PSA activities @ IP2I Lyon:
Scanning data analyzed with machine learning techniques

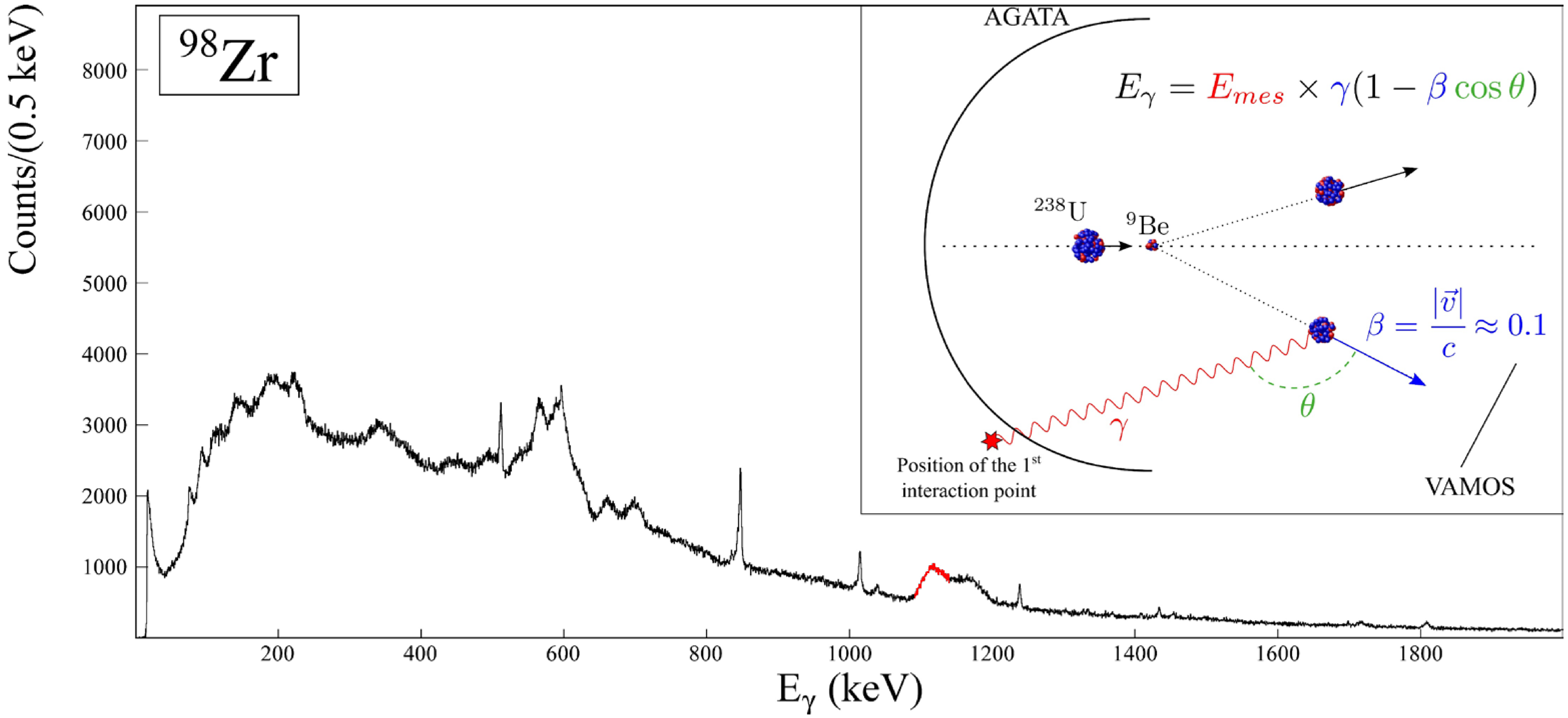
Mojahed Abushawish's PhD work, presented by Jérémie Dudouet

Institut de Physique des deux infinis de Lyon **(IP2I)**

AGATA week 2024, Milano

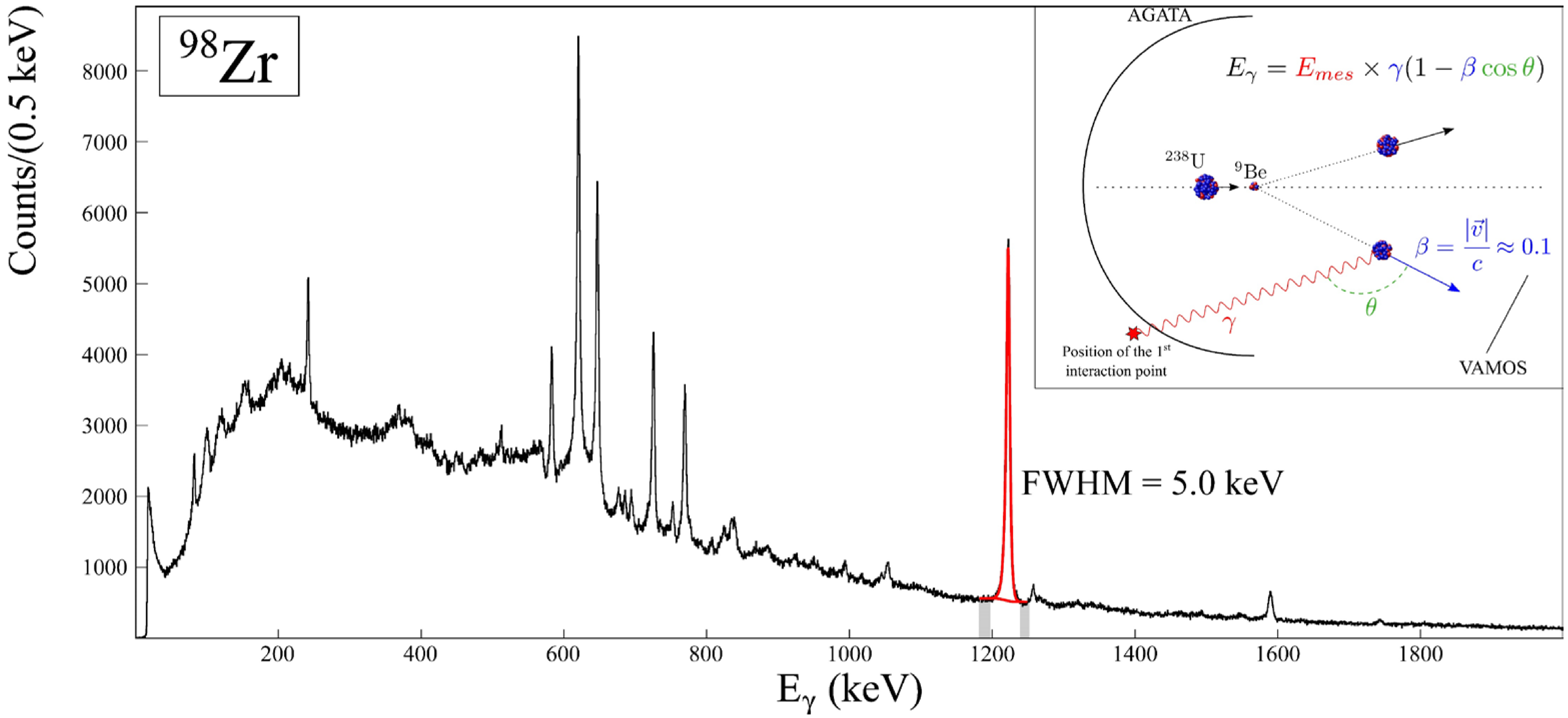
Why AGATA is the best detector of the world ?

➤ Because of its position resolution !



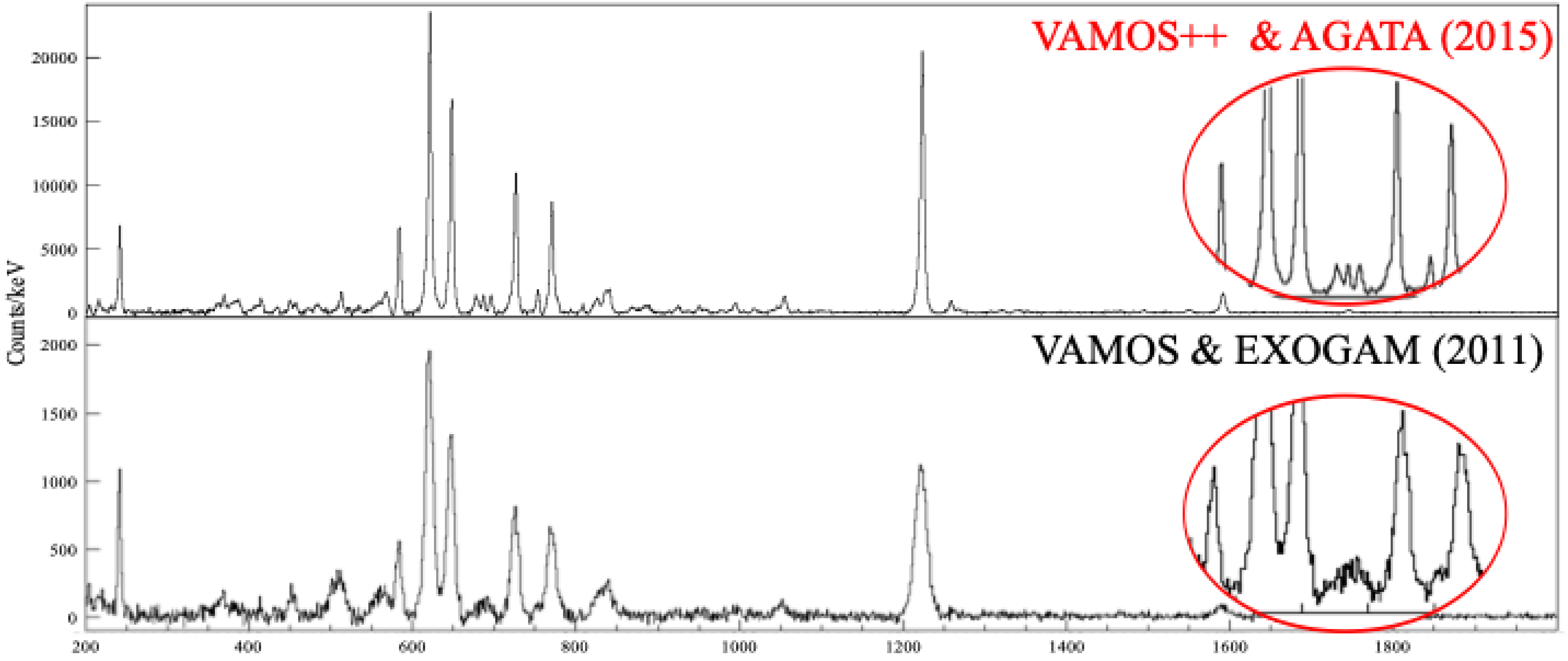
Why AGATA is the best detector of the world ?

➤ Because of its position resolution !



Why AGATA is the best detector of the world ?

➤ Because of its position resolution !



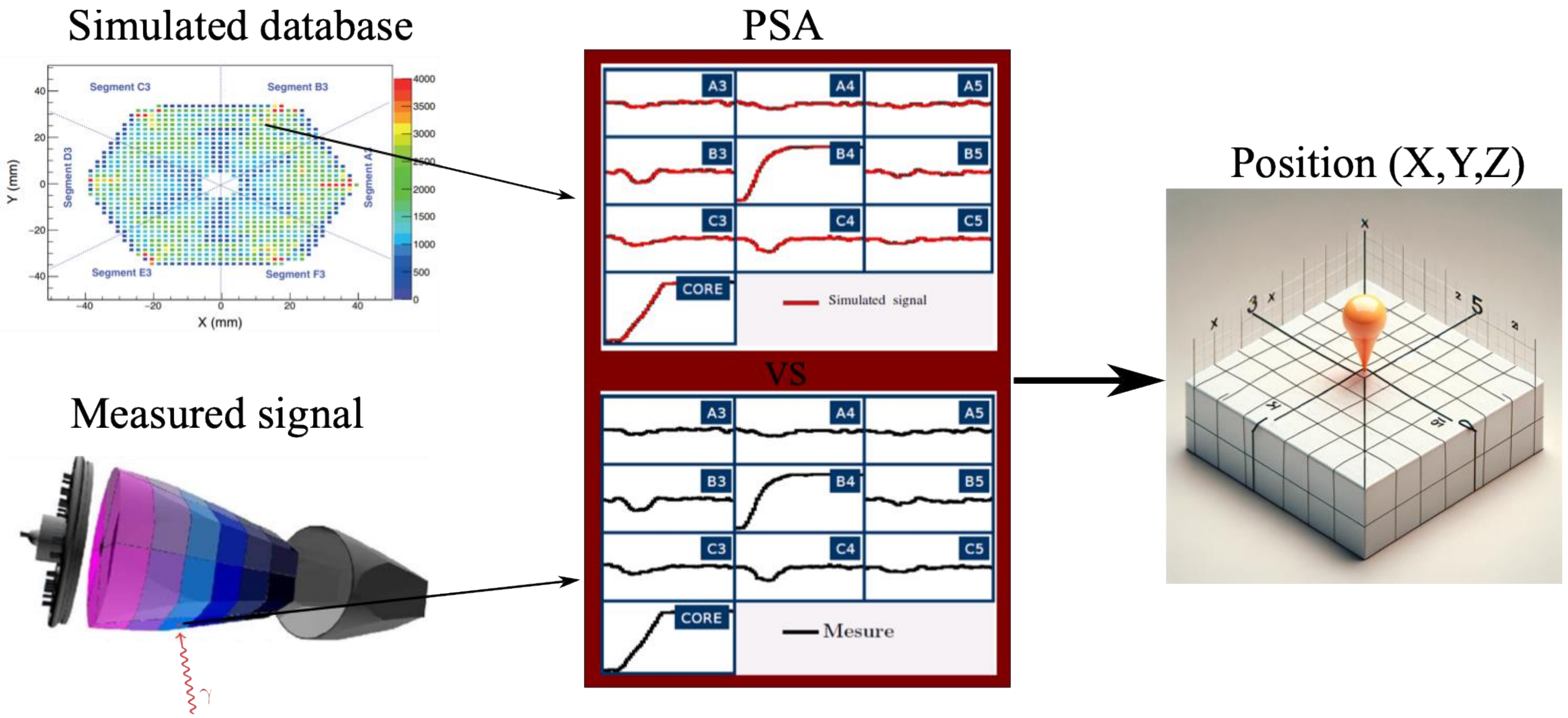
Excellent Doppler correction quality !

How is obtained this position resolution ?

➤ The Pulse Shape Analysis algorithm (PSA)

Simulated databases available:

- ➤ Agata Detector Library (ADL):
  - ➡ Used by default online.
  - ➡ No more ADL experts in the collaboration.
- ➤ AGATAGeFEM:
  - ➡ Results as good as ADL have been shown
  - ➡ Recently developed by J. Ljungvall



What can be improved ?

➤ Is the PSA working ? YES

What can be improved ?

➤ Is the PSA working ? YES

➤ Can we characterize its performances ?

➥ Is the position resolution uniform in the crystal volume ?
➥ What can be improved ? (PSA algorithm, Simulations)

What can be improved ?

➤ Is the PSA working ? YES

➤ Can we characterize its performances ?

➥ Is the position resolution uniform in the crystal volume ?
➥ What can be improved ? (PSA algorithm, Simulations)

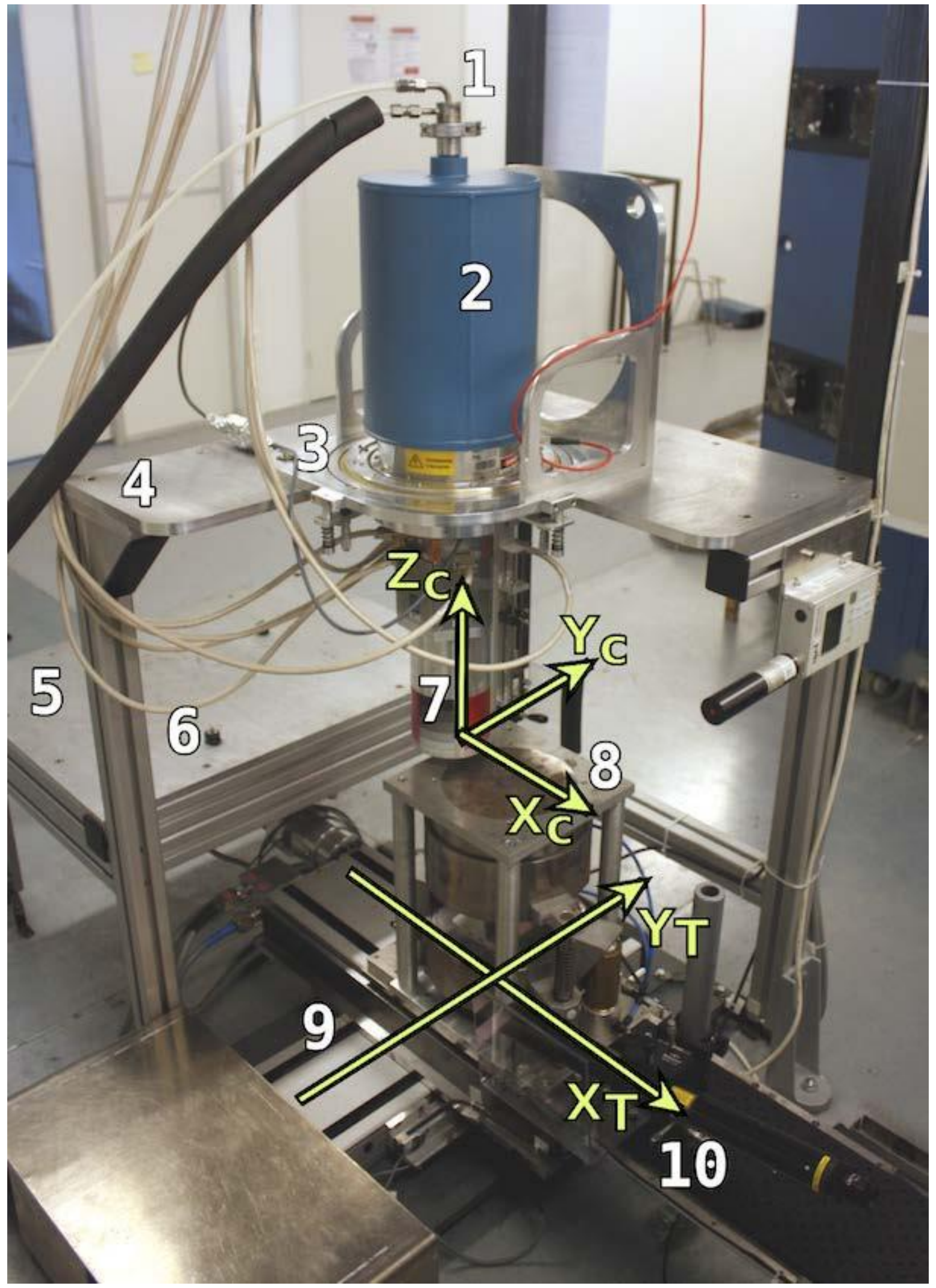
YES, with a scanning table !

