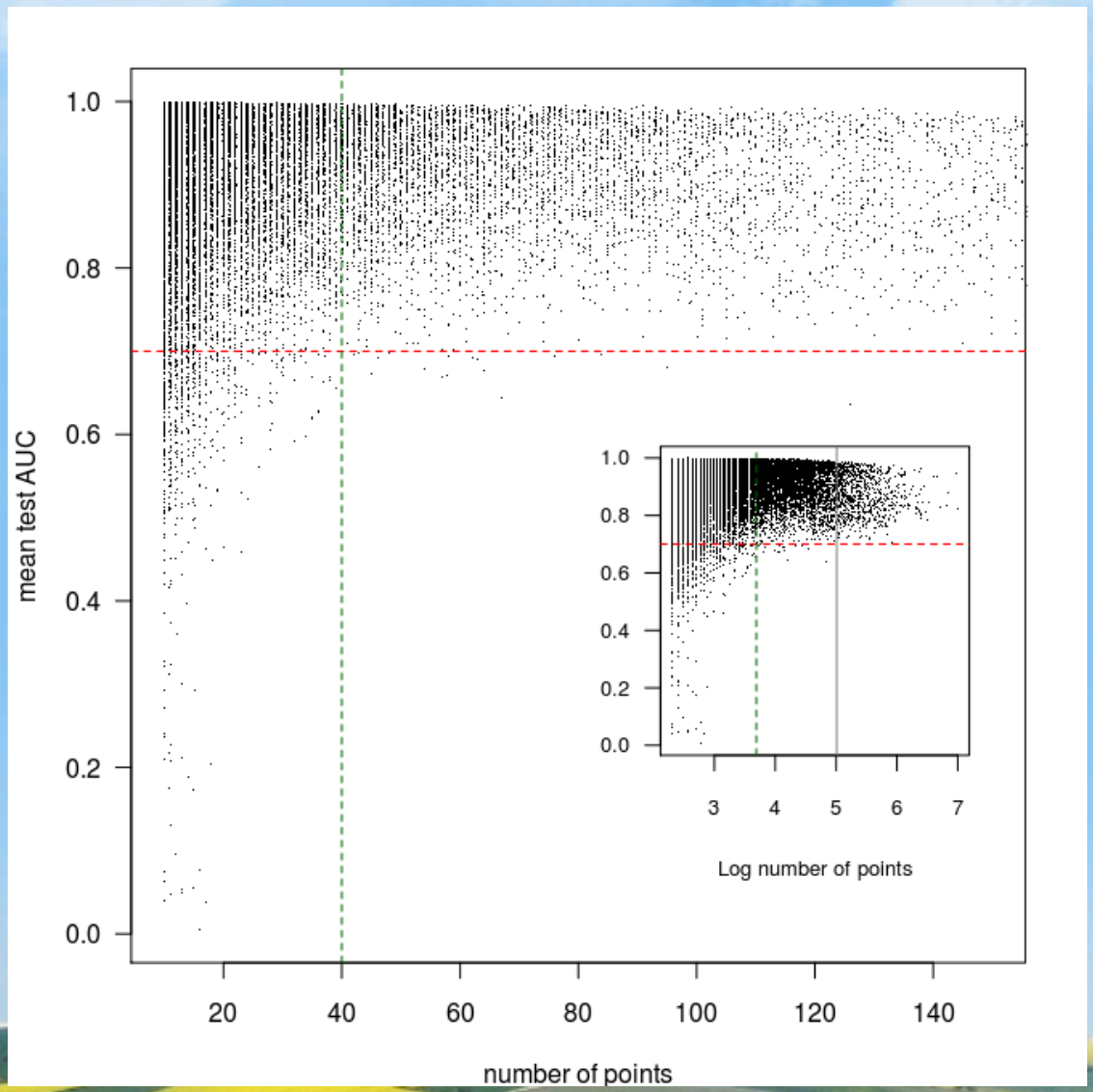
@Niche modelling: Task

- Model the potential distribution of all species with >= 10 unique points of occurrence in IABIN SSTN database and GBIF.
- Assess short and mid-term threats for each species within their potential range.
- Assess the conservation status (percentage of potential habitat protected for each species)

