(Sub-)mm data reduction and analysis
Direct detection

Alexandre Beelen – LAM
AtLAST-fr-days - May 13th 2025

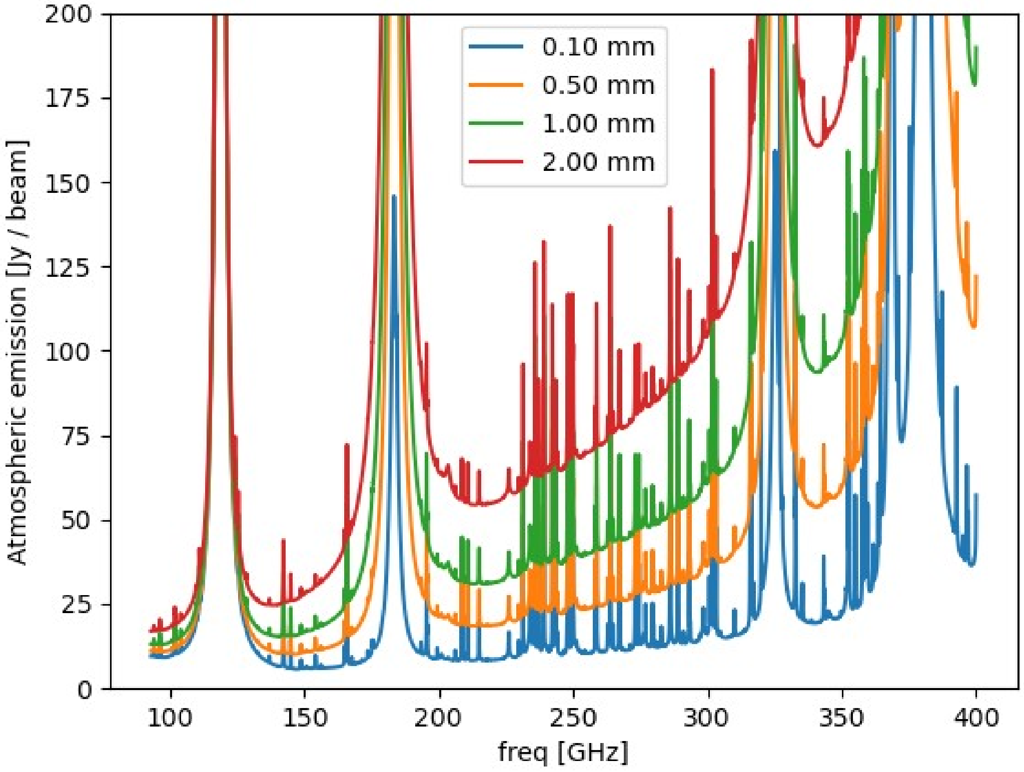
AtLAST wide-field Motivation

- **“To perform the deepest, widest (100 – 1000 deg^2), and most complete imaging (…) surveys” (Reichert+24)**
- **“A worked example for a 1000 deg2 continuum survey in 1000 hours was presented” (van Kampen+24)**
- **“[…] a Galactic Plane survey would need to cover up to 540 deg^2.” (Klaassen+24)**

(Sub)-mm Direct Detection Challenges

- Atmospheric Emission
- Instrumental/Electronic Drifts
- Telescope Observing Patterns
- Optical Aberrations
- Data Rate & Volume