The Strasbourg scanning table

- **Scanning capabilities:**
  - motorized collimator with a precision of 10 µm
  - system allowing the placement of the detector in vertical and horizontal position
  - laser alignment system
  - digital electronic (TNT2)

- **Scanning concept:**
  - not performing a real 3D scan (too long), but two 2D scans (vertical and horizontal)
  - 3D basis obtained by Pulse Shape Comparison Scanning (PSCS) method

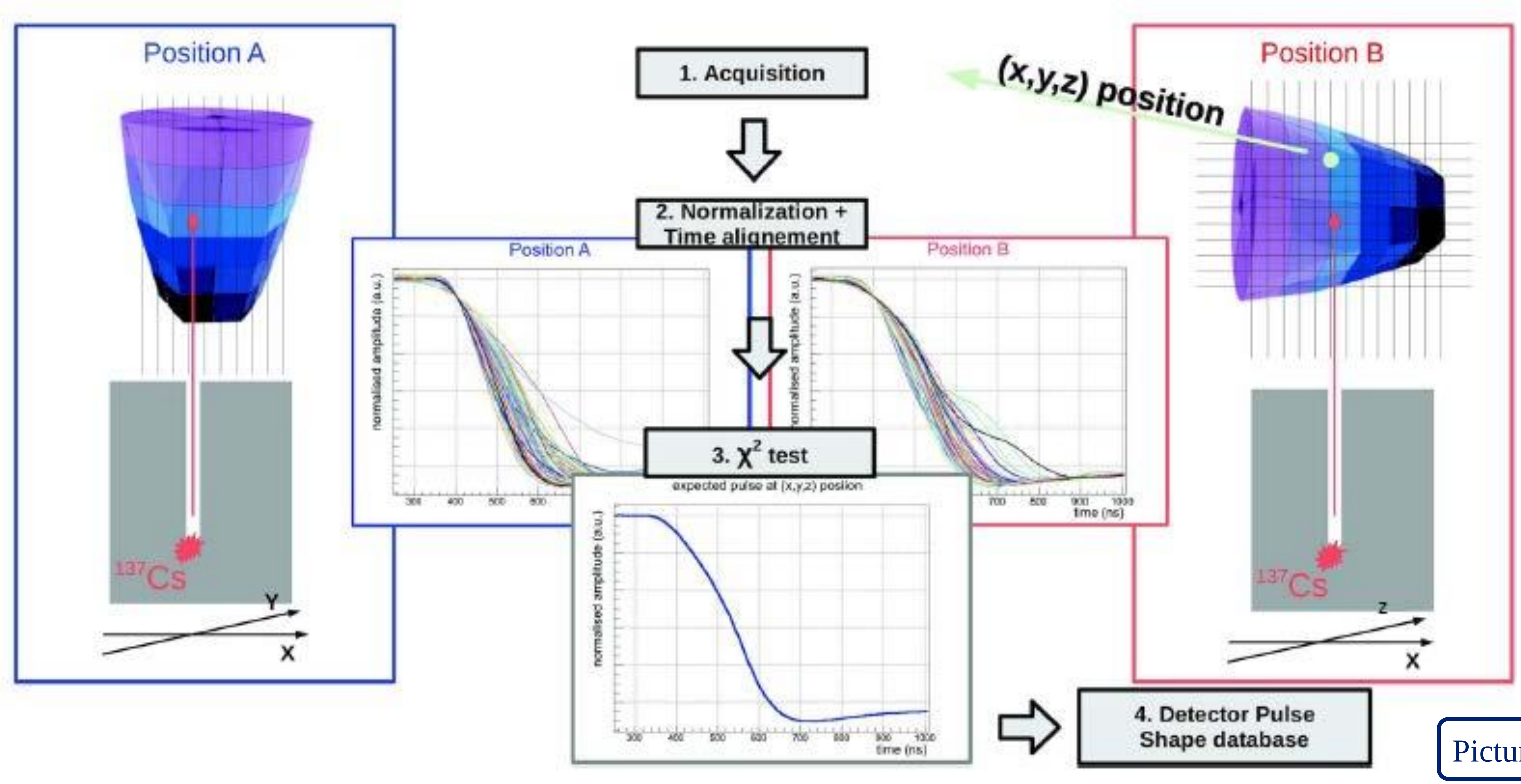
- **Detector scanned:**
  - S001: a prototype symmetric detector
  - A005: scan finalized this summer



The PSCS principle

- ➤ 1 horizontal scan + 1 vertical scan,
- ➤ the 3D basis is obtained by a combined analysis of both data-sets.
- ➤ Validated and published method, but time consuming (5 days for the PSCS analysis)

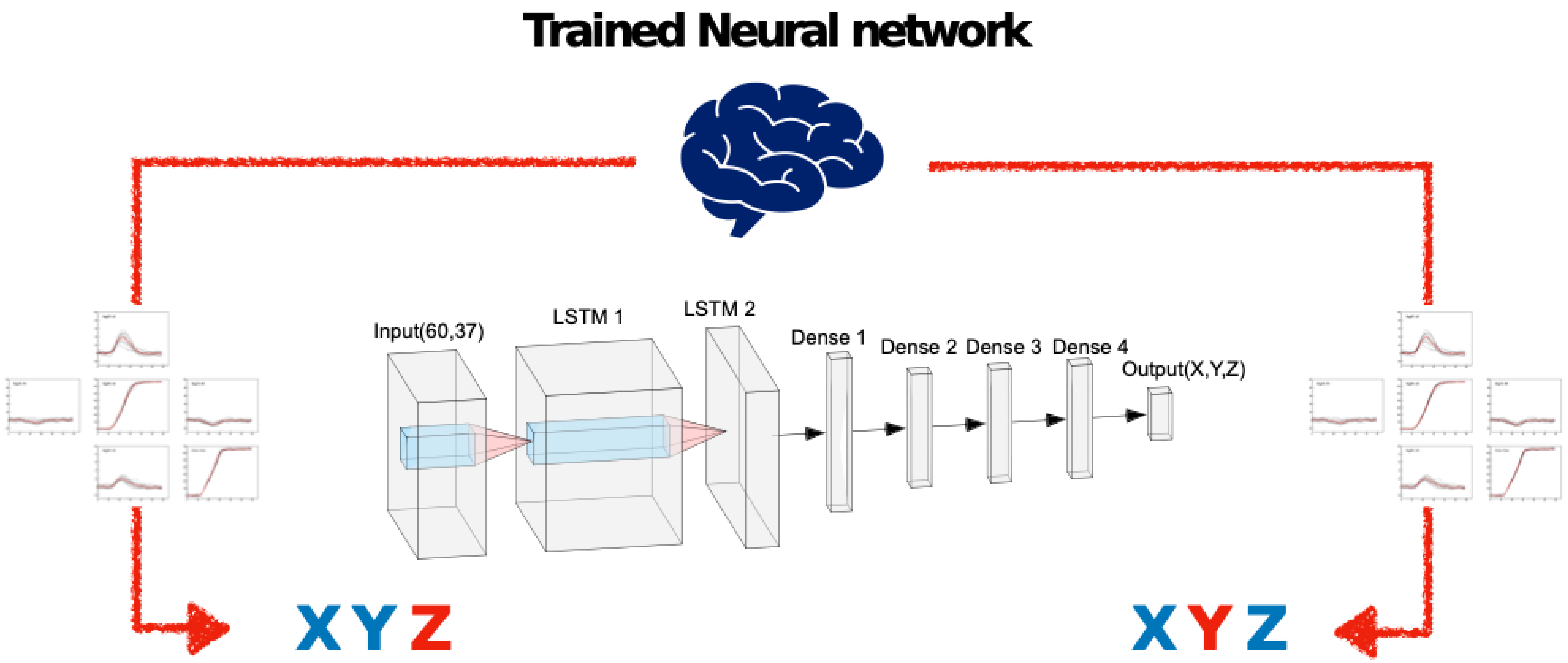
B. De Canditiis et al., Eur. Phys. J. A 57 (2021), B. De Canditiis and G. Duchêne, Eur. Phys. J. A 56 (2020)



Picture from Michael Ginsz's PhD thesis

New method proposed @ IP2I based on neural networks

- 2 Long short-term memory (LSTM) layers were used as starting point.
  - ➥very robust against time misalignment
- 4 dense layers are added to obtain a 3D output (X,Y,Z)
- The loss function is calculated only for the two known axes
  - ➥ this allows the network to learn patterns of each dataset without affecting the other.



Data processing

➤ **To be as close as possible to the data taken online with AGATA, a complete analysis chain has been developed:**

➤ Raw data are converted from TNT2 to ADF using the new *Scanning Table Data Processing* (STDPro) package
- ➥ produce compressed ADF raw data, including scanning meta data
- ➥ produce basic histograms for processing checks and calibration

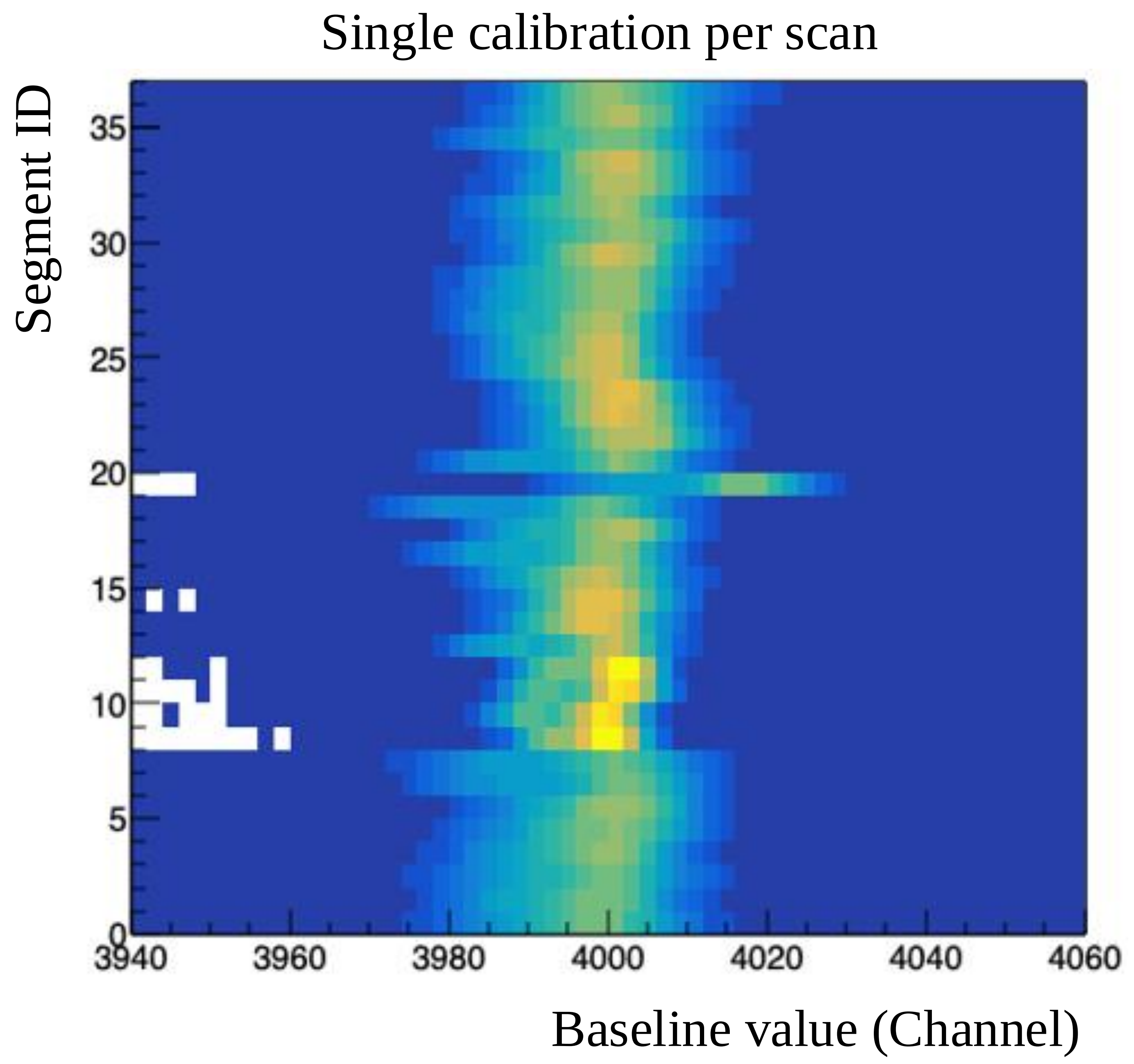
➤ Standard AGATA tools are then used to calibrate the scanned data:
- ➥ traces and energy calibrations (Cubix)
- ➥ x-talk and time alignment (RecalEnergy)
- ➥ data stored to ROOT Trees and calibrated ADF files (after preprocessing)
- ➥ processing is done using the FEMUL emulator

➤ The processing is dispatched using the IP2I SLURM farm (more than 3000 runs to be processed in ~10 minutes)
- ➥ batch processing system based on a docker image containing all AGATA software

Data processing

➤ Baseline fluctuations (critical for comparing traces):

➡ Not possible to use one baseline calibration for a full scan

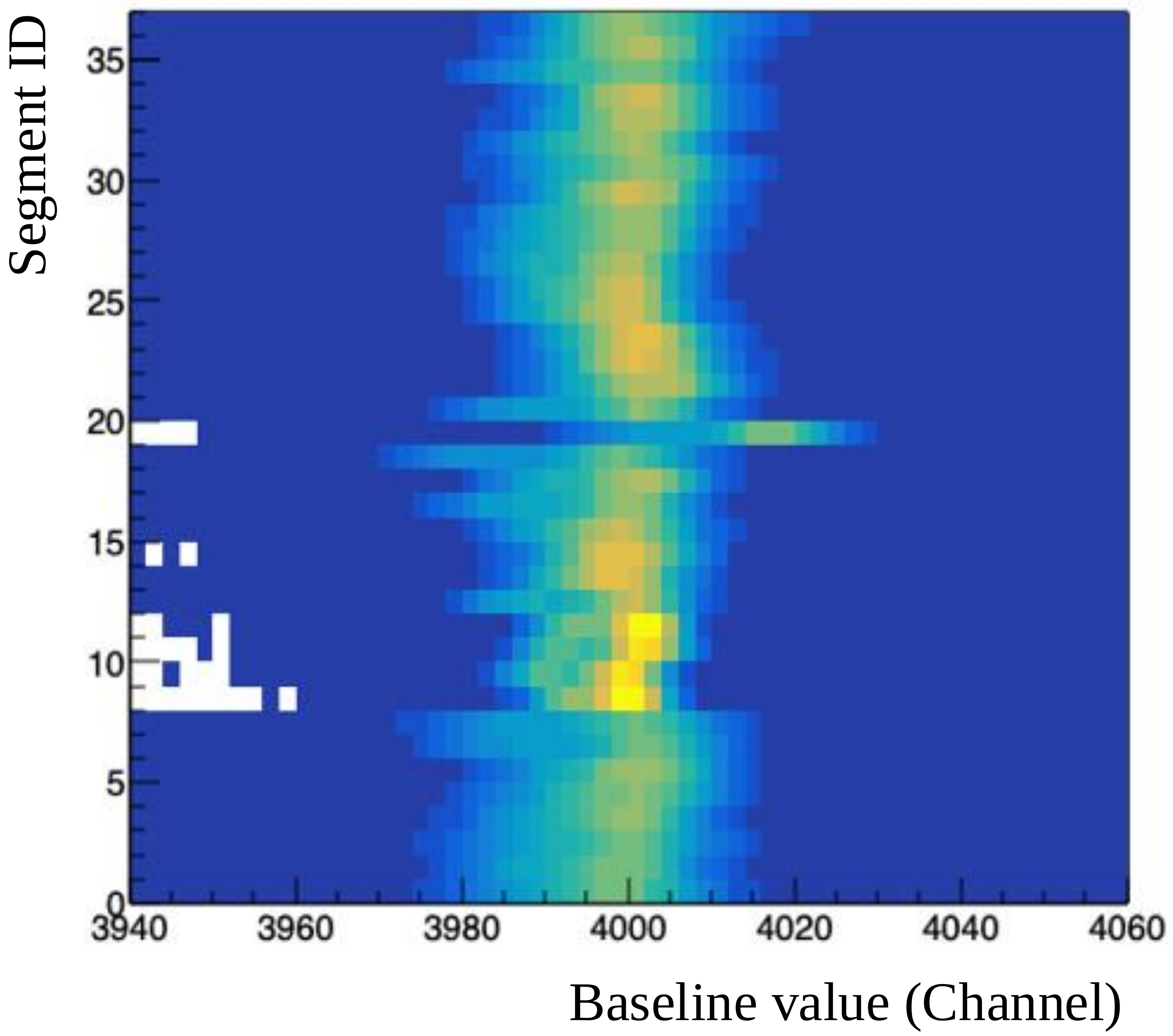


Data processing

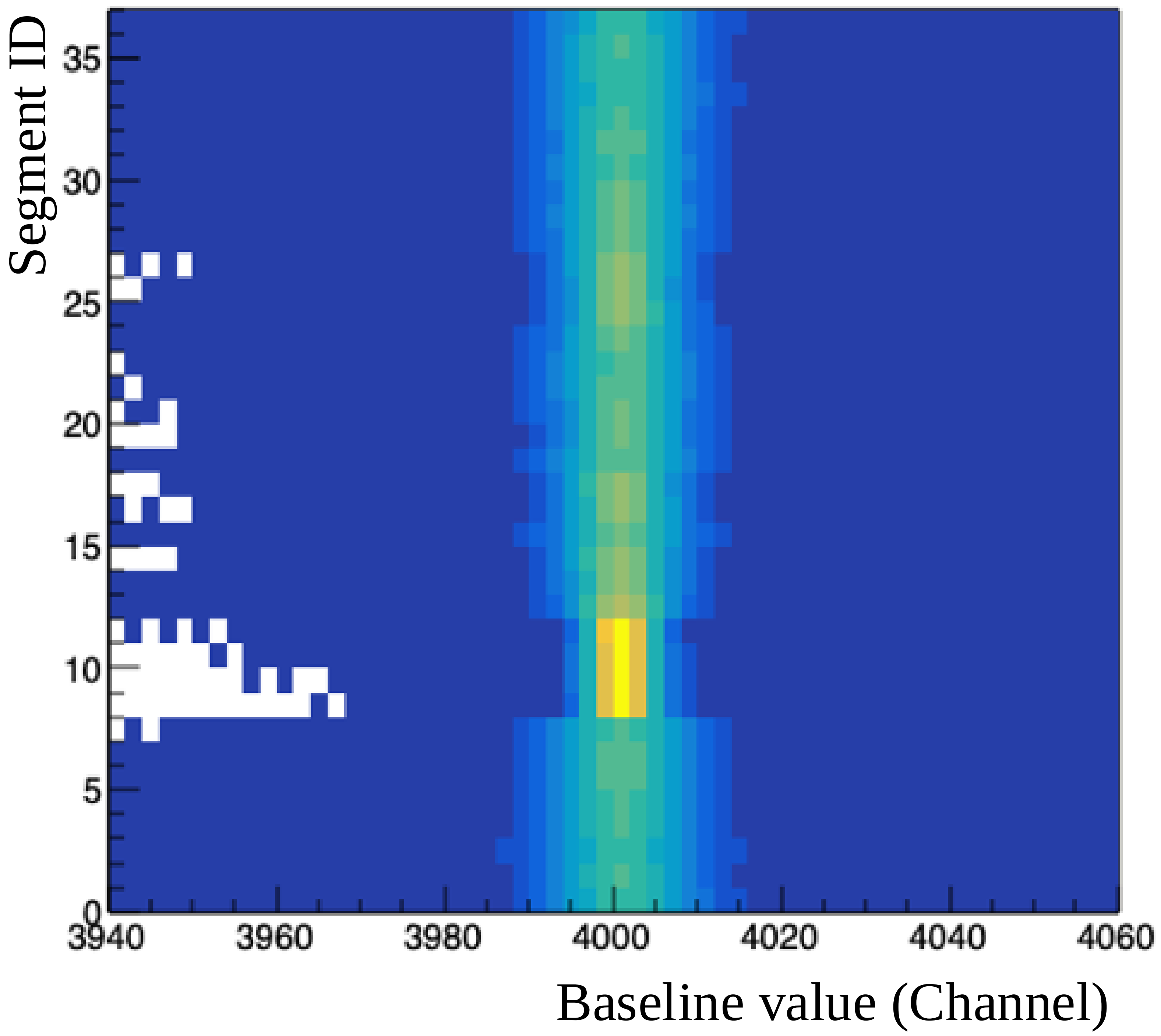
- ➤ Baseline fluctuations (critical for comparing traces):
  - ➥ Not possible to use one baseline calibration for a full scan