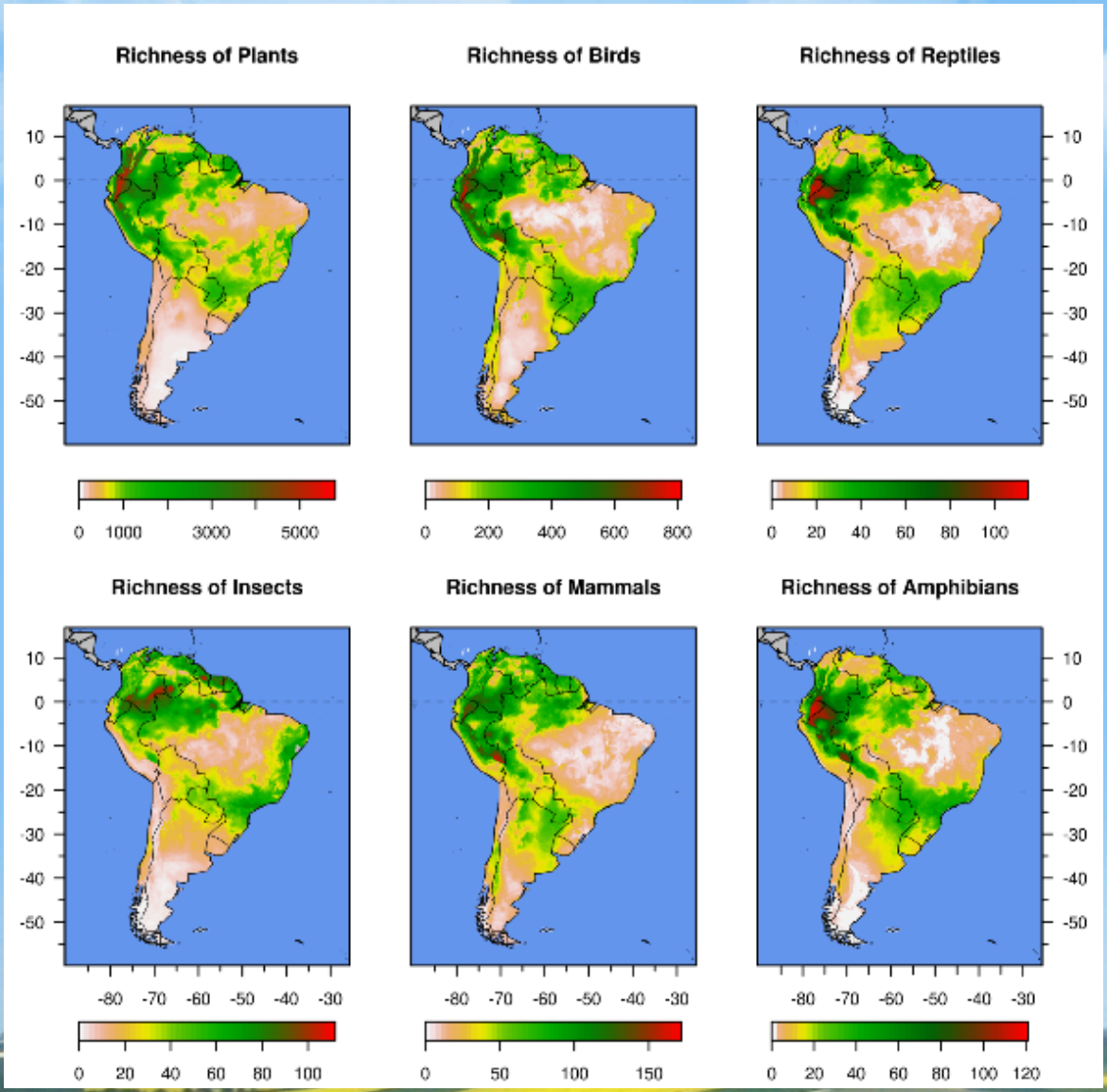
How many species were modelled?

Species	# Families	# Genera	# Species	# AUC > 0.7
Amphibia	18	79	400	384 (96.00 %)
Birds	93	750	2122	2044 (96.32 %)
Insects	37	169	474	407 (85.86 %)
Mammals	60	229	531	493 (92.84 %)
Plants	421	2566	15658	15225 (97.23 %)
Reptiles	28	113	309	295 (95.46%)
Total	657	3,906	19,494	18,848 (96.86 %)