Atmospheric Emission

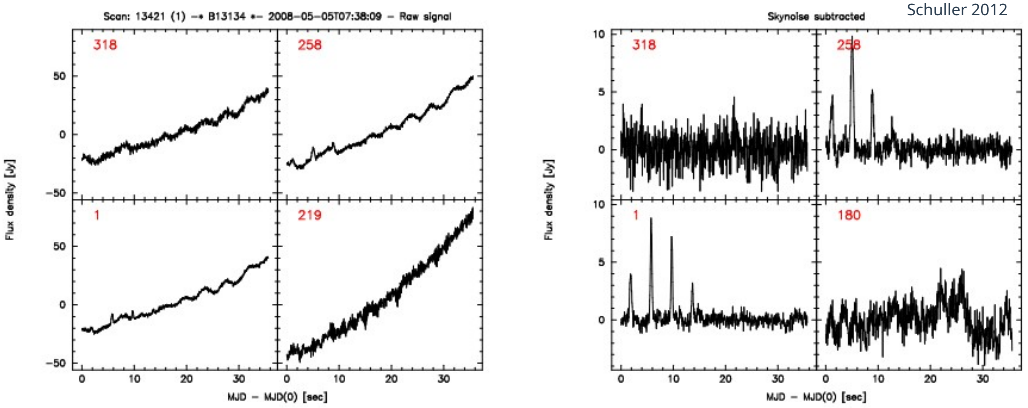


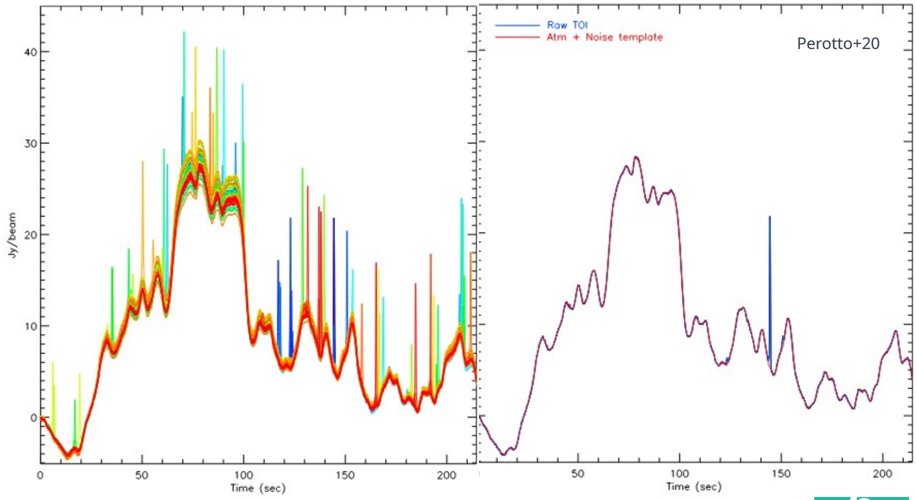
- **Main contributor :**
  - Water vapor is the dominant absorber/emitter in sub-mm/mm windows
- **Characteristics:**
  - Highly variable on timescales of s to min.
  - Sky brightness ∝ precipitable water vapor (PWV); often > 10 to 1000 times brighter than sources.
- **Mitigation:**
  - Observing from the highest, dryest sites
  - **Observing pattern with sky redundancy !**
  - Skydips, WVR-based corrections, atm. Models (Pardo+01, Paine+19).
- **Impact:**
  - Strong influence on raw noise power spectra and **correlated signal across detectors**.
  - Impact on calibration accuracy.

Atmospheric Emission Mitigation

- **Strongly correlated**
  - MedianNoise / *Most Correlated Pixels*
  - PCA/ICA
  - Template/model fitting
  - Scanamorphos
  - ML : UNIMAP/SANEPIC/TOAST/Minkasi
  - Inverse Problem (SUPREME/TAMASIS/…) …
- **Filters the sky**
  - MC simulations → Transfert function
  - All scales !
- **Trade-off**
  - Point Source vs Extended Emission
  - Iterative Processing could be needed…