- ➤ **Automatic baseline calibration for each scan position**

Single calibration per scan



Auto calibration per position



Data processing

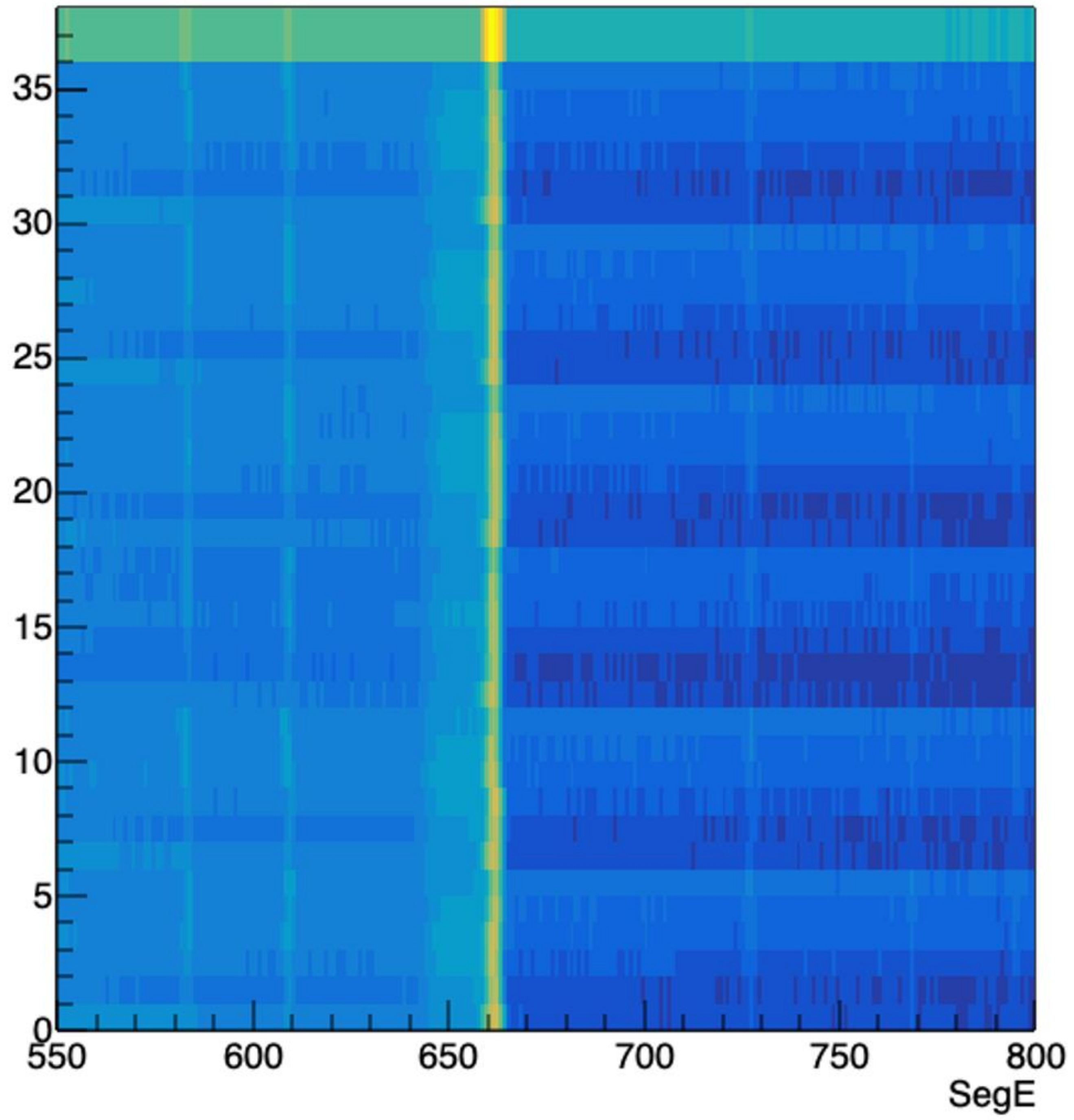
- Energy calibrations:
  - ➡ With AGATA electronics (ATCA/GGP), no need to calibrate traces, this is automatic. Not the case for TNT2
  - ➡ Need to play artificially with the **tfall/trise** values of **PreprocessingFilterPSA.conf** file

		tfall	trise	gain		
segm	0	10.0	6.148618e-01	5.649478e-02	15	0
segm	1	10.0	6.317184e-01	5.558346e-02	15	0
segm	2	10.0	6.262809e-01	5.310815e-02	15	0
segm	3	10.0	6.336661e-01	5.976445e-02	15	0
segm	4	10.0	6.205523e-01	5.447304e-02	15	0
segm	5	10.0	6.200081e-01	5.590805e-02	15	0
segm	6	10.0	5.958490e-01	5.982570e-02	15	0
segm	7	10.0	6.367053e-01	5.183969e-02	15	0
segm	8	10.0	6.289934e-01	1.090906e-01	15	0
segm	9	10.0	6.099112e-01	1.158522e-01	15	0
segm	10	10.0	6.138186e-01	1.127747e-01	15	0

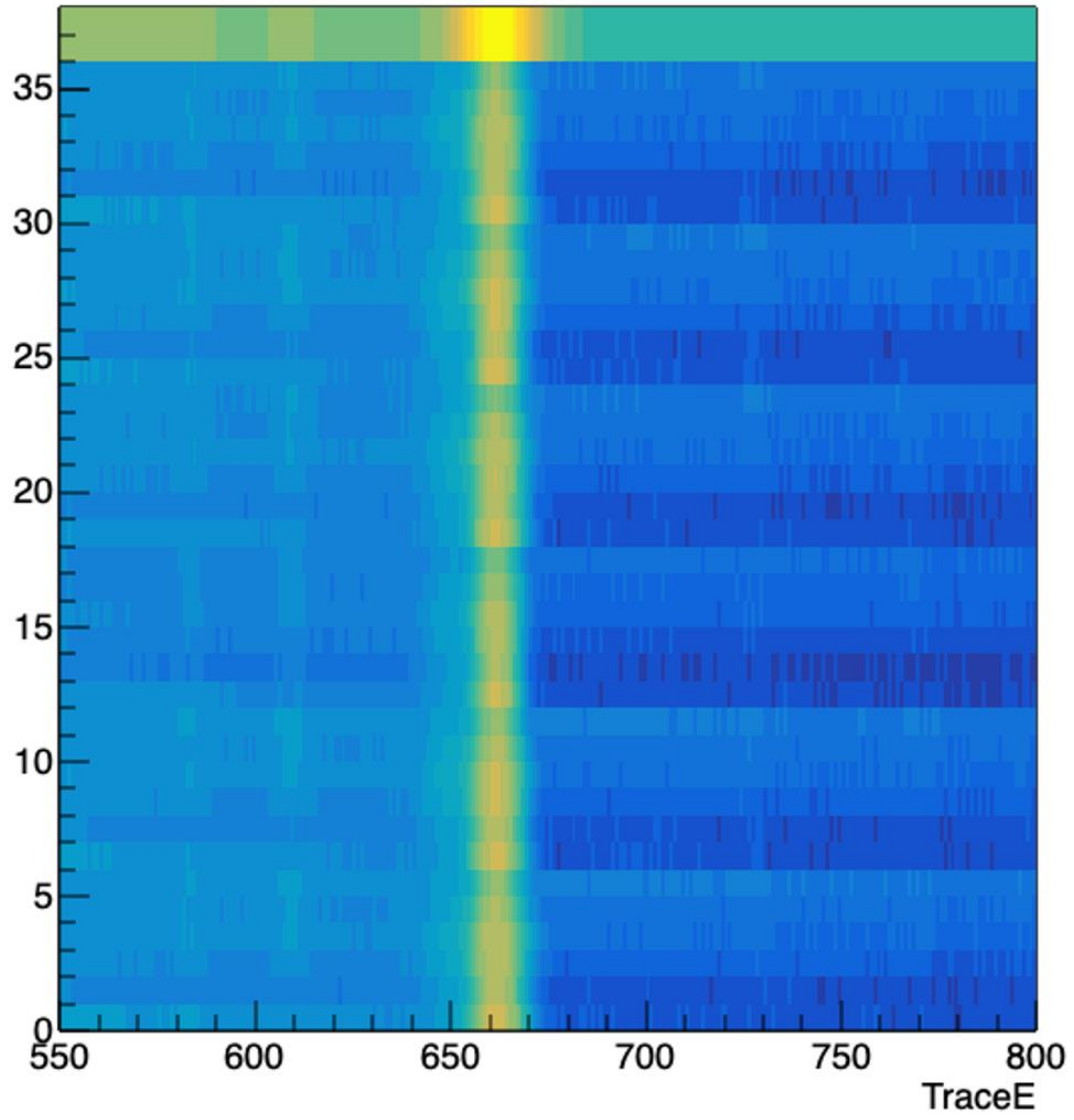
Data processing

- ➤ Energy calibrations:
  - ➥ With AGATA electronics (ATCA/GGP), no need to calibrate traces, this is automatic. Not the case for TNT2
  - ➥ Need to play artificially with the **tfall/trise** values of **PreprocessingFilterPSA.conf** file
- ➤ **Segments/cores traces are calibrated**

Energies

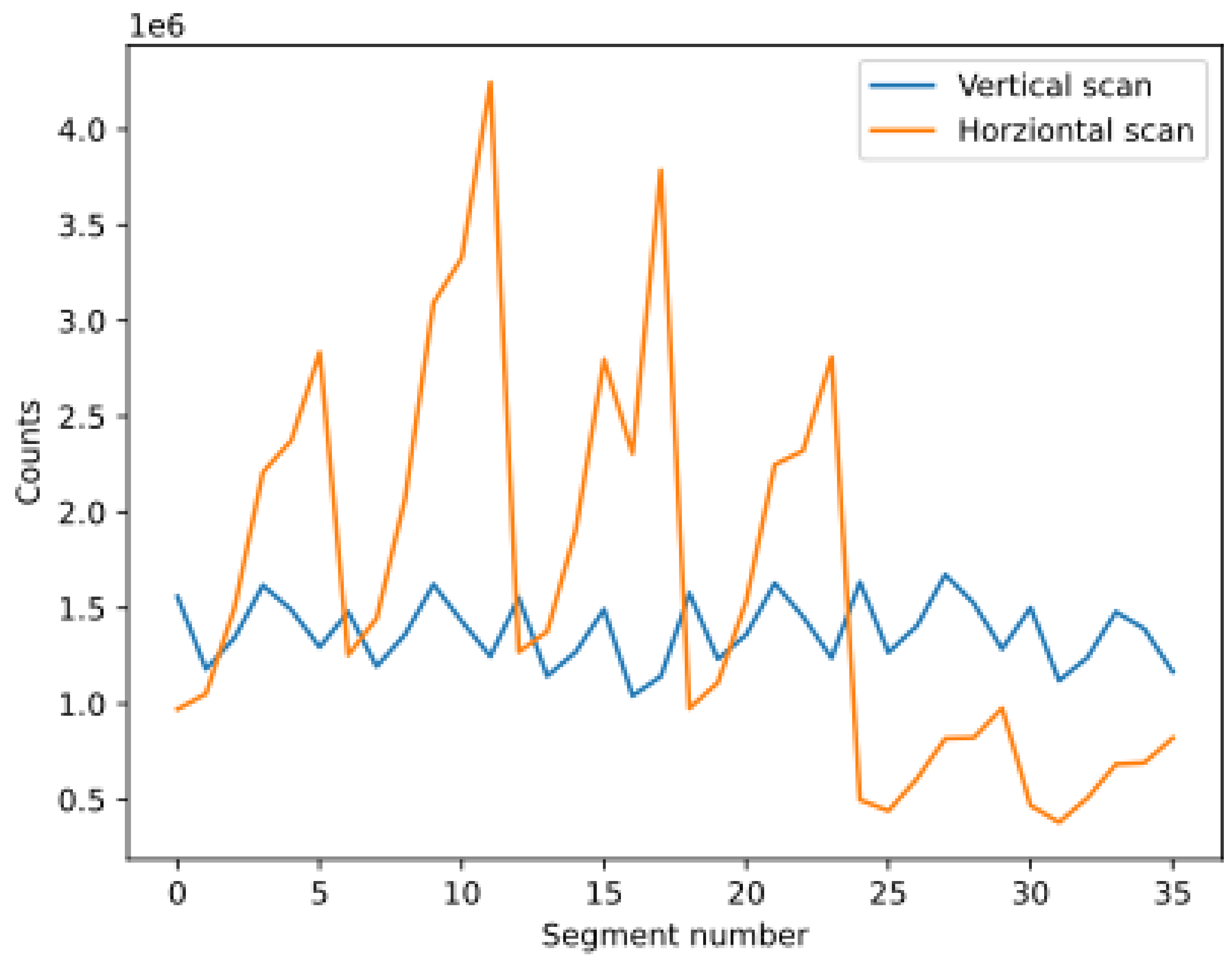


Max of the trace



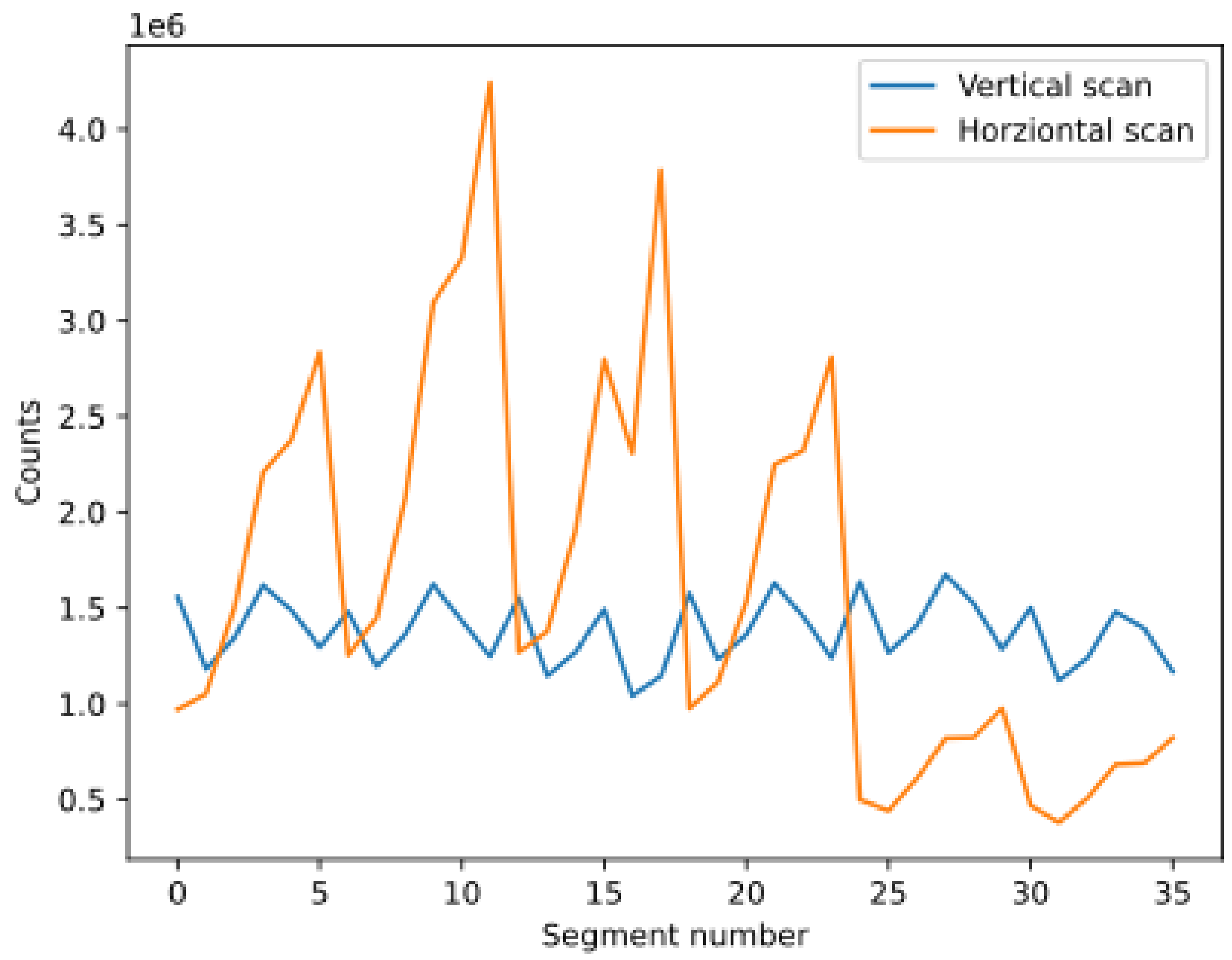
S001: Neural network training

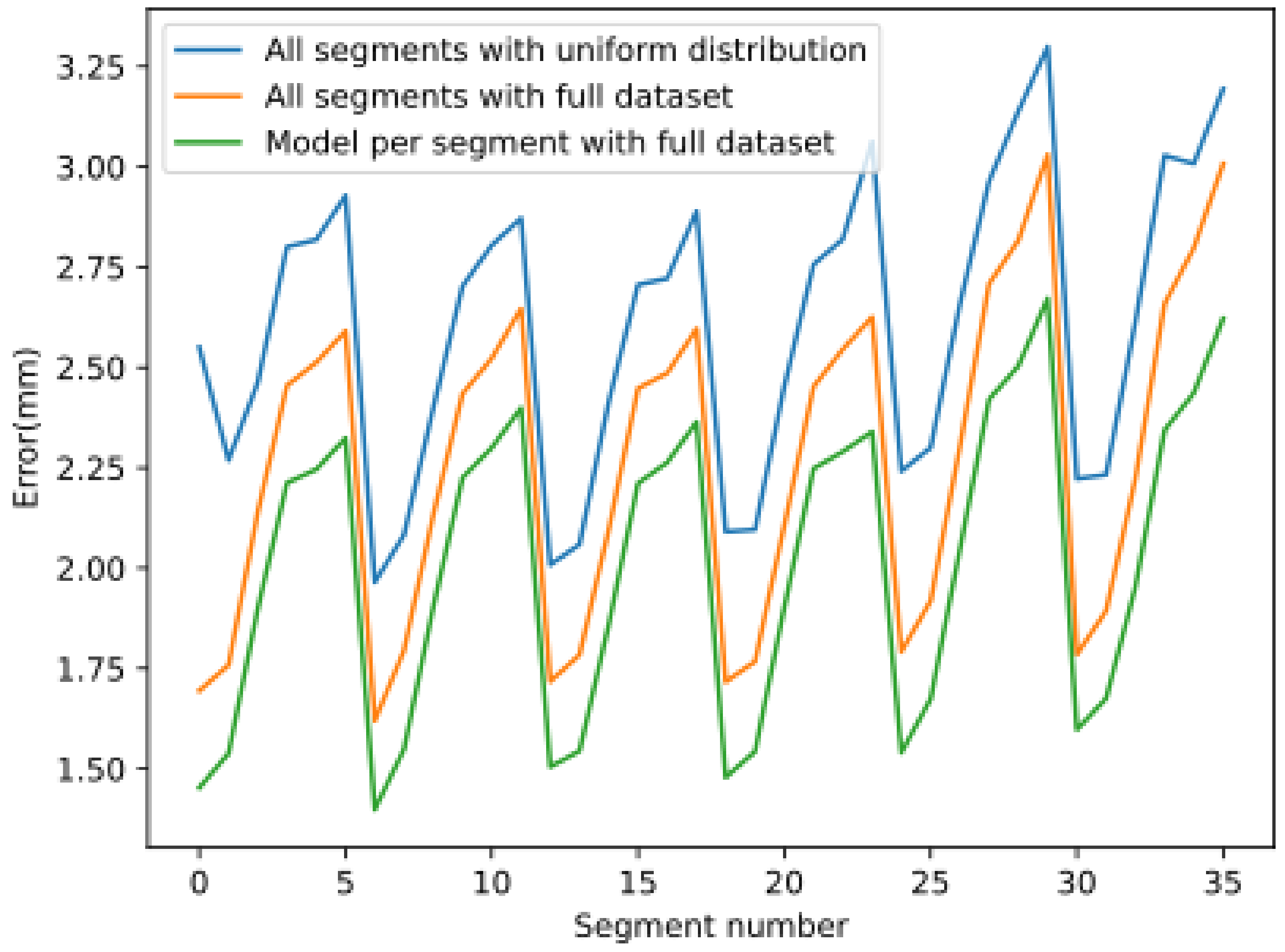
- Large statistic fluctuations per segment:
  - ➥ Vertical scan quite homogenous: attenuation is compensated by the larger size of backward segments
  - ➥ Horizontal scan: large discrepancies, more statistic in the larger layers, and in the segments closer to the source