Big issue: quality of SDMs



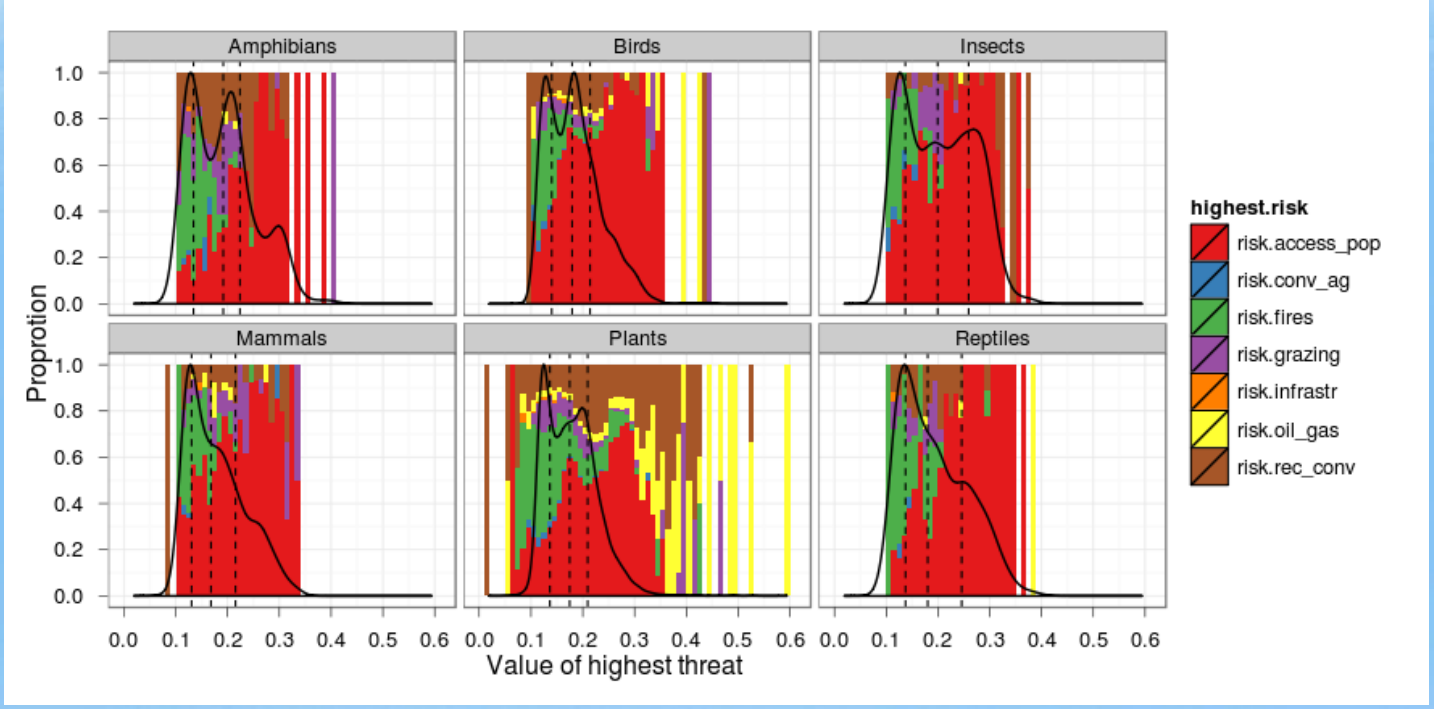
Where are species rich areas?



- Hotspots distribution and relevance depends upon species groups
- Andes: important diversity spot for all species and greatest in relation to all other spots
- Amazon: lack of data impedes from realising the existent richness
- Guyana shield and Brazilian Atlantic forest: less important than andes but still highly diverse, particularly for insects.

Most important threats?

A threat index relating the modelled distributional range (probabilities) and the threat distribution was created and each species was assessed