Schuller 2012



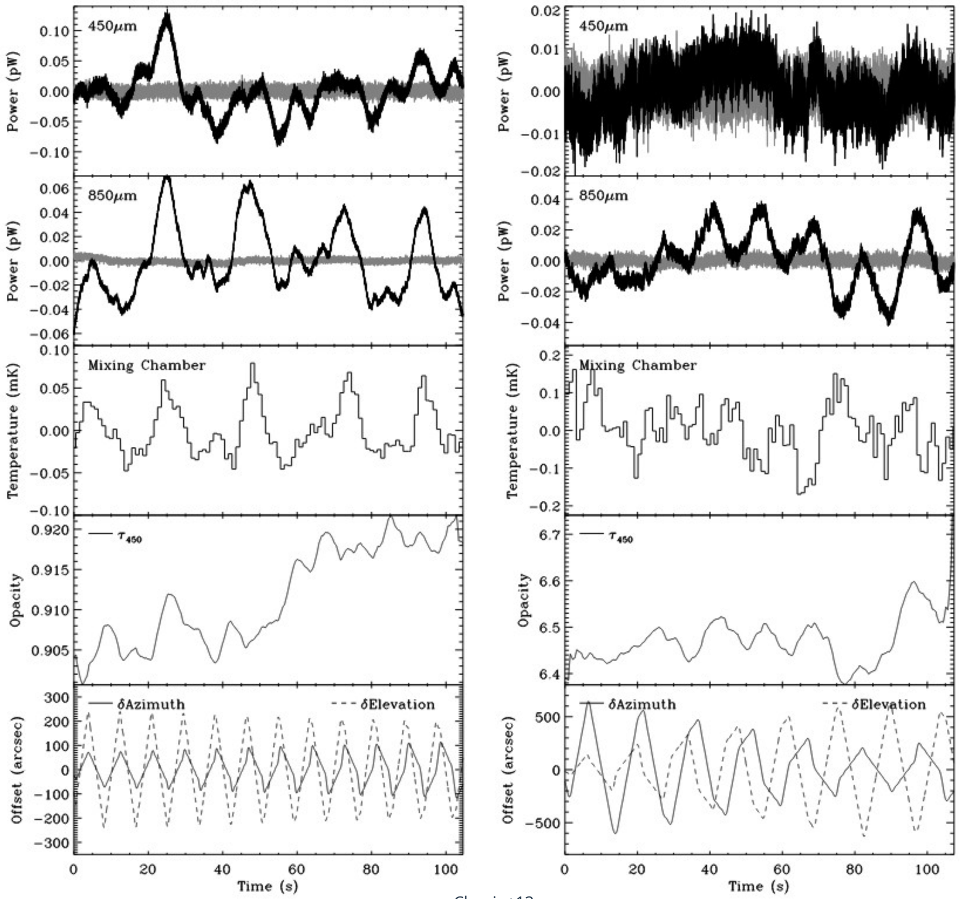


Perotto+20

Atmospheric Mitigation in the AtLAST Era

- **Large FoV and small PSF : 2° vs 1.5"**
  - Is the atmostphere uniform at those scale ?
  - Anomalous refraction
- **Multi-λ camera**
  - Can we properly use the multi-λ information ?
- **LOS PWV radiometer**
  - Several water lines ?
  - Several positions on the FoV ?