S001: Neural network training

- Large statistic fluctuations per segment:
  - ➥ Vertical scan quite homogenous: attenuation is compensated by the larger size of backward segments
  - ➥ Horizontal scan: large discrepancies, more statistic in the larger layers, and in the segments closer to the source
- Training process:
  - ➥ Best results with one dedicated model per segment
  - ➥ 90% of data for training, 10% for validation

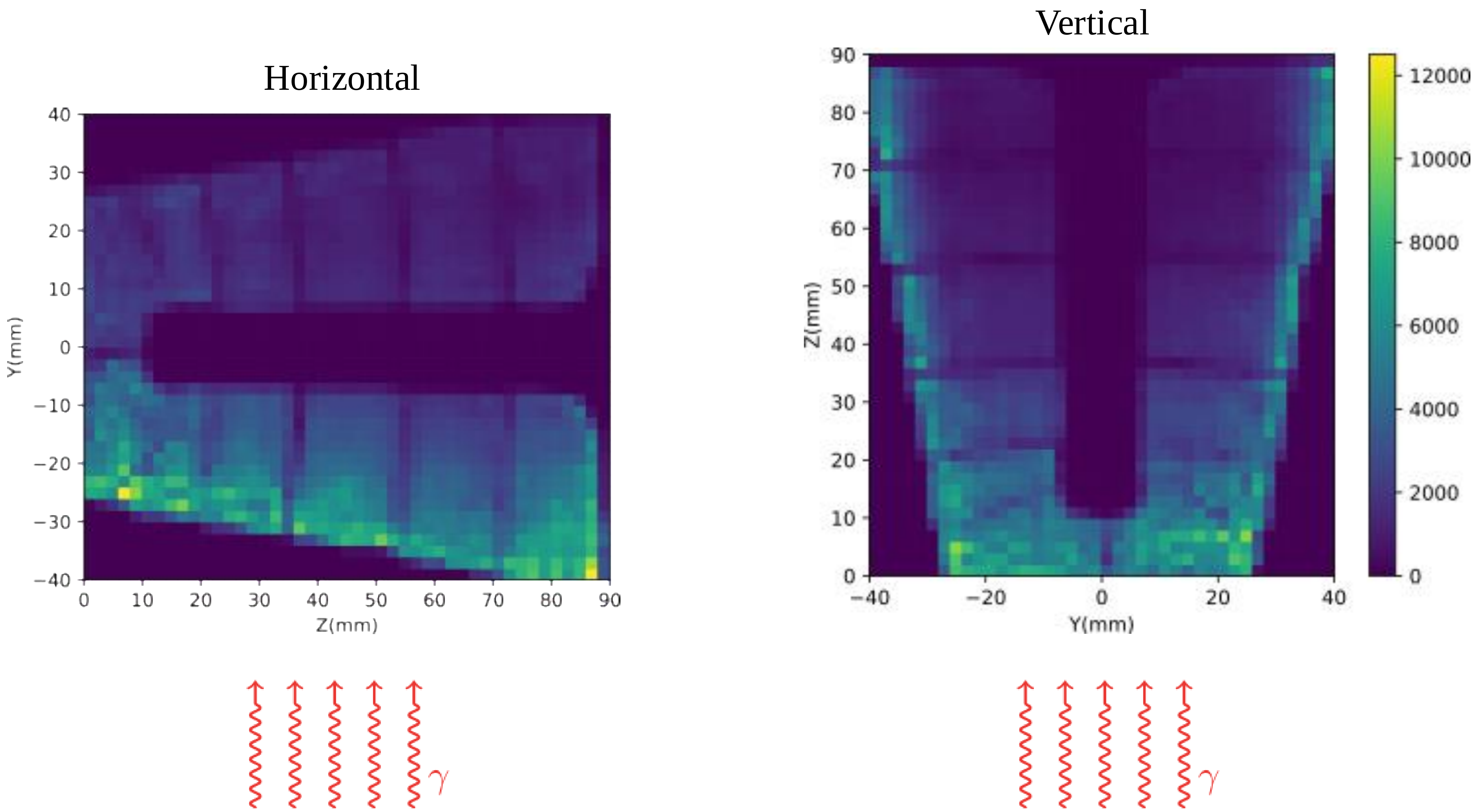




S001: Neural network results

➤ The distribution of the predicted positions conforms with the attenuation of the gamma rays.

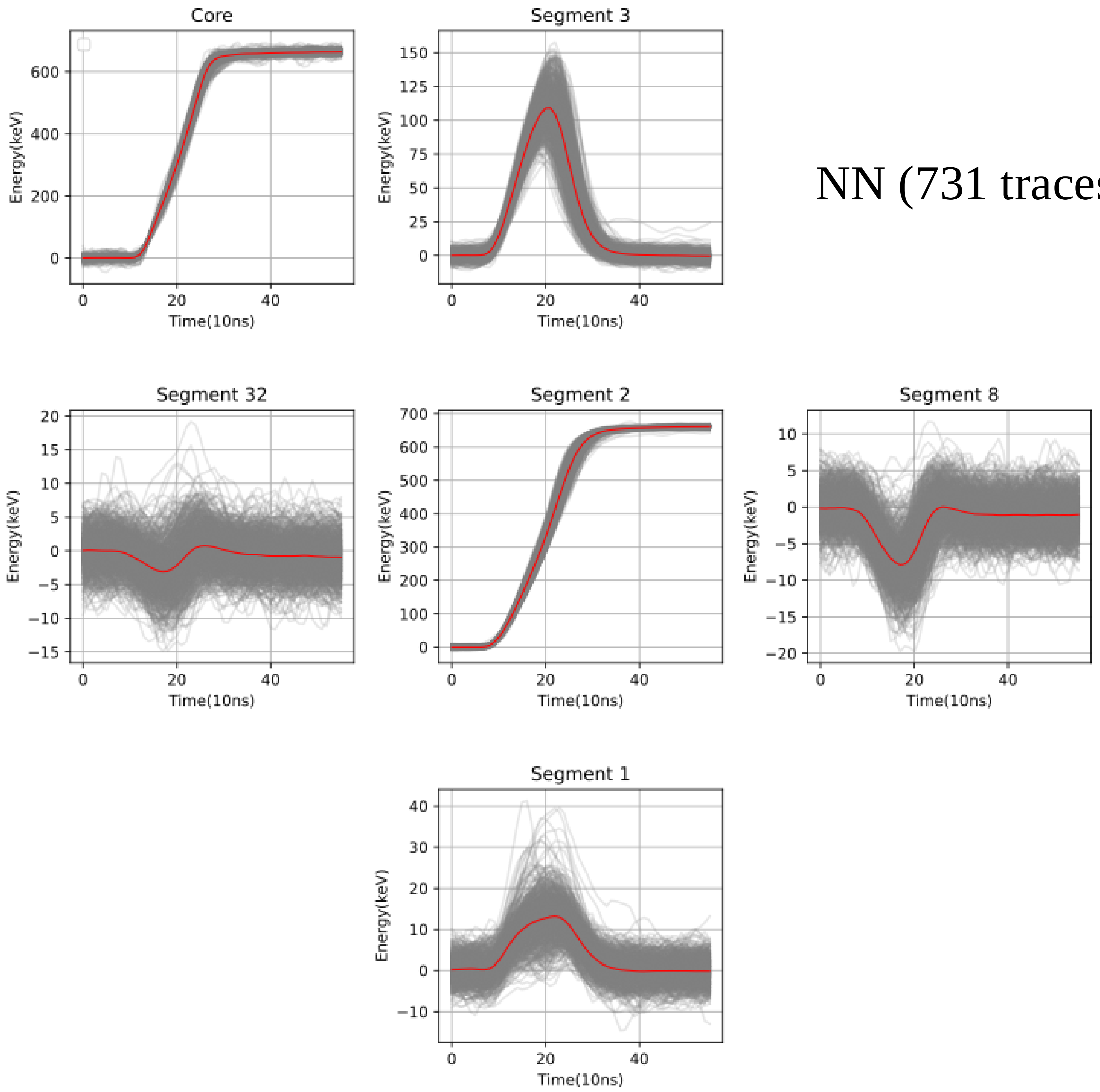
Predicted positions in (Y,Z) plane for X in [-3 mm ; 3 mm]



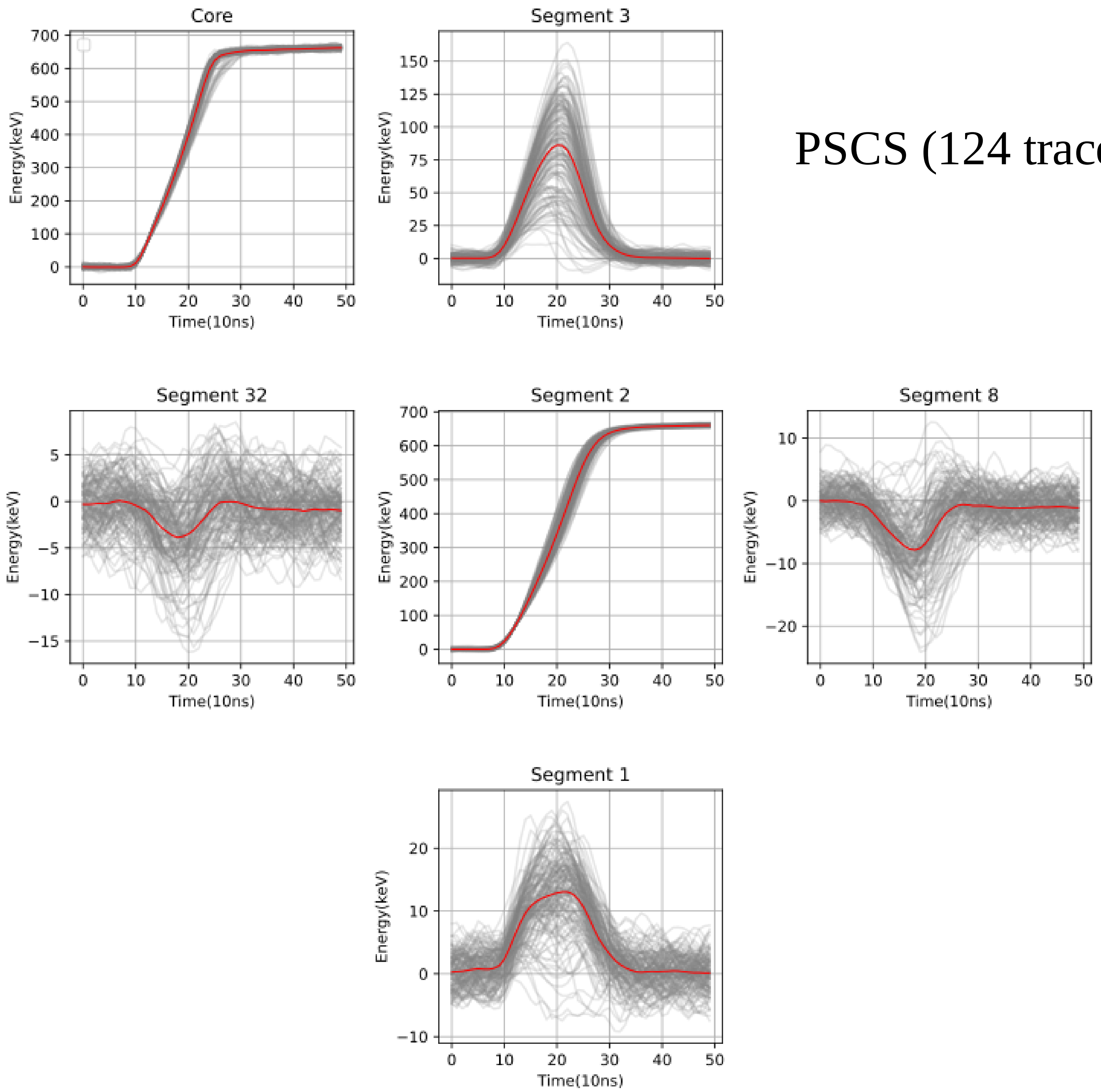
S001: Model consistency

➤ Average trace between Neural network and PSCS looks similar but more statistics and less fluctuations in NN

Traces predicted at position (22,0,34) in segment 2



NN (731 traces)

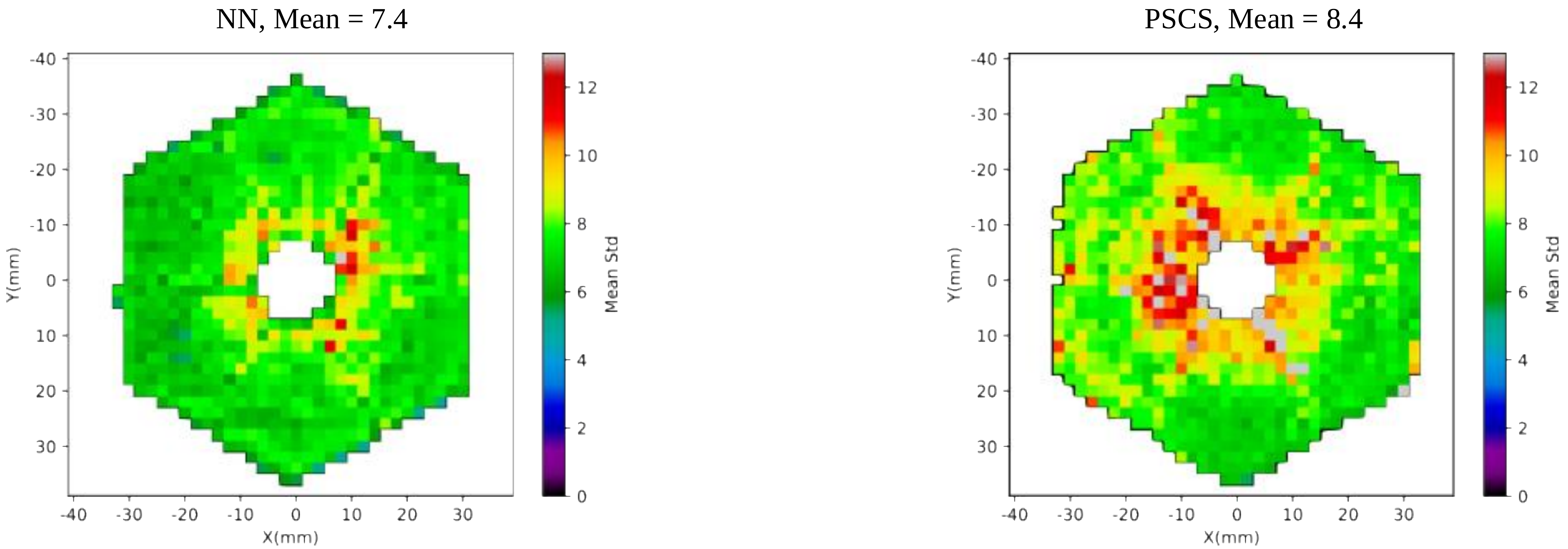


PSCS (124 traces)

S001: Model consistency

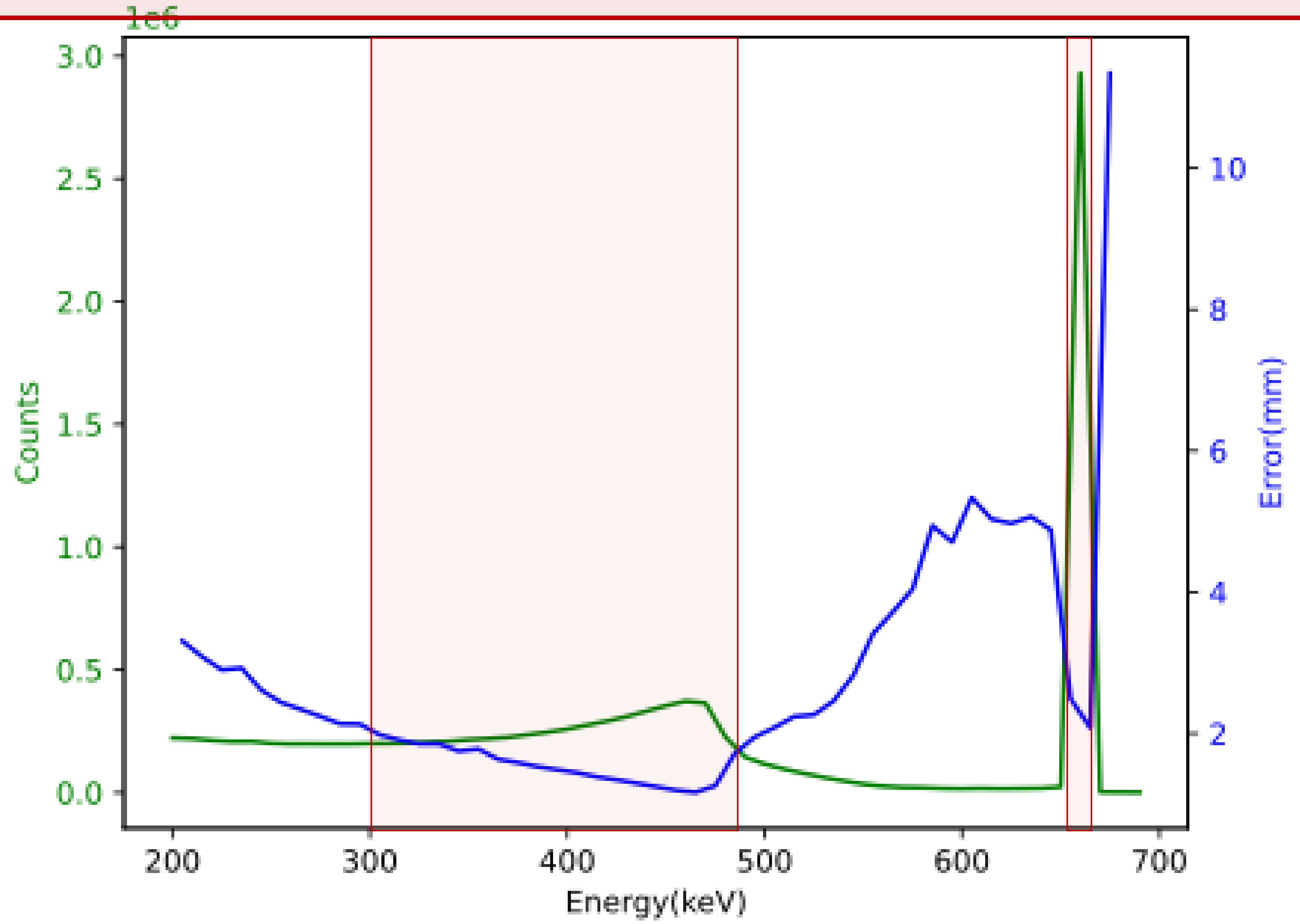
➤ Average trace between Neural network and PSCS looks similar but more statistics and less fluctuations in NN