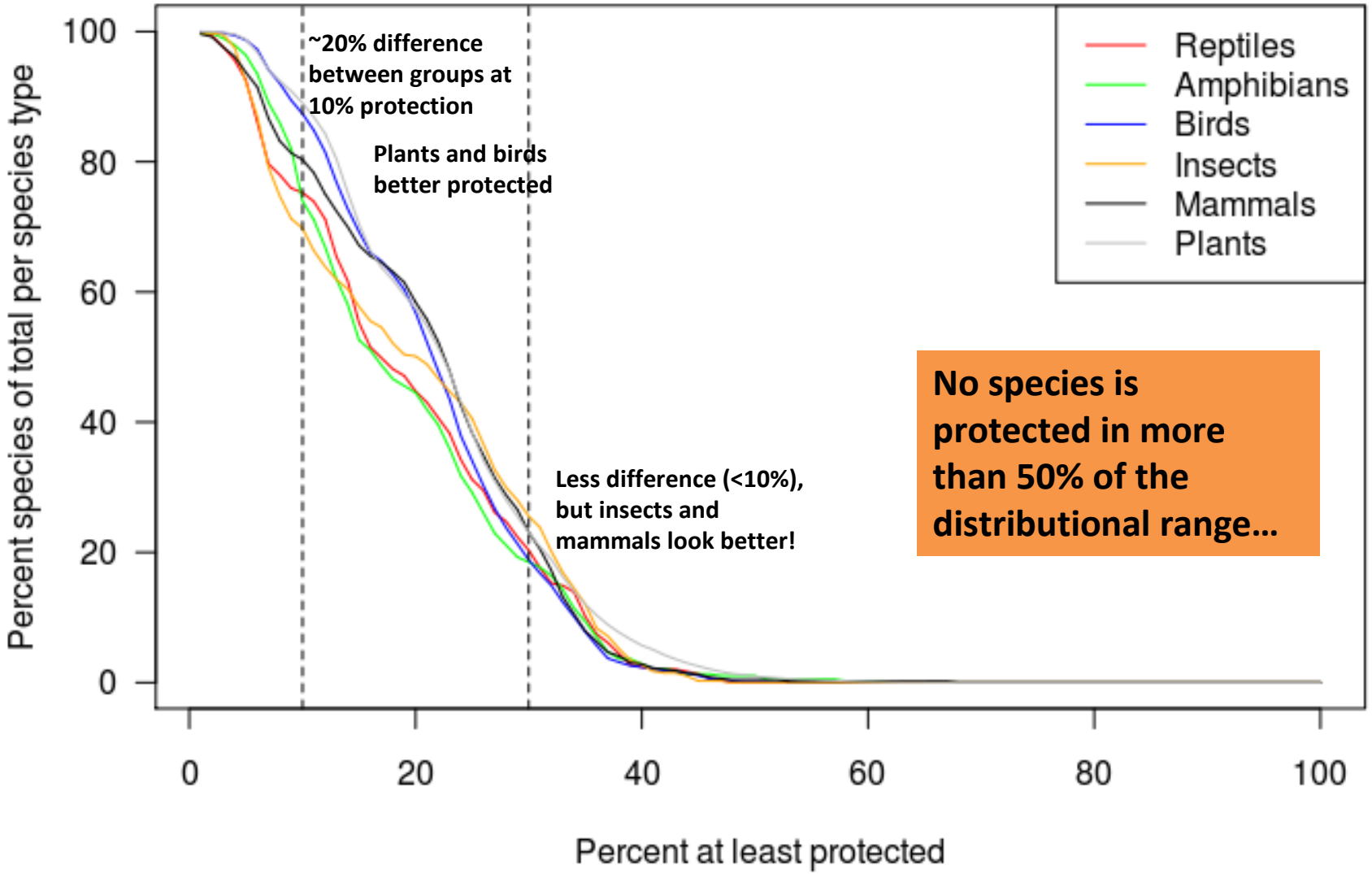
- For each species the most important threat is shown. The bold line indicates a rescaled species density and quantiles are shown by dashed lines.

- High threat is commonly associated to accessiblity, recent conversion (deforestation), and for plants also fires (probably due to a higher species density).

Conservation status

For each species type we calculate the percentage of area under each richness quartile that is protected.

Species type	First Quartile [%]	Second Quartile [%]	Third Quartile [%]	Forth Quartile [%]
Amphibians	15	18	20	27
Birds	19	14	19	27
Insects	17	14	19	29
Mammals	13	16	18	33
Plants	10	18	19	32
Reptiles	18	15	17	31


Percent species of total per species type
100
80
60
40
20
0
0
20
40
60
80
100
Percent at least protected
~20% difference between groups at 10% protection
Plants and birds better protected
Less difference (<10%), but insects and mammals look better!
Reptiles
Amphibians
Birds
Insects
Mammals
Plants
No species is protected in more than 50% of the distributional range…

@Niche modelling: Conclusion

- Species Distribution Models performed well, for some species types better (birds, plants) than for others (insects).
- Main threat to biodiversity are accessibility by population, recent conversion (deforestation) and fire.
- Protected areas are located in places with an above average species richness under current conditions.
- Species are generally well protected, next big question is whether they will work under future climates

@Navigation Tool

- Google-maps based
- Uses ready-to-use inputs such as PNGs, and KMLs
- Stand-alone, but easy to couple with the portal, if required
- Using Jquery, Maven, and other technologies suited for visualisation and data-manipulation purposes
- Displaying all the modelling results, as well as the primary data
- Easy transferability