Instrumental/Electronic Drifts

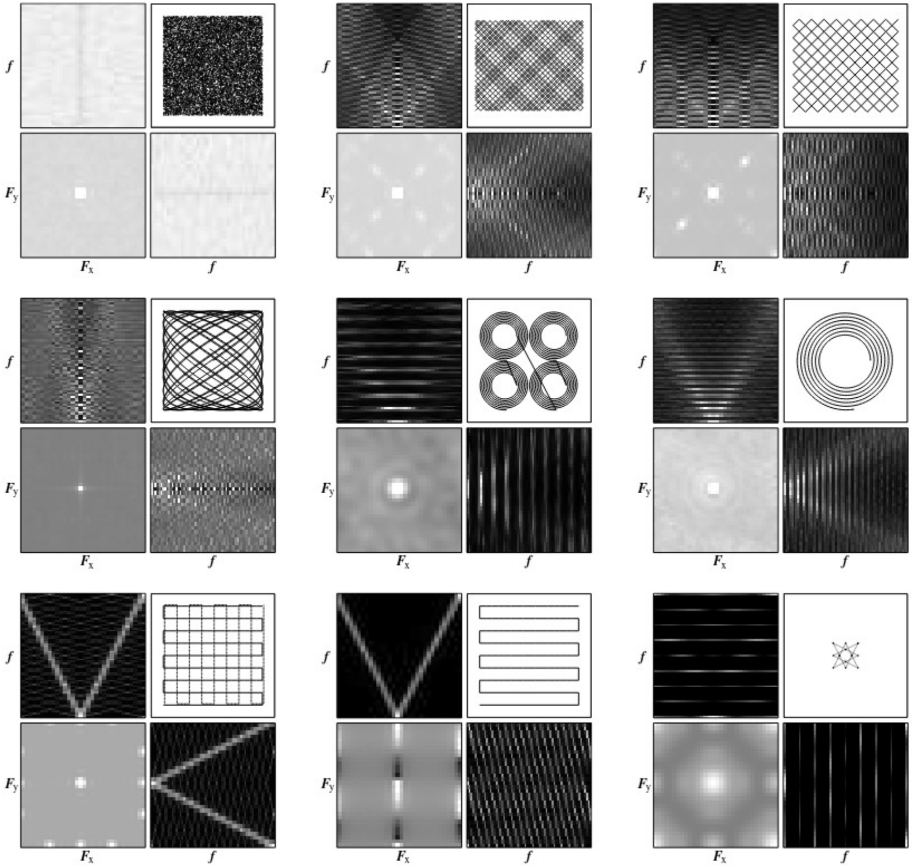


Chapin+13

- **Noise Sources:**
  - Electronics thermal fluctuations,
  - Readout instabilities,
  - Magnetic field pick-up, Varying optical load…
- **Manifestation:**
  - 1/f pink noise, slow drifts, scan-synchronous patterns.
- **Corrections:**
  - Iterative Common-Mode subtraction
  - Electronic boxes decorrelation
  - High-pass filtering / Template fitting
  - (Deeper) PCA/ICA
- **Even More Sky Filtering**