Mean Std of the traces per pixel at Z=30mm



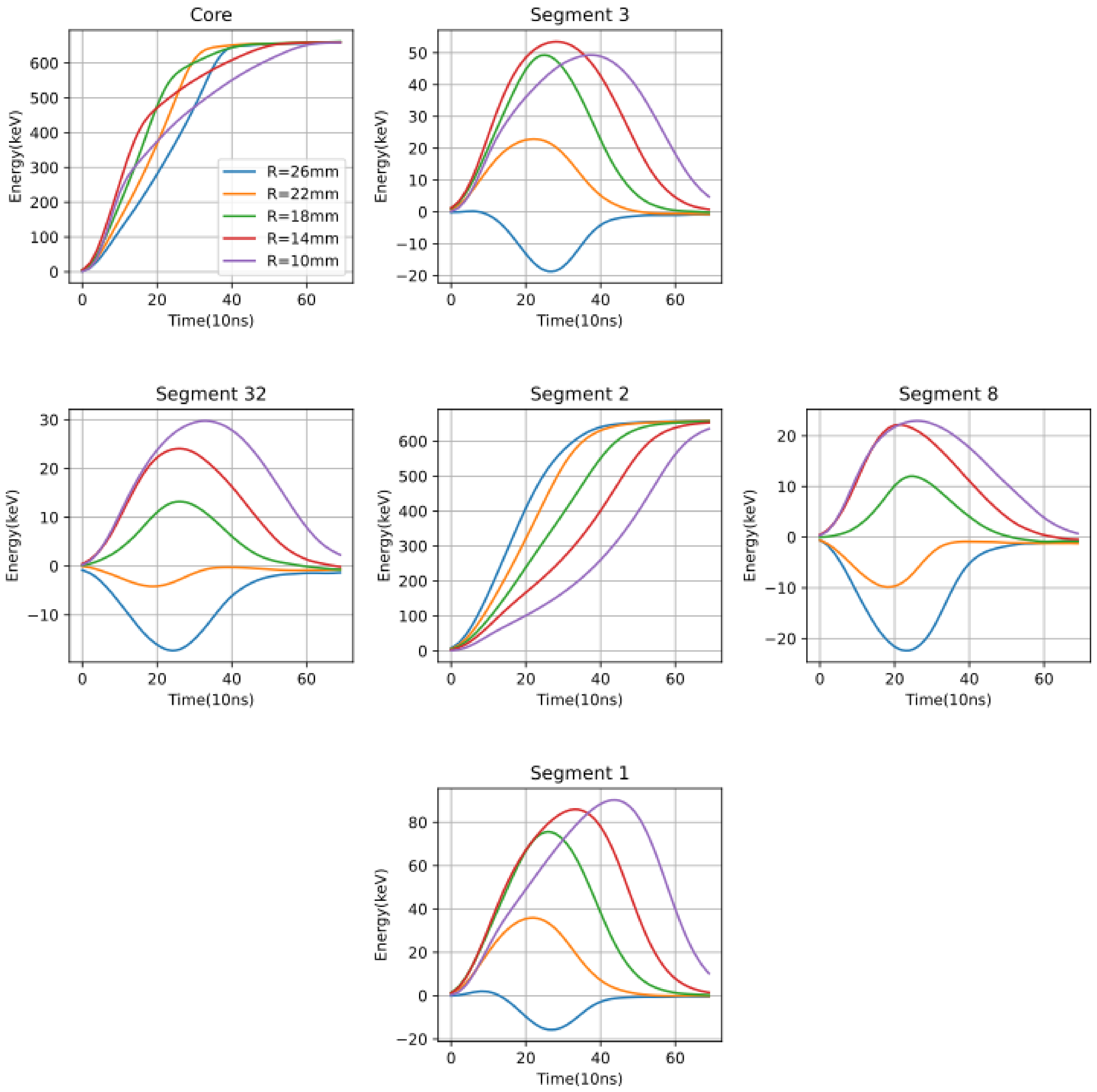
S001: The NN Experimental basis

- Filtering of the data to keep the best traces:
  - ➥ limitation on the energy range with the best results: **300-480 keV** and **650-670 keV**
  - ➥ in this energy range, only traces with **error < 2mm** are kept
  - ➥ an **iterative method** based on the trace std then **removes the remaining noisy signals**
- Remaining traces in the basis: 43% of the full scan



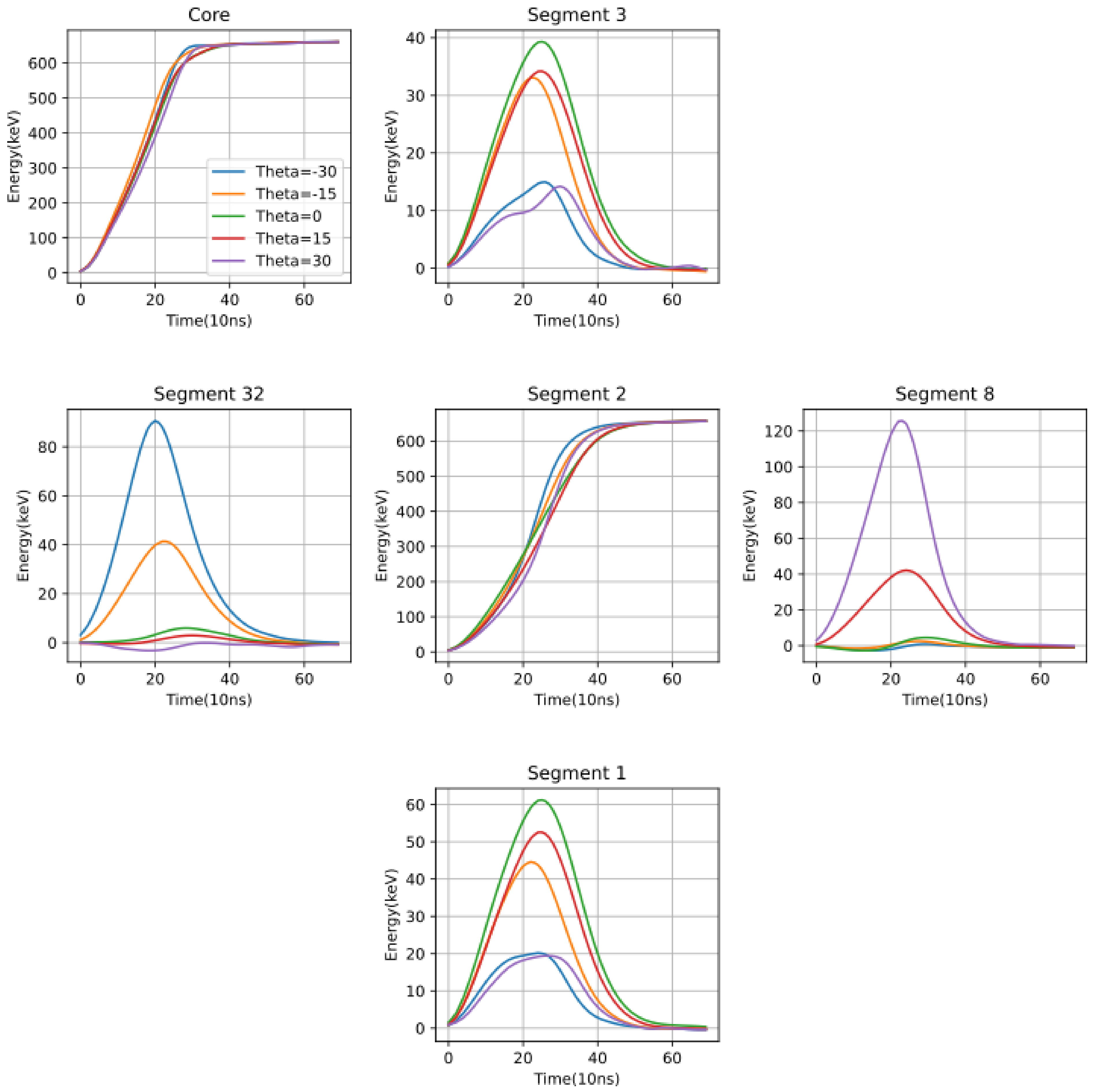
S001: The NN Experimental basis

➤ Experimental basis as a function of radius



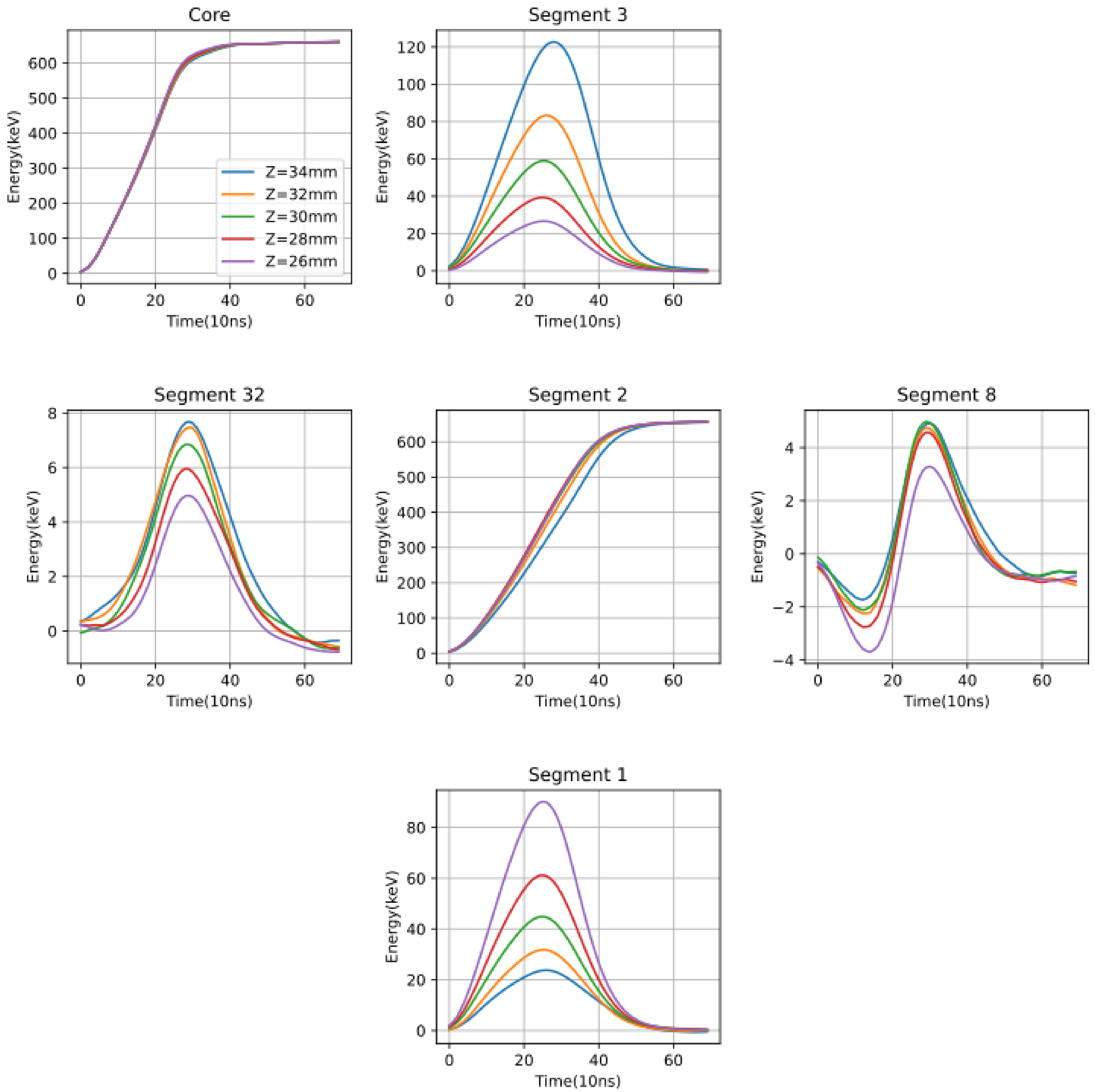
S001: The NN Experimental basis

➤ Experimental basis as a function of theta



S001: The NN Experimental basis

➤ Experimental basis as a function of Z



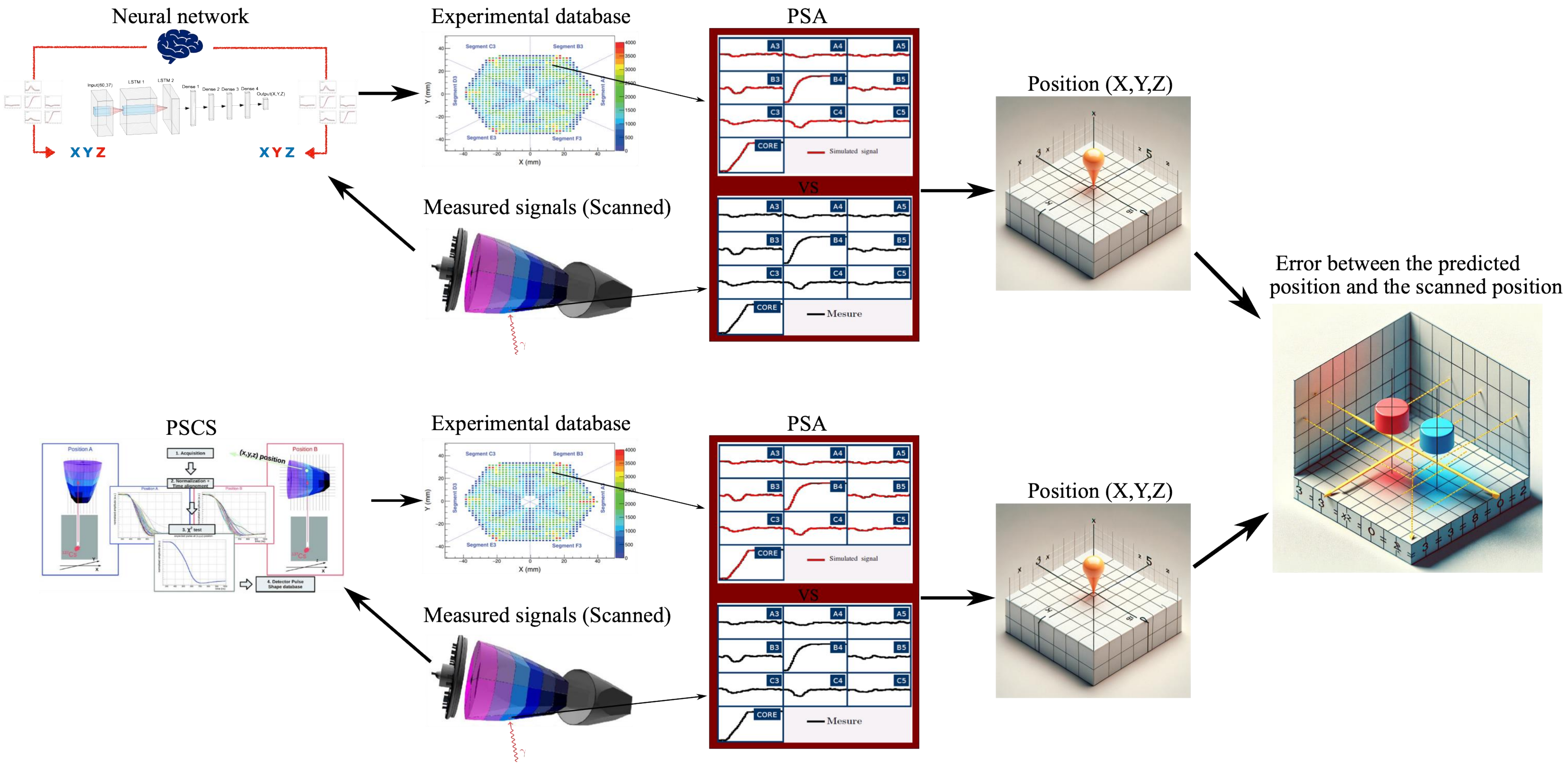
S001: PSA results

➤ We are now ready to perform PSA !!!