@Navigation Tool

- Live Demo Available in

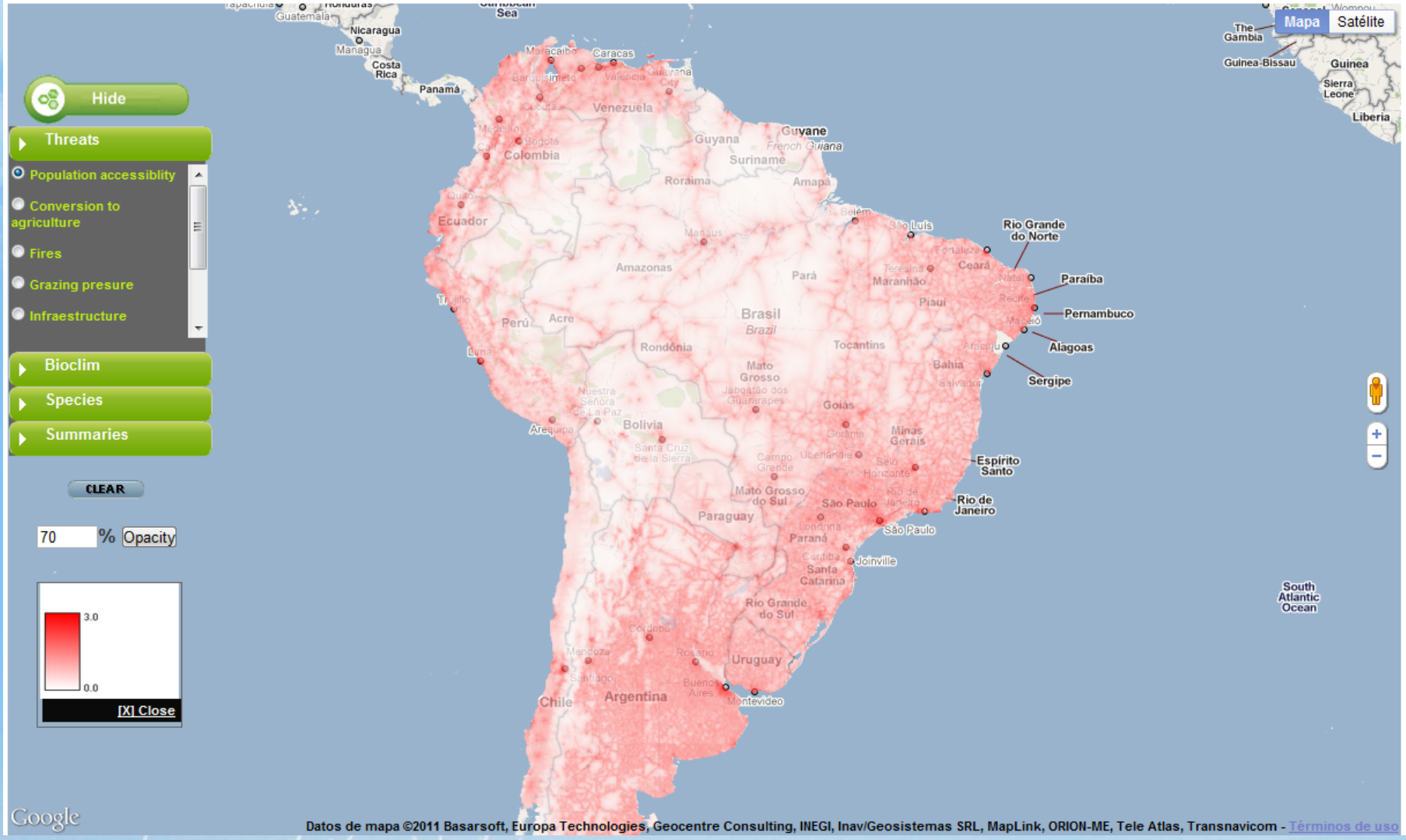
http://gisweb.ciat.cgiar.org/ita/menumap1.html