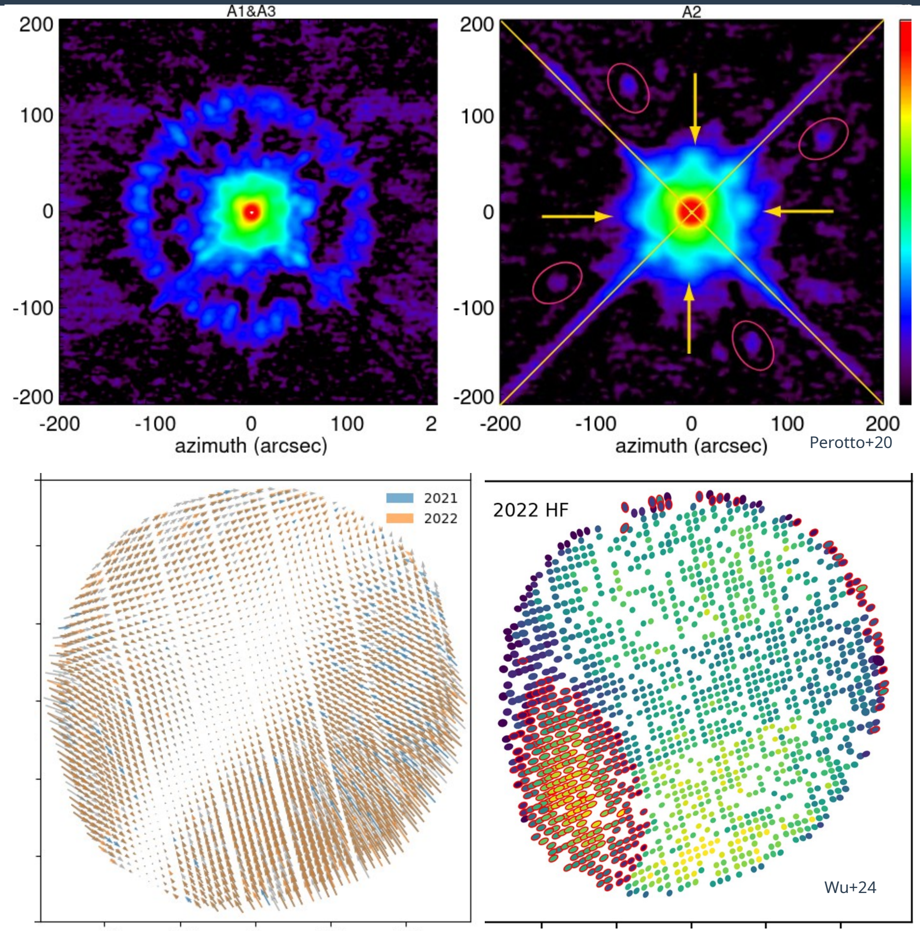
Single Dish Observing Modes

- **Optimizing for Sensitivity and Sky Coverage**
  - Raster scans, Spirals, Lissajous, on-the-fly (OTF) mapping
  - Maximize cross-scan directions
  - Constraints from atmospheric decorrelation time
  - Telescope accelerations
- **Sky filtering**



Kovács+08

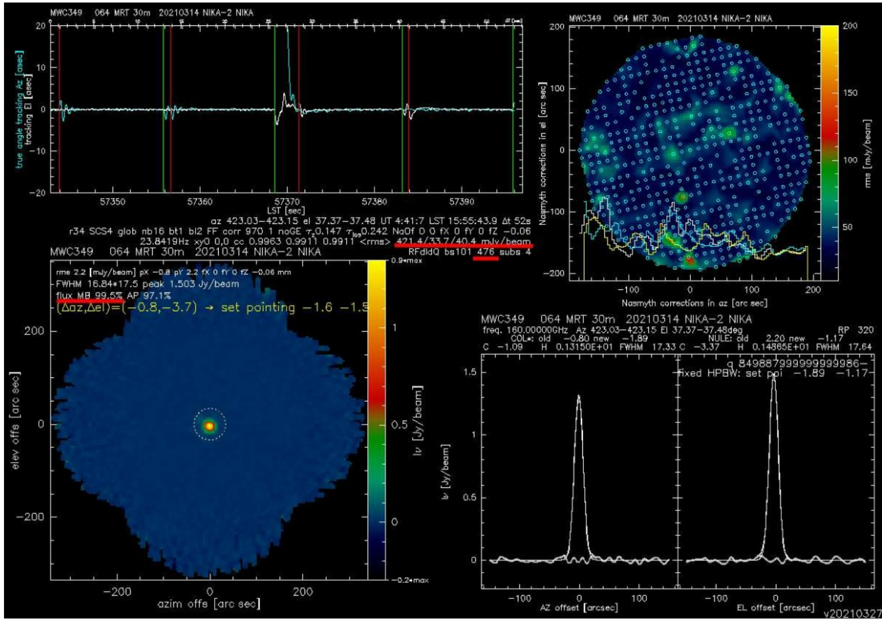
Single Dish Optical Aberrations



- **Optical Quality and Accurate Pointing**
  - Beam properties and calibration
  - Beam fidelity critical for photometry and spatial analysis
- **Aberration Types**
  - Coma, Astigmatism, Field curvature, Spherical aberration
  - Elevation dependant gain
- **Mitigation Strategies:**
  - Best possible optical design
  - Calibrated effect