➥ PSA performed with the standard AGAPRO/femul environment (Full Grid Search used by default)

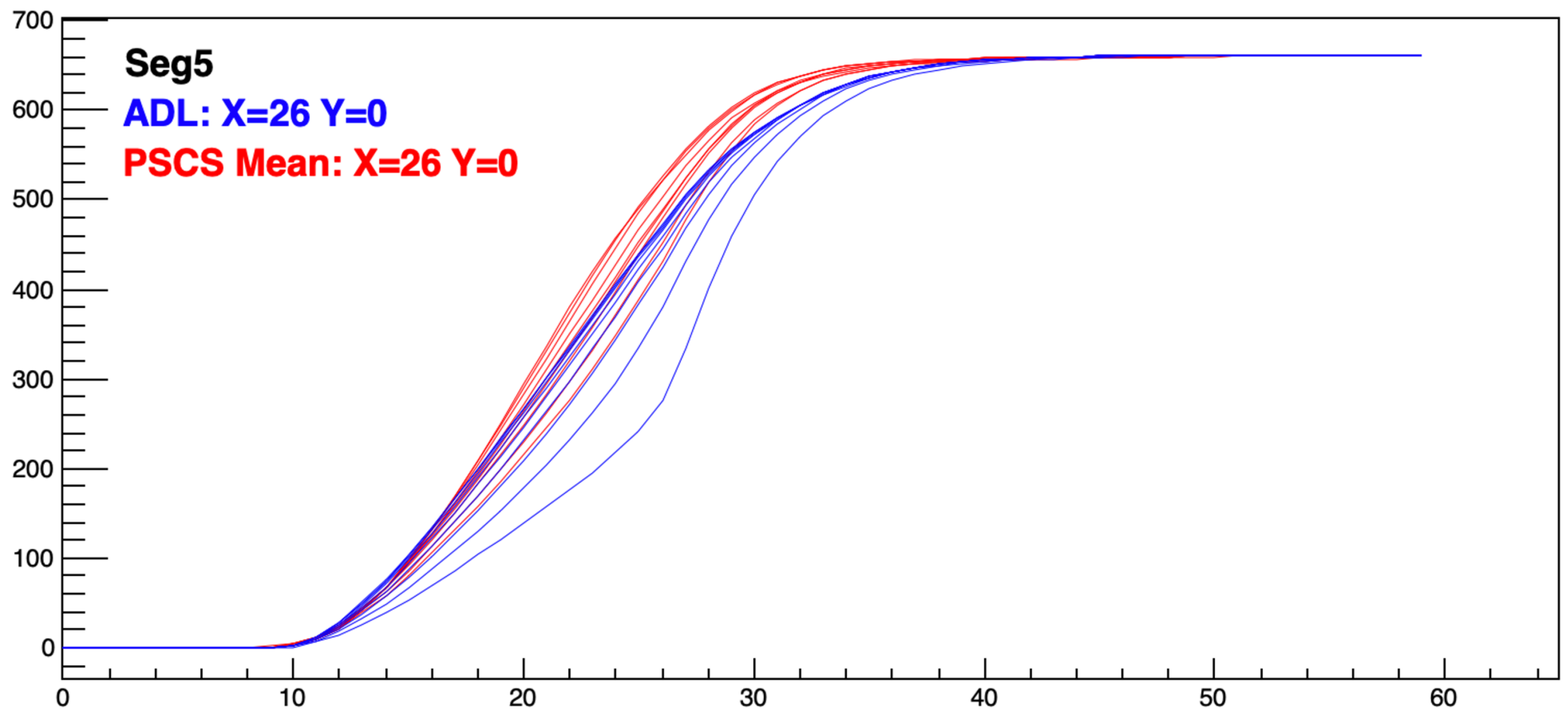
➥ The basis is switched between ADL, NN and PSCS (for NN and PSCS, preamp response removed in PSA)

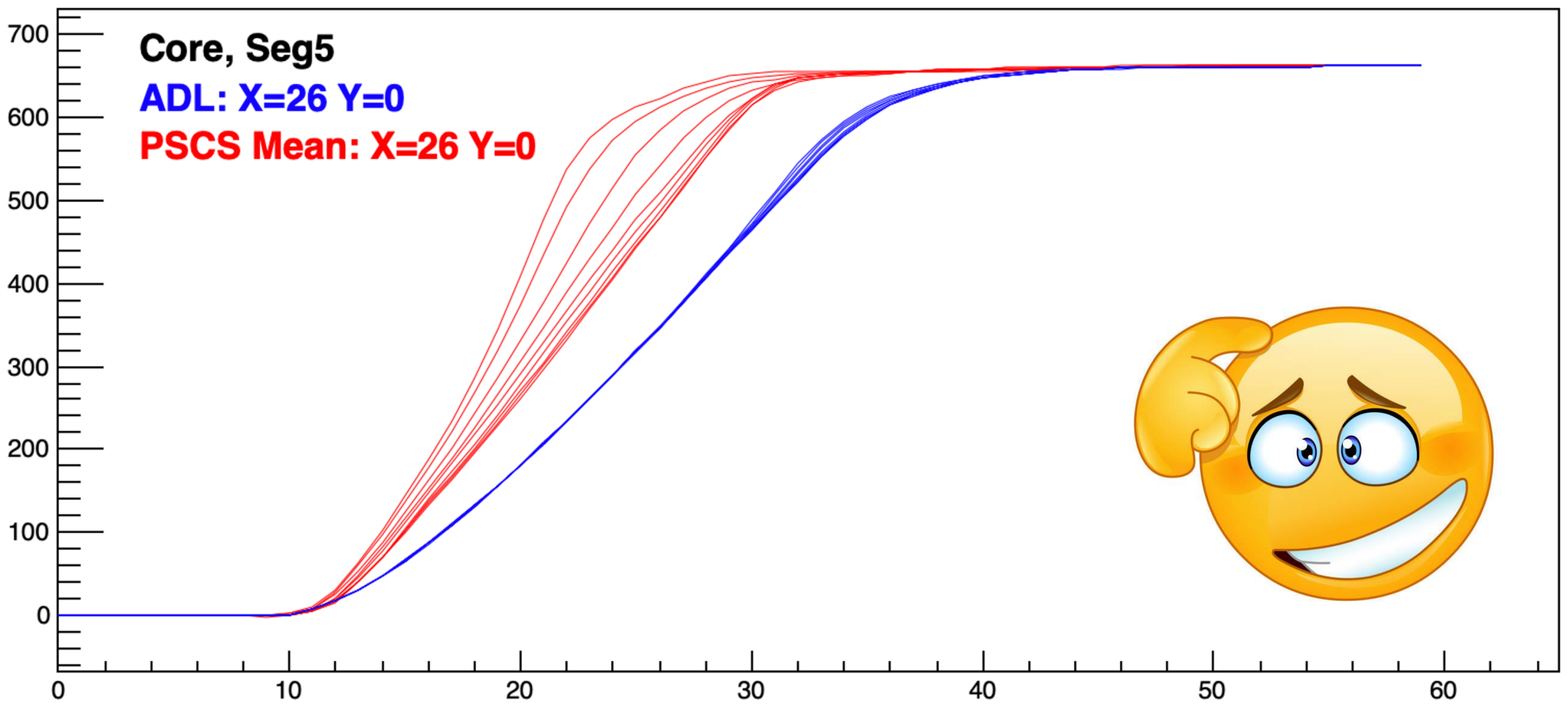
➥ PSA results are then compared



S001: PSA results

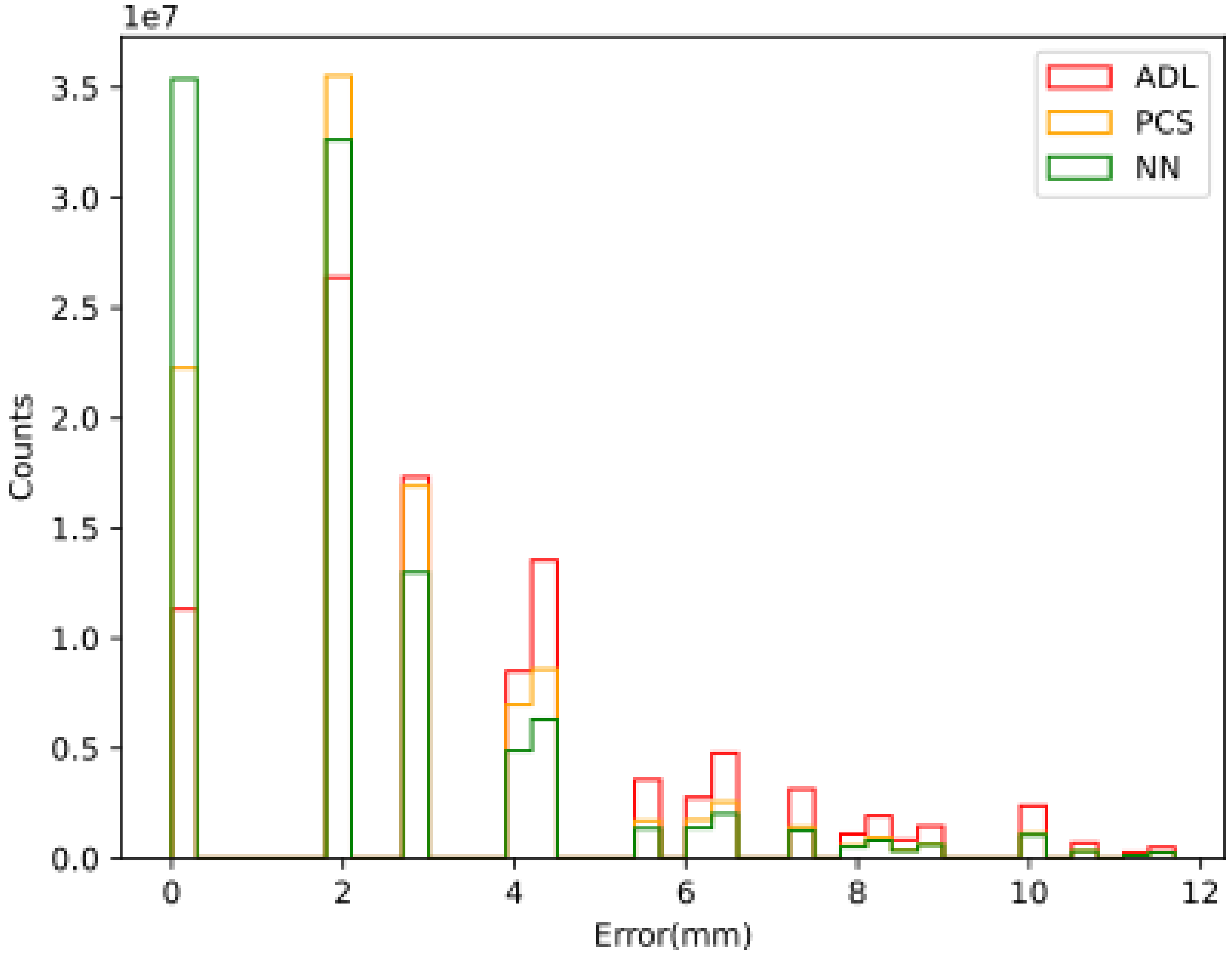
- Keep in mind that for S001, ADL comparisons are not fair !
  - ➥ The core signal is too fast compared to ADL and we didn't succeeded to understand why





S001: PSA results

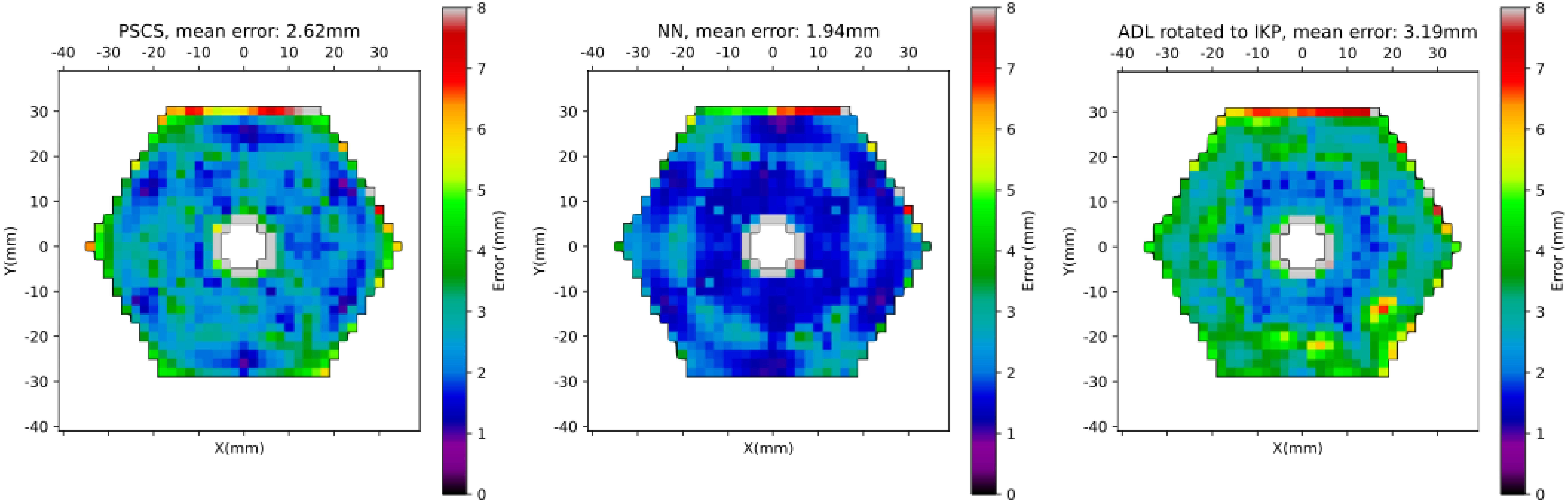
- PSA results:
  - NN basis performs the best, with a majority of events at exactly the good position (mean error = 2.59mm)
  - PSCS gives slightly worse results than NN (mean error = 3.12mm)
  - ADL is the worst (but unfair comparison) (mean error = 4.26mm)



S001: PSA results

➤ PSA results:
- ➥ Clear patterns appear at middle radius
- ➥ The error increase with depth (so with the volume of the segment)