Currently in development

iabin
Red Interamericana de Información sobre Biodiversidad

Google-maps based navigation tool for all the modelling results

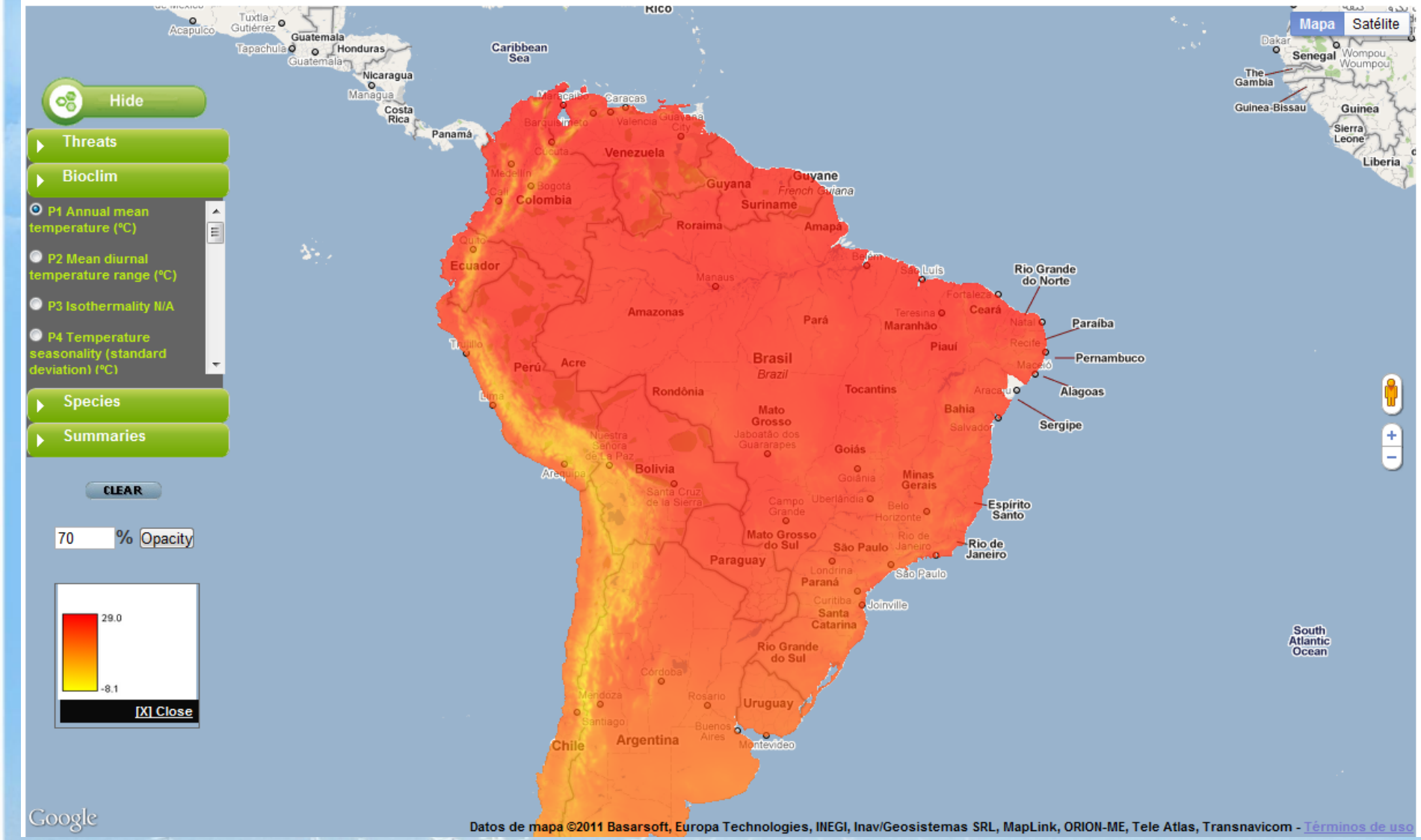
CIAT



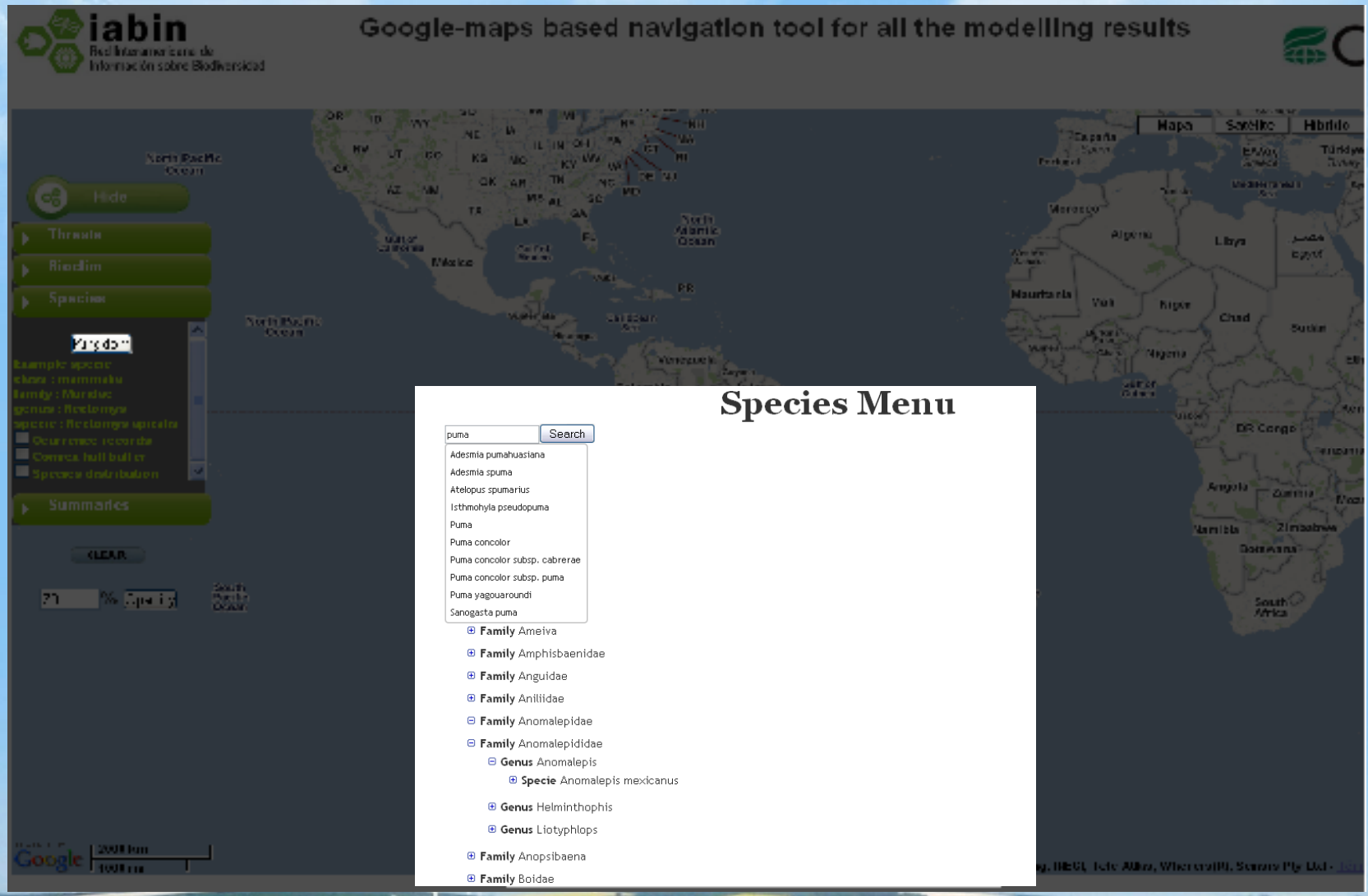
iabin
Red Interamericana de Información sobre Biodiversidad