Online vs Offline Data Processing

- **Online processing for operational decisions**
  - Pointing scans / Focus scans
  - Monitoring weather and instrument status
  - **Real or near real-time** processing
  - Integration with the telescope control system
  - Edge computing is necessary (Commissioning)
- **Offline processing**
  - High fidelity data reduction for science-grade products
  - Allow for parameters tuning depending on the science case
  - Coadding of multiple scans, iterative processing
  - **Calibrated reproductible products**



NIKA2 QL, Berta & Zylka 2025

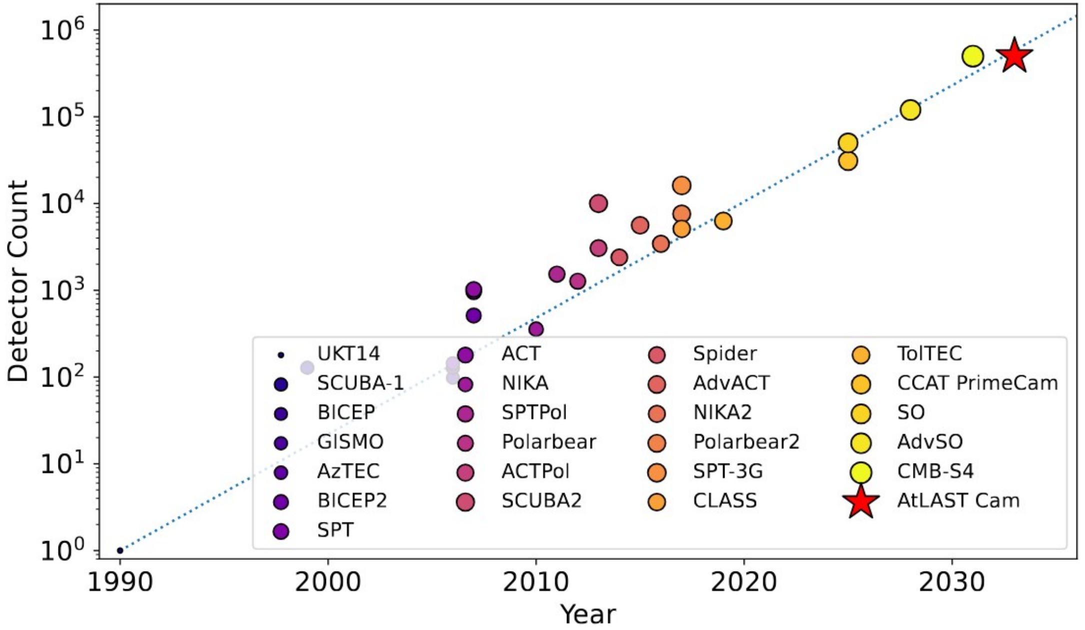
Raw Data Rate & Volume

- **MAMBO2 117 detectors : ~ 1 MiB / h**
- **SCUBA2 : 4 MiB/s, 12 hours → 100 GiB**
- **SPT-3G : 1.2 TiB /day (compression ~ /5) → 14 MiB /s**
- **NIKA2 : ~18 MiB/s (+lossy compression /5)**
- **CONCERTO : ~116 MiB/s (+lossy compression /5)**
- **TolTEC : ~30 MiB/s or 2 TiB per observing night**
- **(LSST: 20 TiB / night = 480 MiB/s, SKA 700 PiB /yr = 23 GiB / s)**

- **Data Rate and Volume are/were manageable**
  - Data compression is sometime mandatory
  - Disks at the telescope are sometime necessary

AtLAST L0 Date & Volume

- **Scanning speed**
  - 1.5 ~~to 3~~ deg/s
- **Telescope Resolution**
  - 1.5” @350µm to 10” @2mm
- **Data Sampling (¹/₃ beam)**
  - 2 to 11 kHz
  - Downsampling ? & **detector time response !!**
- **Data throughput (~~10^6 pix~~. 300k@ 32 bits)**
  - **2 to 12 GiB / s** : ~2 to 11 dedicated 10 Gbps optical fibers
    - 130 MiB / s if downsampled to 120 Hz
  - **1000 h → 6 to 41 PiB** : could we even keep the full raw data ?
    - **471 TiB** if downsampled to 120 Hz