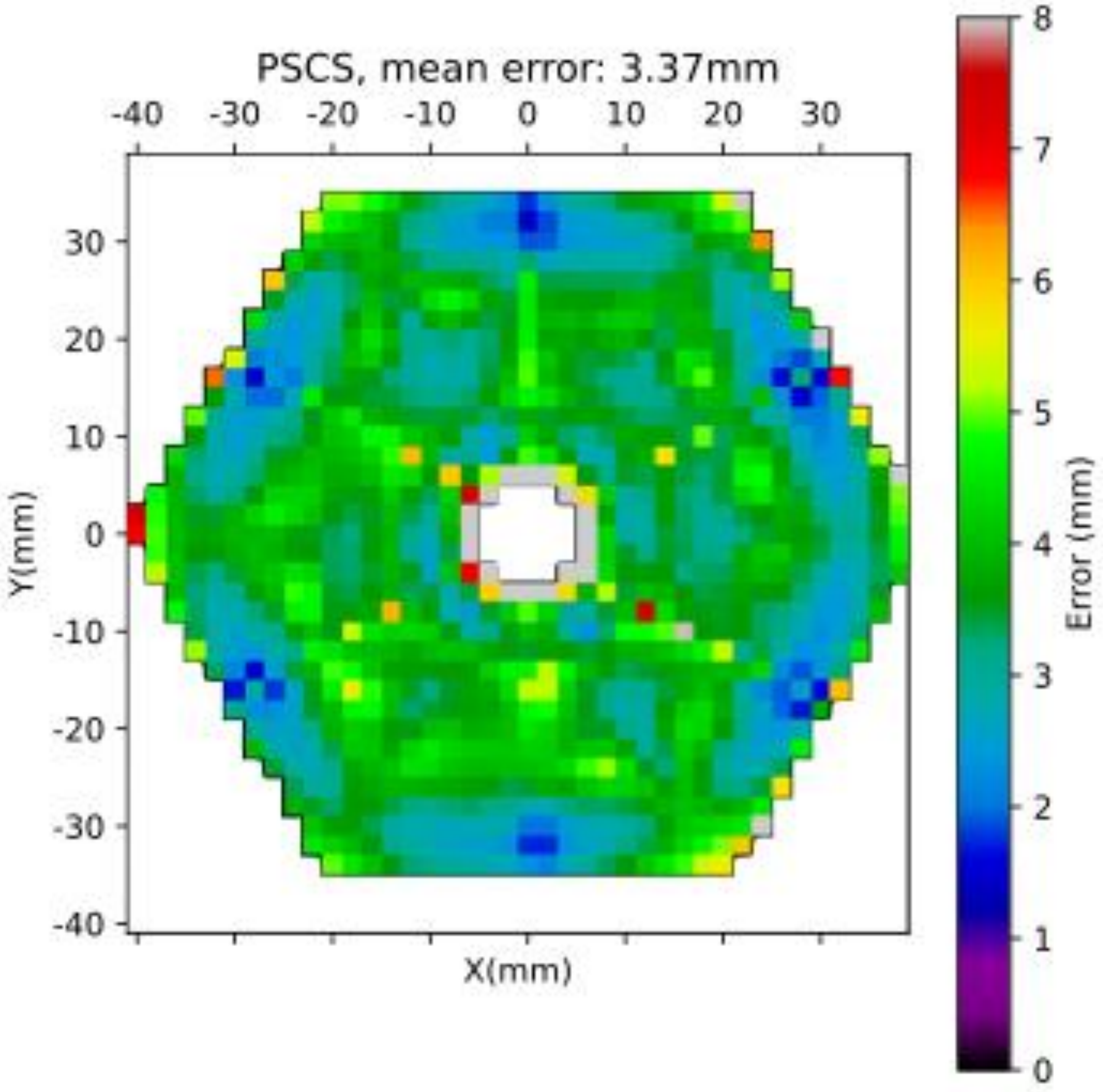
Layer 1

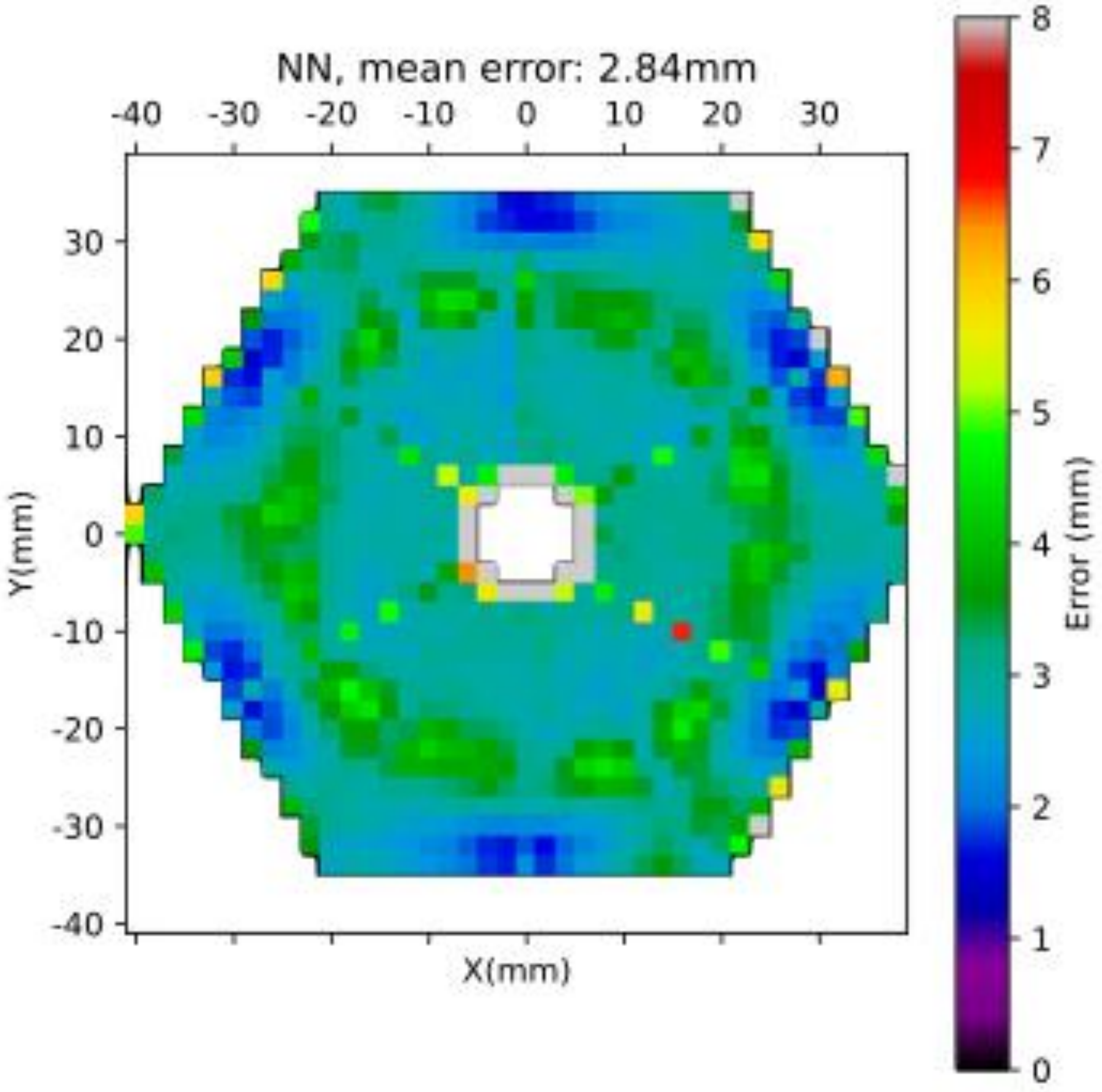


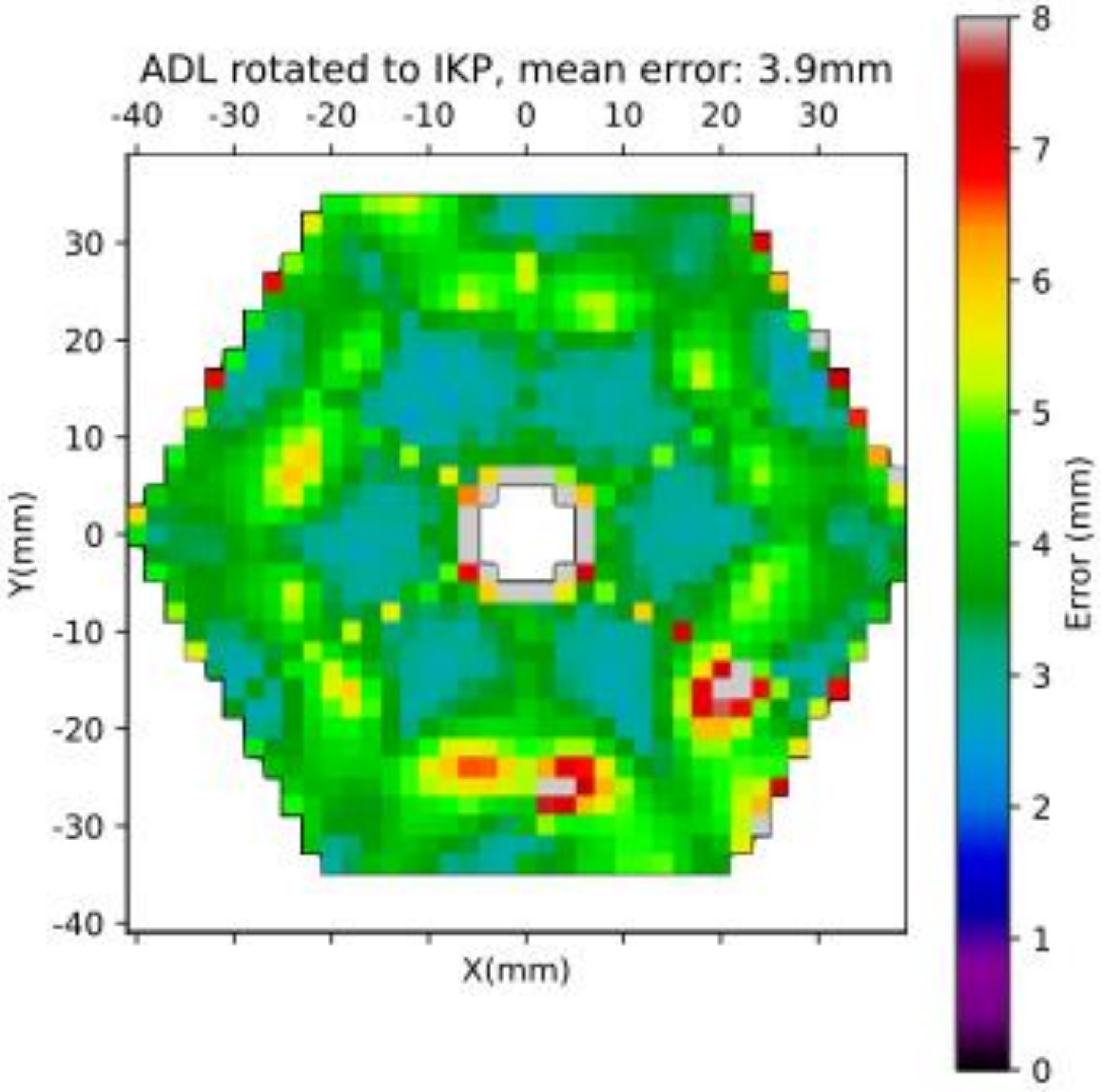
S001: PSA results

➤ PSA results:
- ➥ Clear patterns appear at middle radius
- ➥ The error increase with depth (so with the volume of the segment)

Layer 3



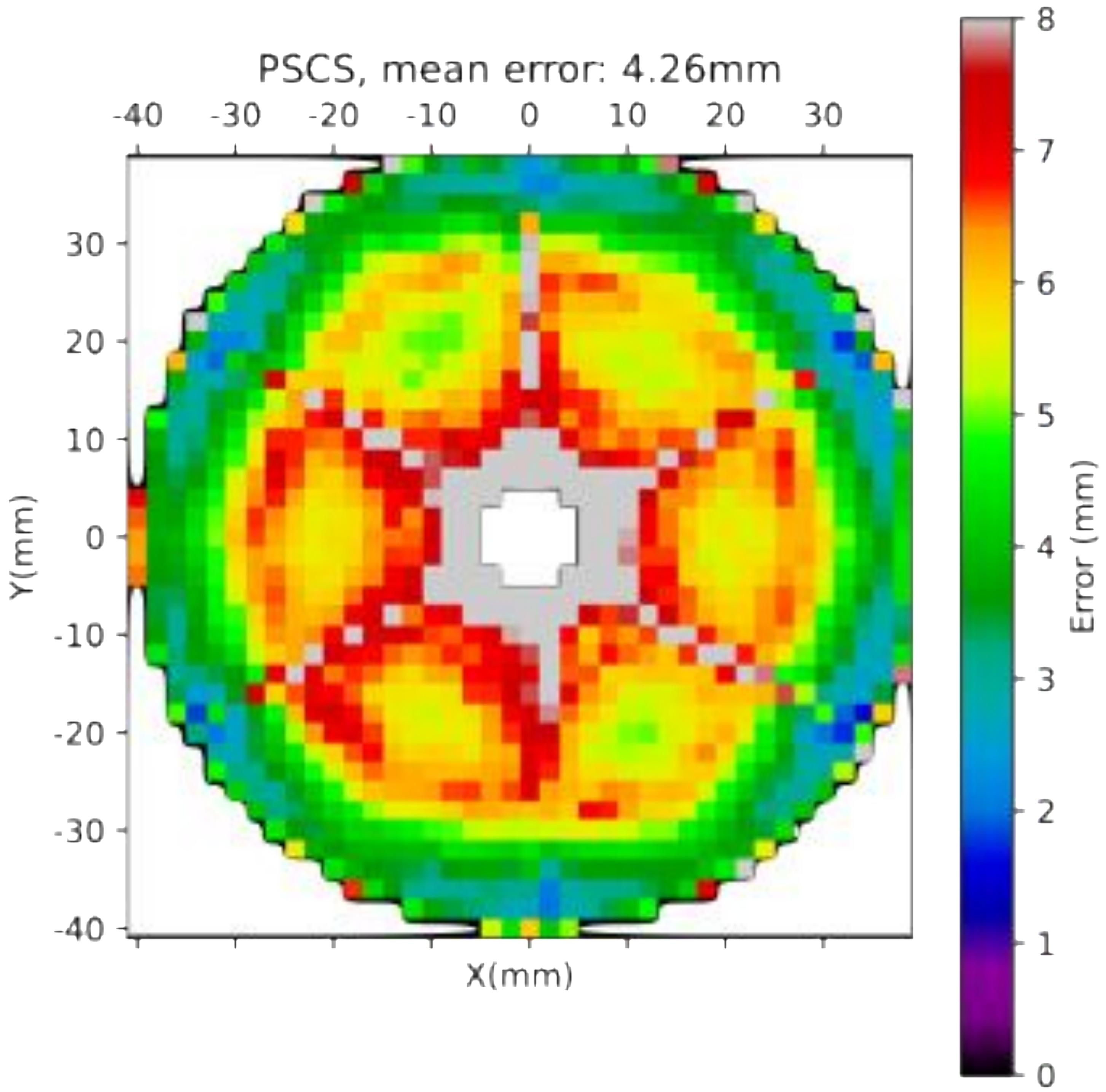


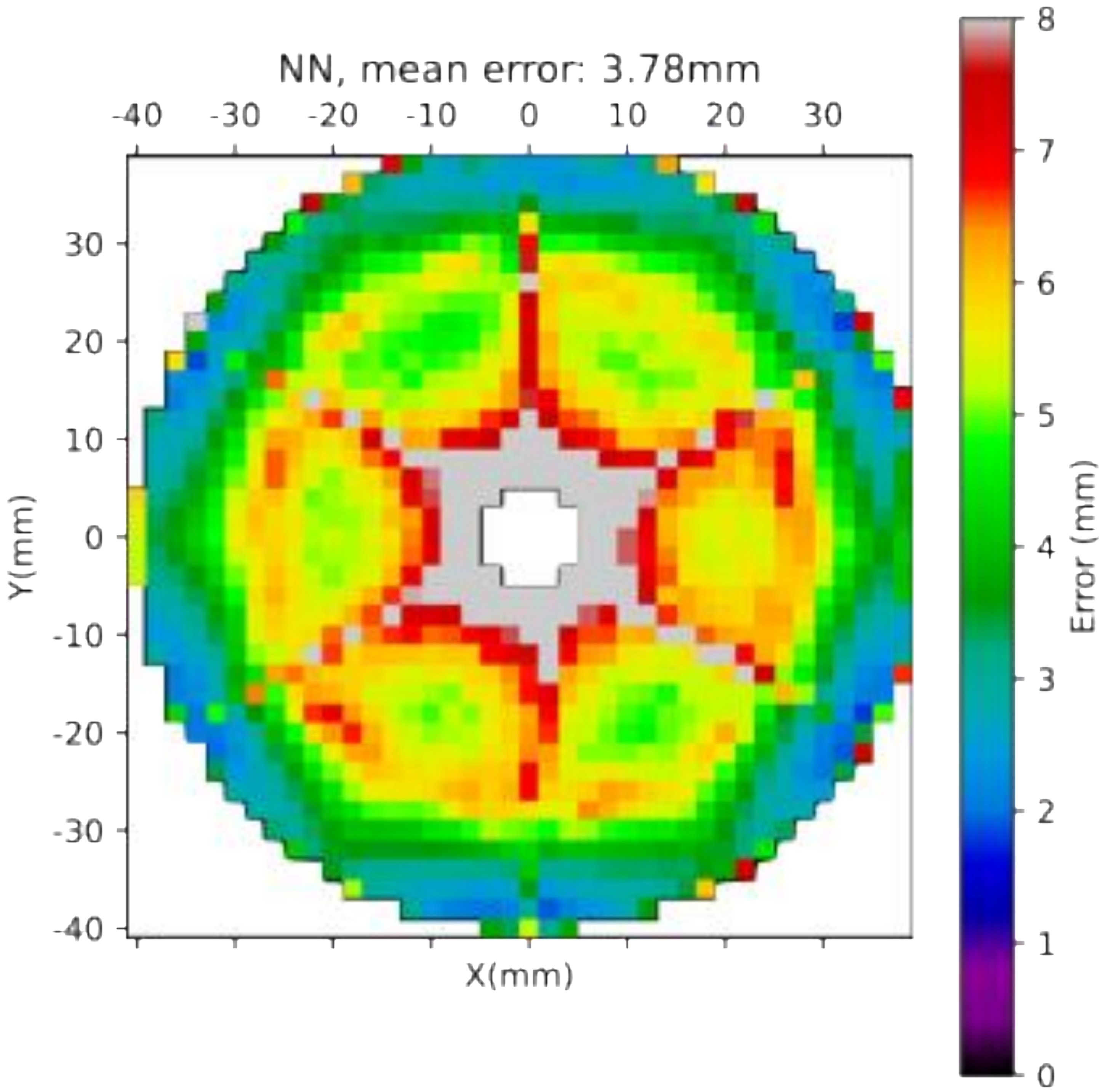


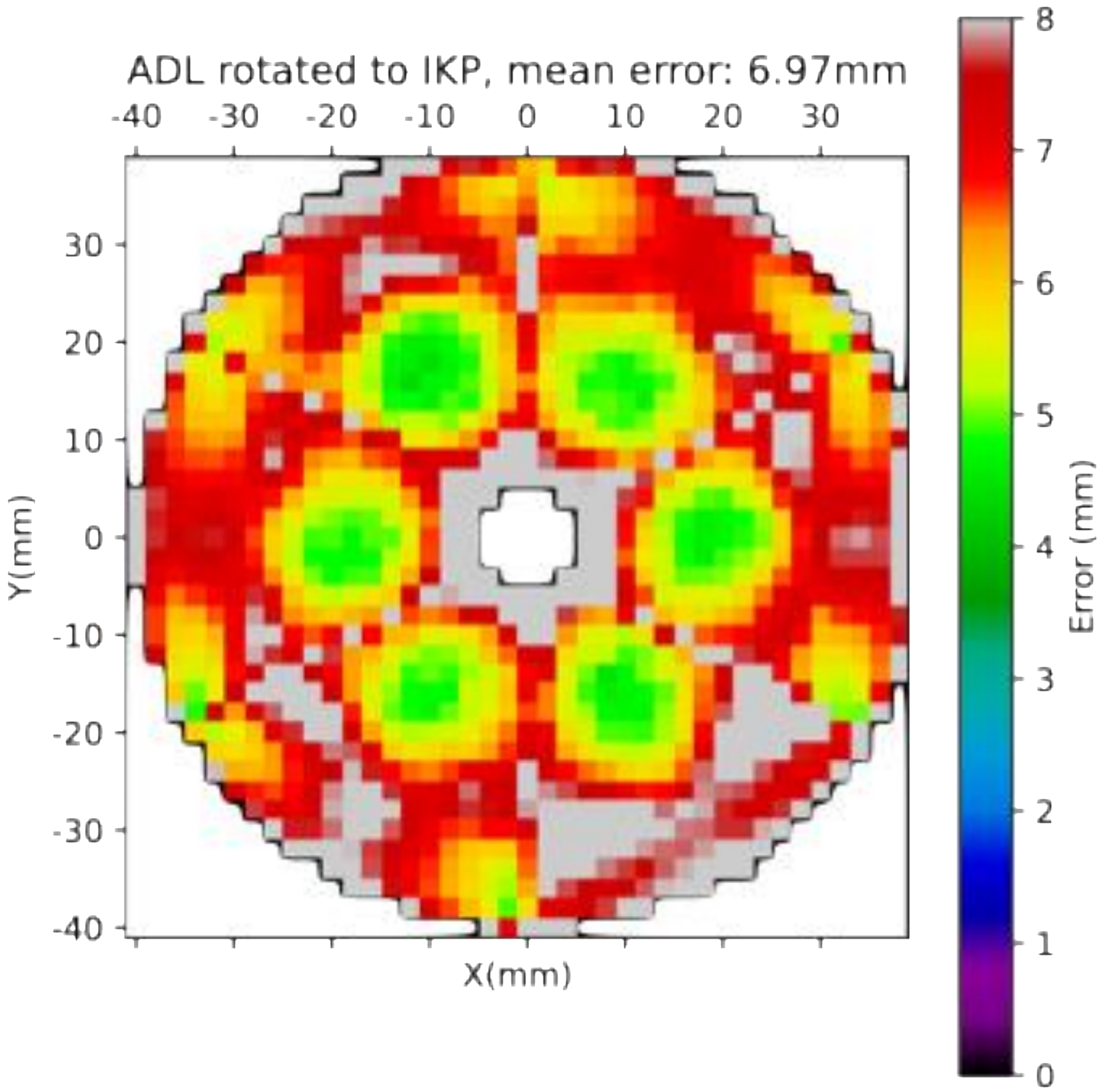
S001: PSA results

- PSA results:
  - ➥ Clear patterns appear at middle radius
  - ➥ The error increase with depth (so with the volume of the segment)

Layer 5



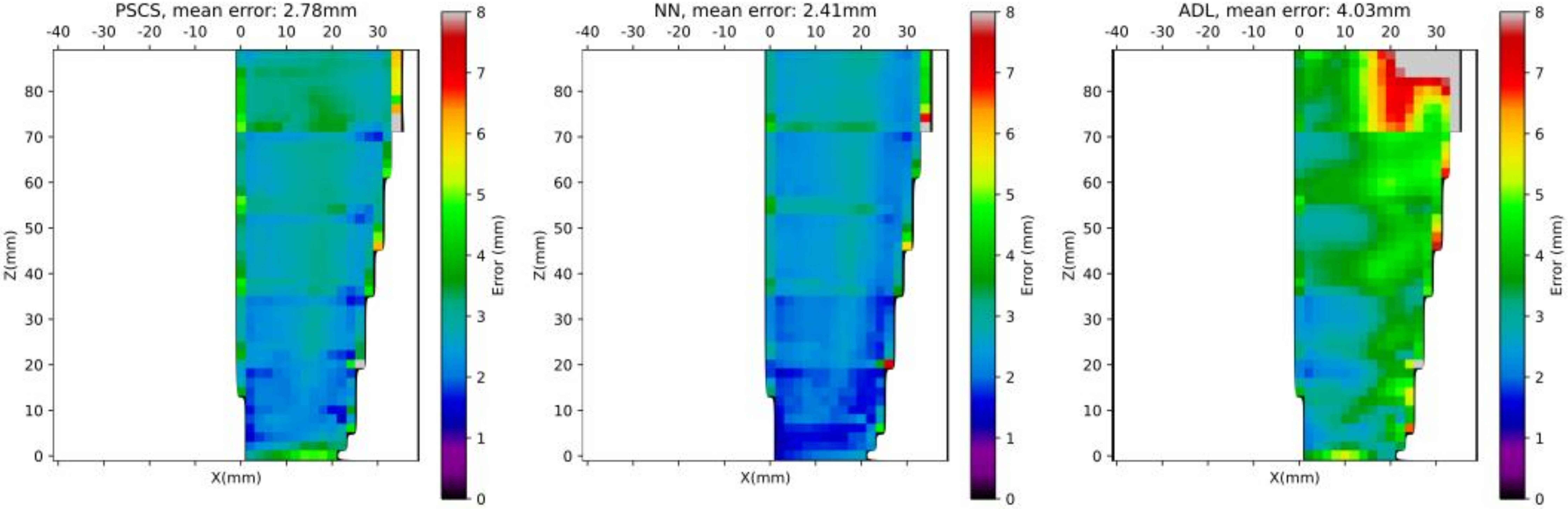




S001: PSA results

- PSA results:
  - Clear patterns appear at middle radius
  - The error increase with depth (so with the volume of the segment)

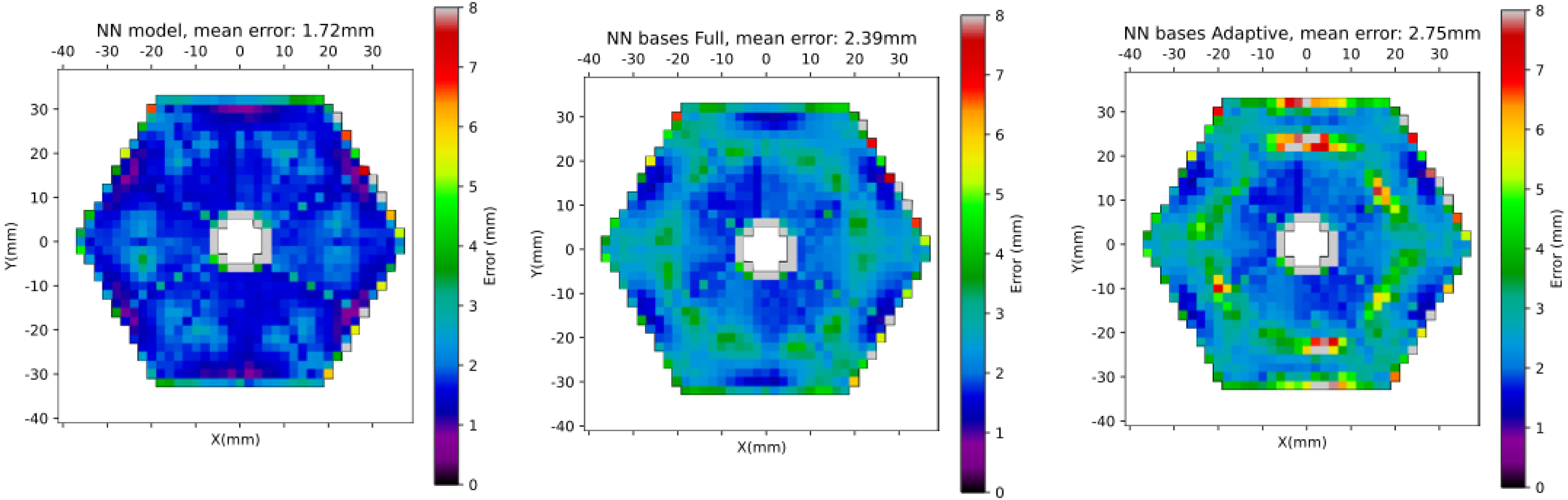
Segments 0 to 5 (A column)