Google-maps based navigation tool for all the modelling results

CIAT



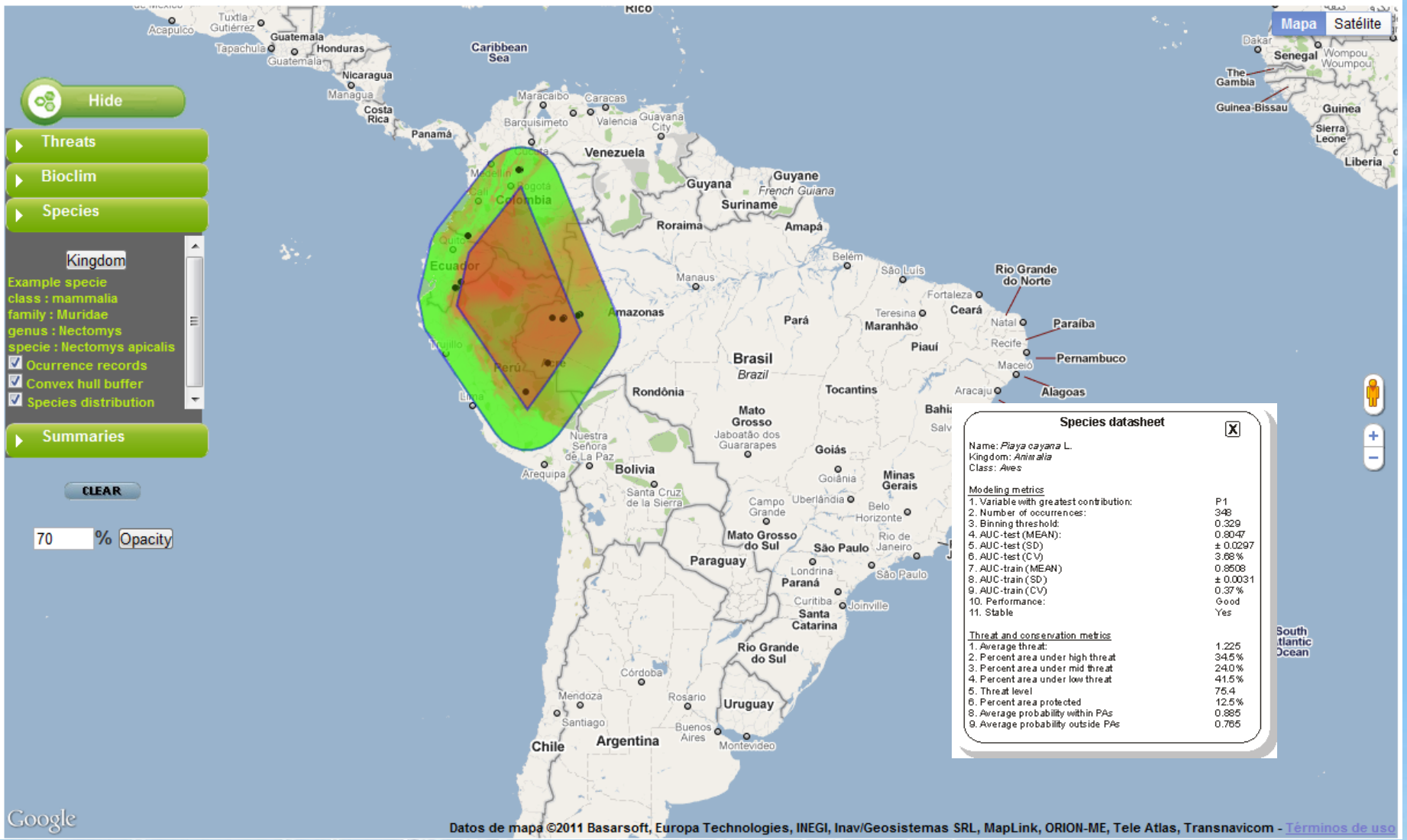
@Navigation Tool, Species search menu



iabin
Red Interamericana de Información sobre Biodiversidad

Google-maps based navigation tool for all the modelling results

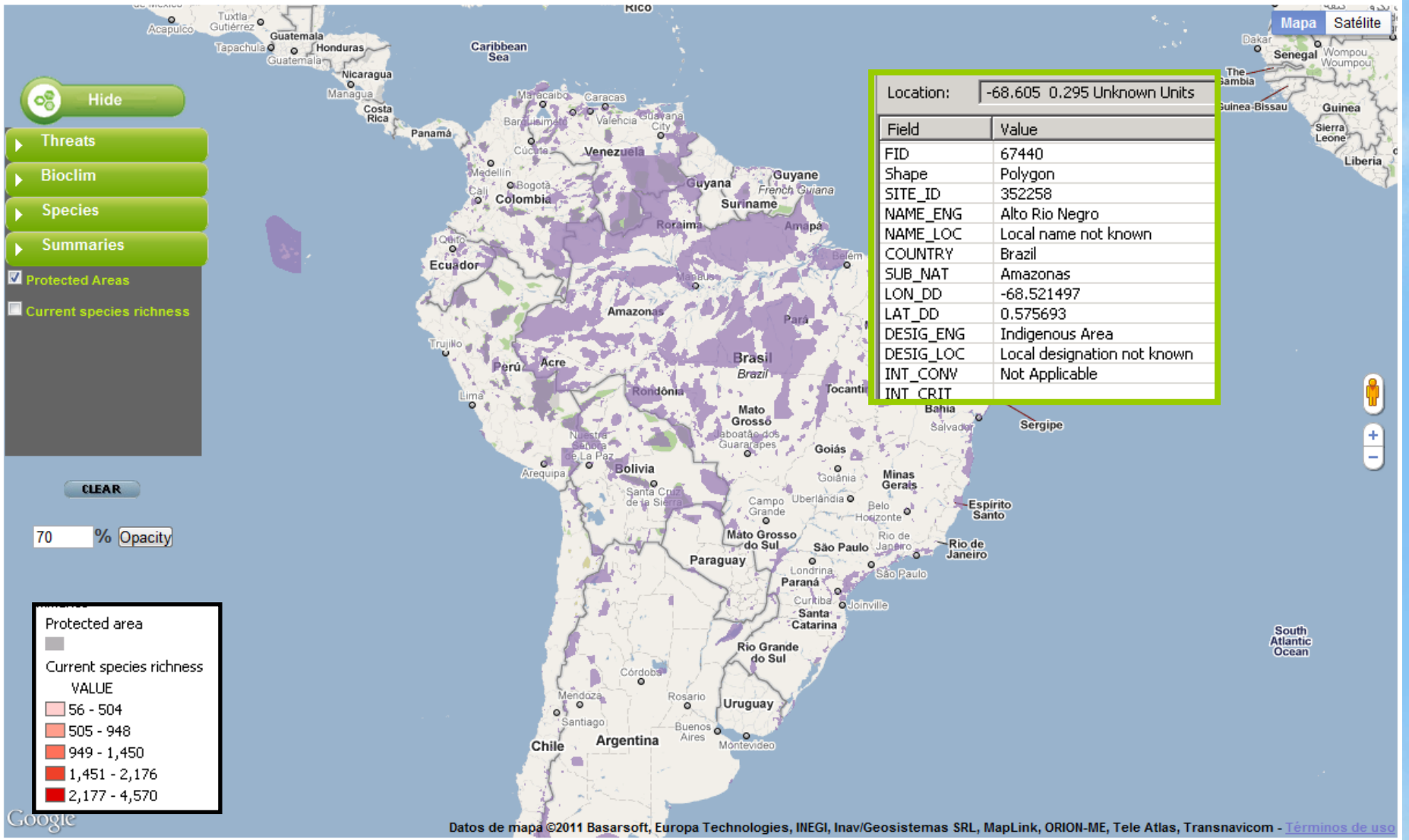
CIAT



iabin
Red Interamericana de Información sobre Biodiversidad

Google-maps based navigation tool for all the modelling results

CIAT



Summarising, in terms of data

- We have:
  - Assessed all three databases and delivered them to CBI
  - Developed automated scripts for cross-checking, georeferencing and species distribution modelling
  - Modelled ~19,000 species from ~3,900 genera
  - Assessed all species with regards of threat distributions and conservation and found that population accessibility, fires and recent conversion are the most important threats to species
  - Started and advanced with the development of a web visualisation interface.