AtLAST Science Overview Report 2024

L2 Data Volume

- **1000 deg^2** @350µm
  - Pixel sizes of 0.5”x0.5” → ~200 GiB x 3 (data/weights/hits)
- Repeat and project that scans many times → x a few thousand
- And/or MonteCarlo Simulation → x 100 to 1000
- **> 1 PiB of storage…**

- **Distributed storage and distributed processing will be mandatory**
- Need for huge temporary storage → Storage policy
- Shared Common Infrastructure : AtLAST DataCenter !

Data Reduction Pipeline Architectures

- **From distributing L0 to L2/L3 data ?**
  - Avoid infrastructure duplication,
  - Open architecture,
  - Centralized knowledge base,
  - **Cost & Energy efficient,**
  - Q/A on distributed data products,
  - Reproductability,
  - Offline vs fully Online processing ? (prevent iterative processing)

Data Reduction Pipeline Architectures

- **Based on past expertise**
  - MAMBO2/SCUBA2/LABOCA/AzTEC/(SPT)/SPIRE/NIKA2/TolTec/ …
  - Comparable processing / recipies, but not scalable to AtLAST
  - Need to handle instrument specific issues
- **Towards a full observatory pipeline**
  - Data reduction pipeline, with open contributions
  - Resilient to data or hardware failures
  - Up to the L3 data product (catalogs / P(k) / …) ?
  - **Simulations are a key ingredient**

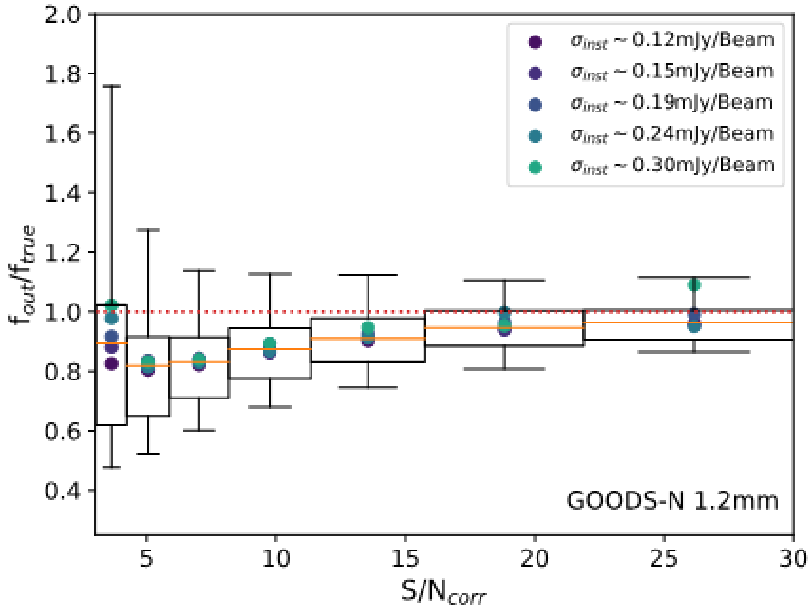
L3 Data Products

- **Any product based on maps :**
  - Transfert functions
  - Insure proper usage of the observatory products
- **Point Source Catalogs**
  - Blind / Prior
  - Galaxies / SF Core
- **Extended Structure Detection**
  - SF Filaments
  - SZ Clusters
- **High order statistic (?)**
  - Power spectra
  - n-pts correlation functions

Simulations

- **Full instrument & pipeline transfert function**
  - *MUST* for most observations
  - Realistic simulated sky
  - Fast(er than observations)
- **Map/Catalog characterization**
  - Purity
  - Completeness
  - Flux bias

Bing+23



Thanks