S001: PSA results

- PSA algorithm effect:
  - The clustering effects in the middle of the segment is enhanced by the algorithm:
    - NN model: present but very limited
    - PSA full grid search: clearer pattern
    - PSA adaptive grid search: stronger effect

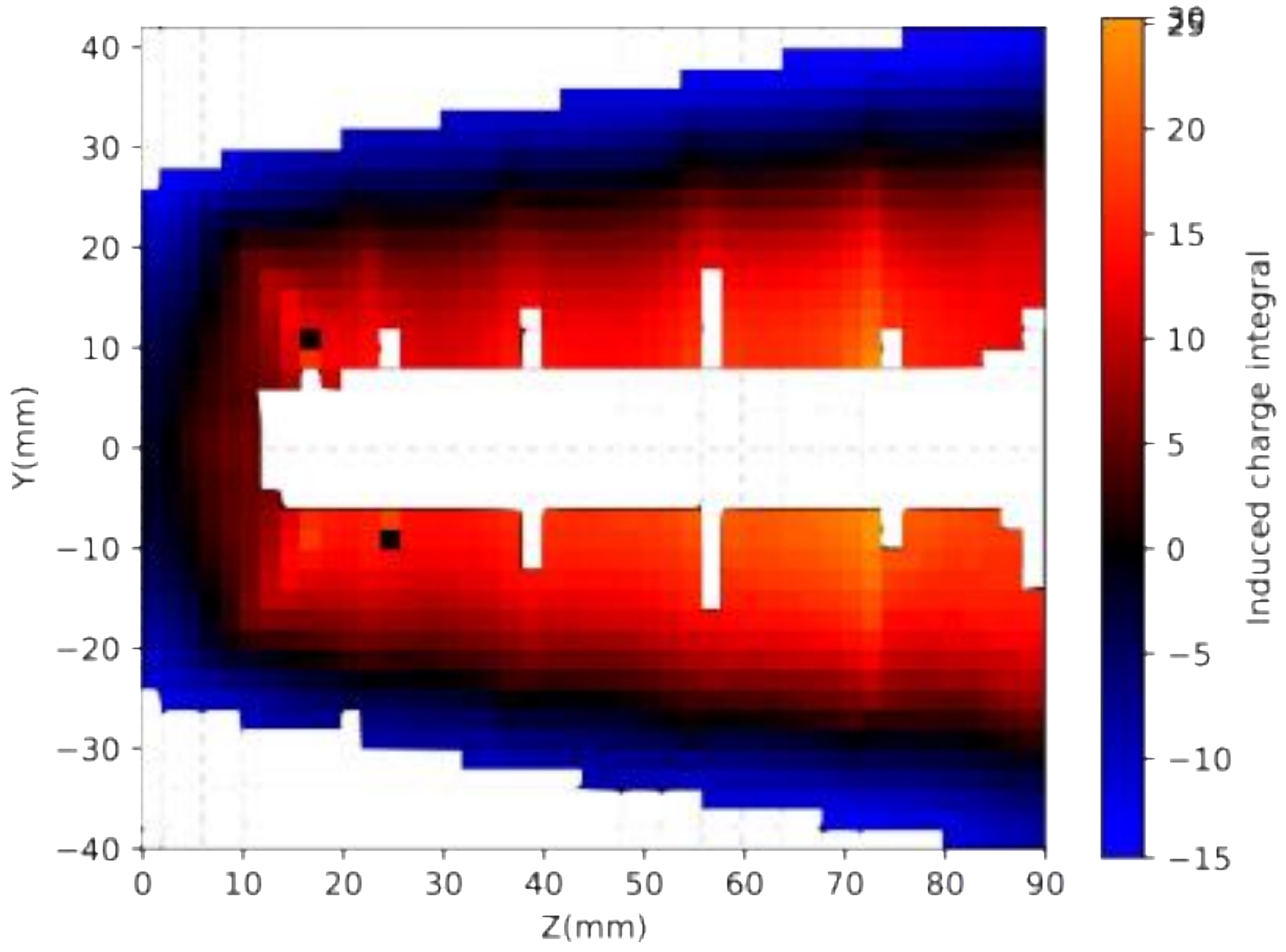
Layer 2

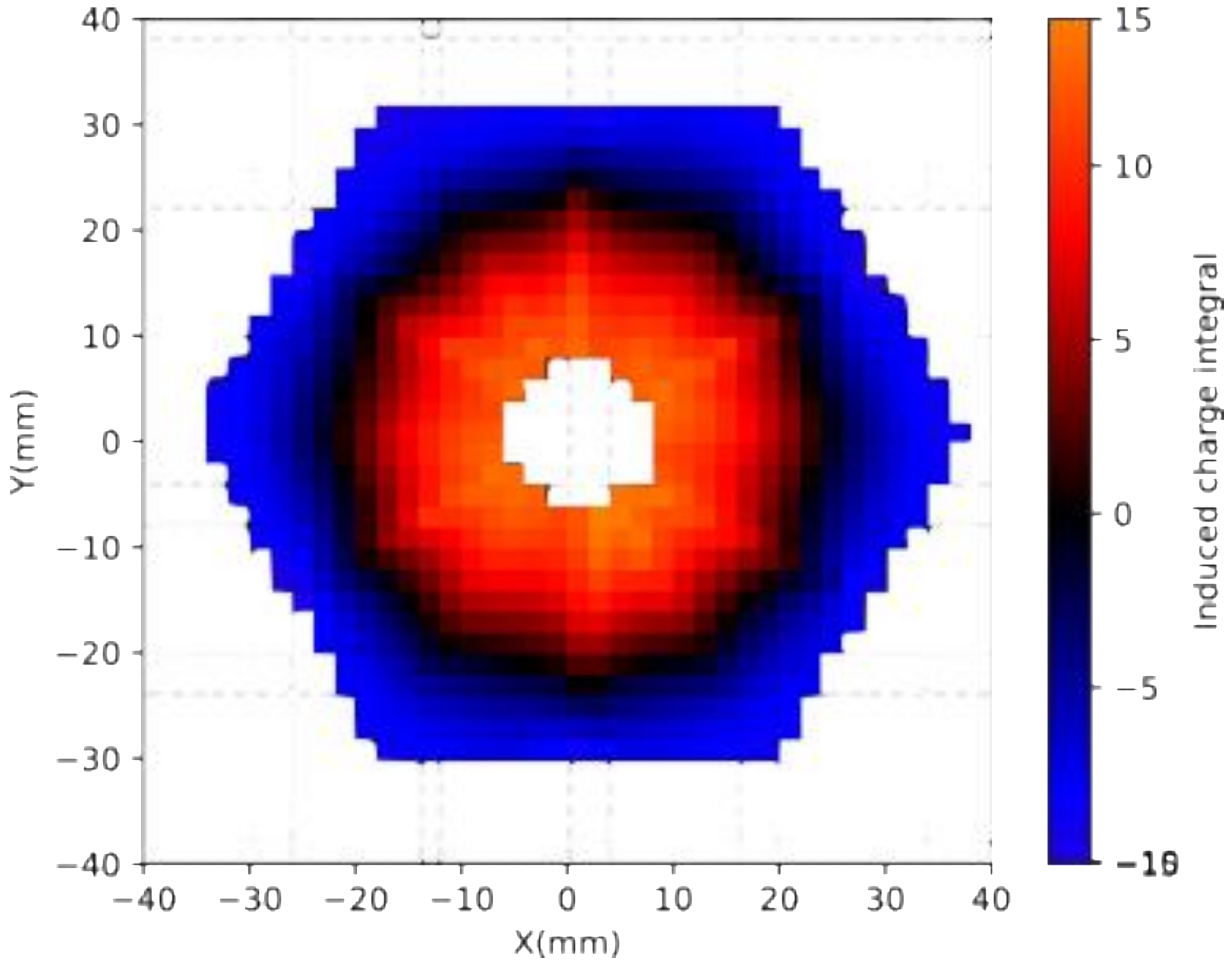


S001: PSA results

- PSA algorithm effect:
  - The clustering effects in the middle of the segment is enhanced by the algorithm:
  - The clustering regions corresponds to voxels where the transient signal area are almost null

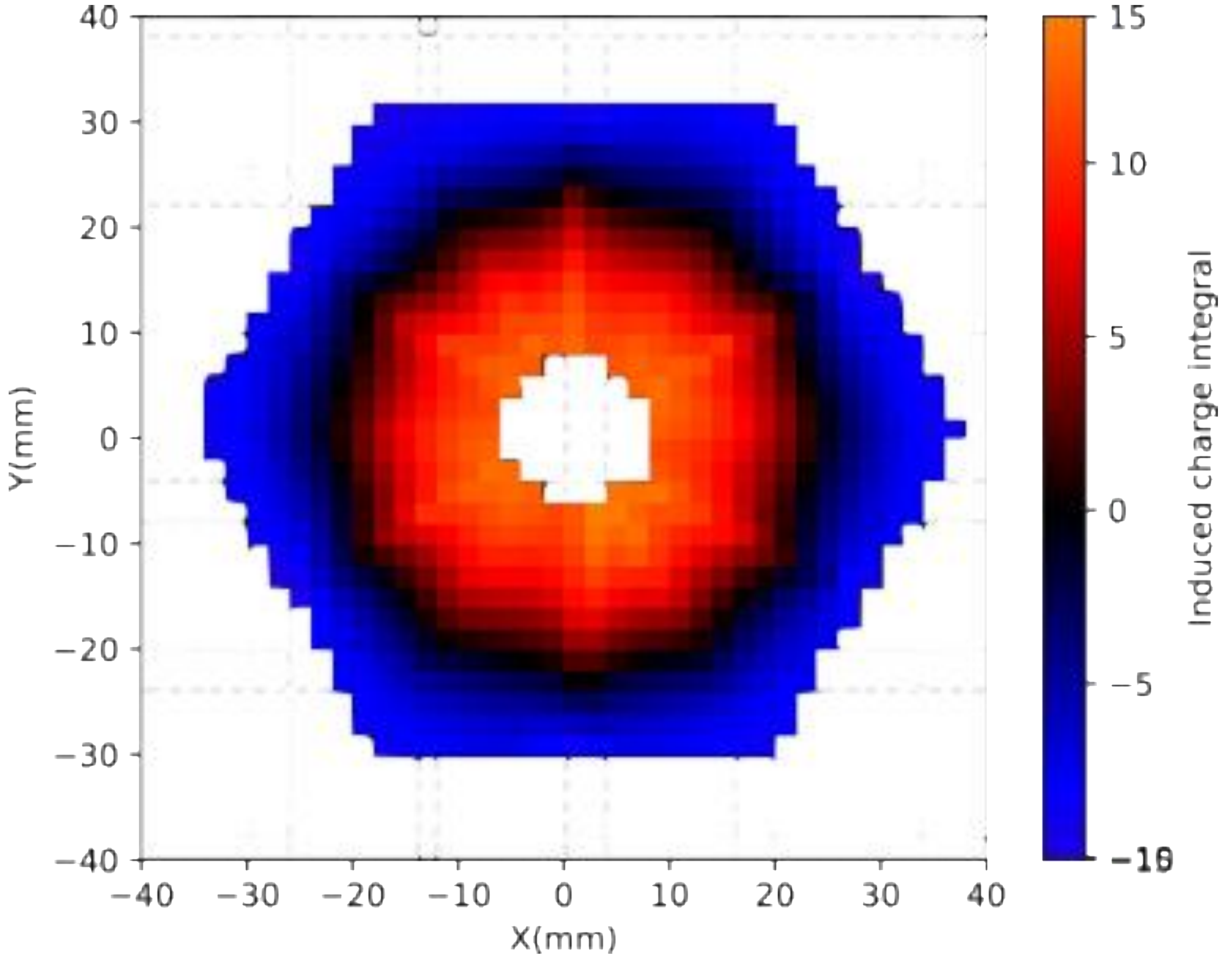
Integral of the transients signals in the first neighbouring segments

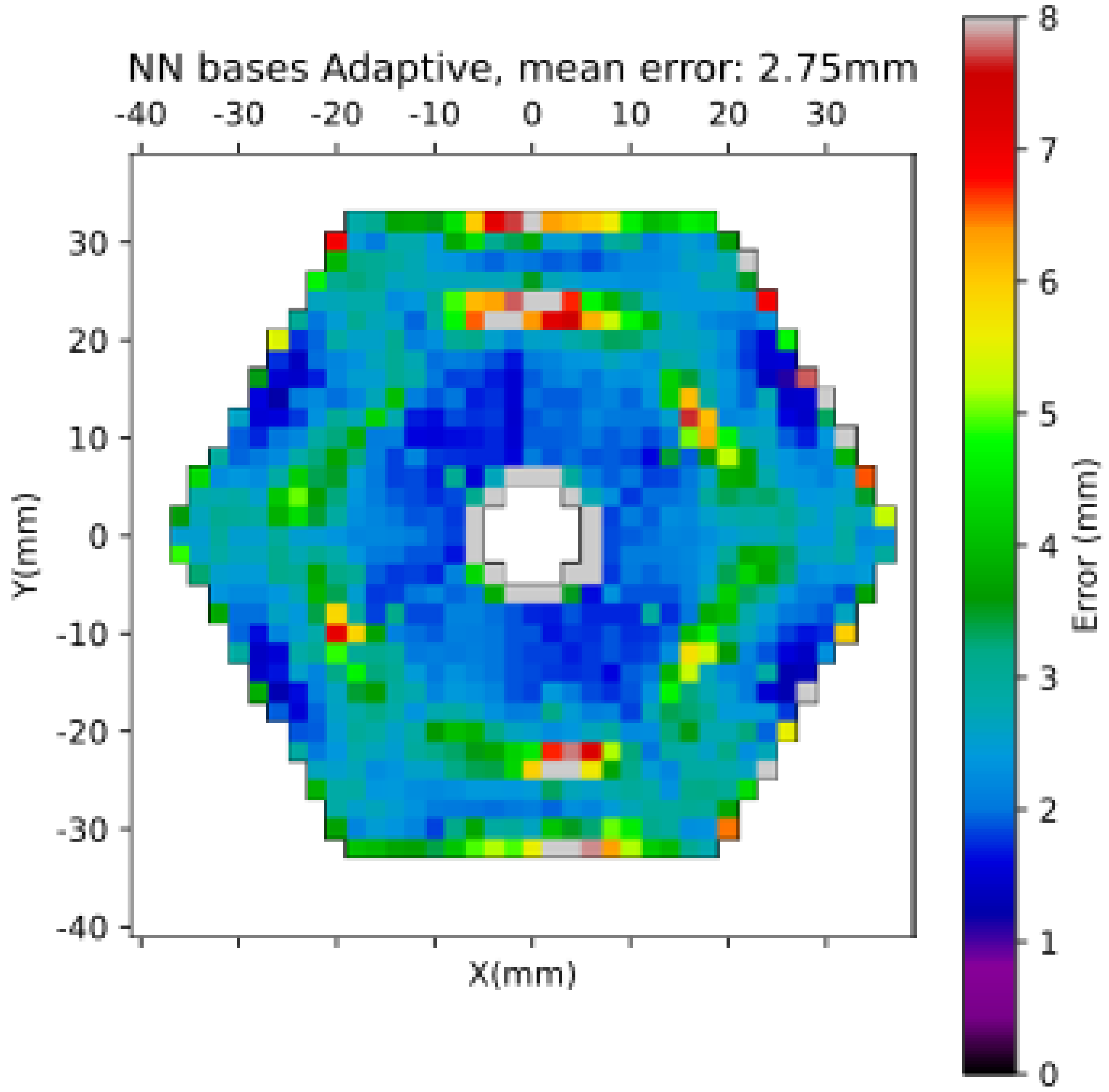




S001: PSA results

- PSA algorithm effect:
  - The clustering effects in the middle of the segment is enhanced by the algorithm:
  - The clustering regions corresponds to voxels where the transient signal area are almost null
  - In this regions, the PSA adaptive grid search is less efficient to find the good coarse grid voxel.

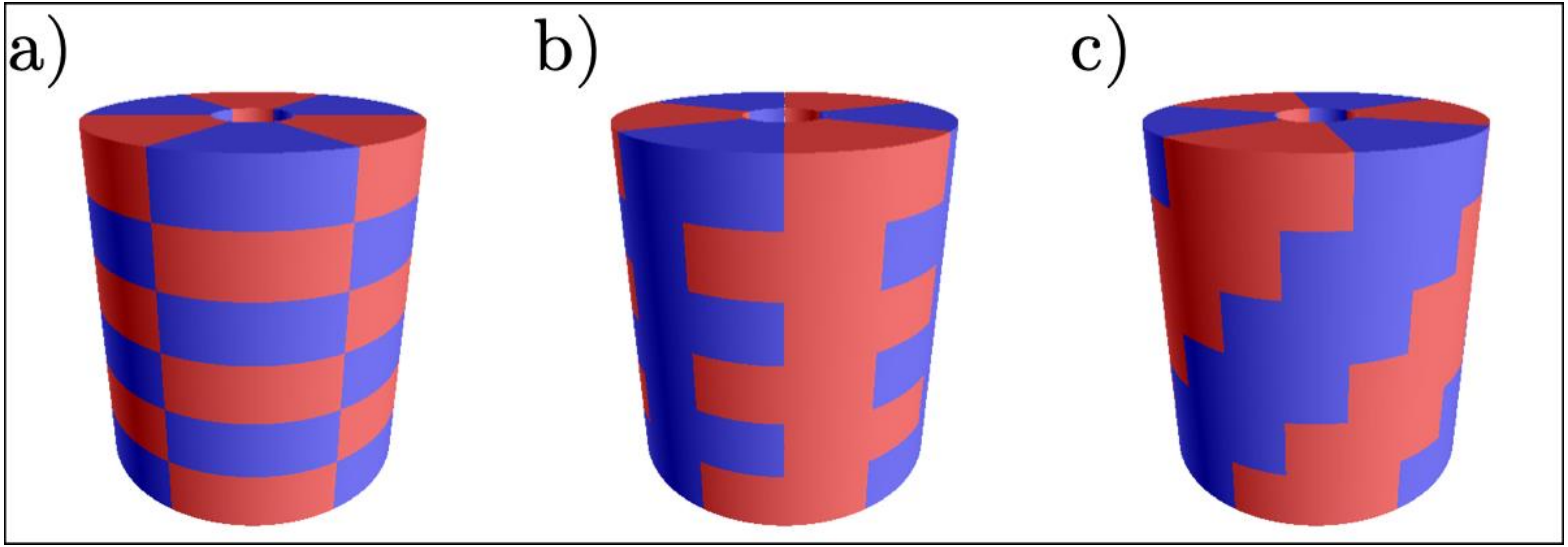




S001: PSA results

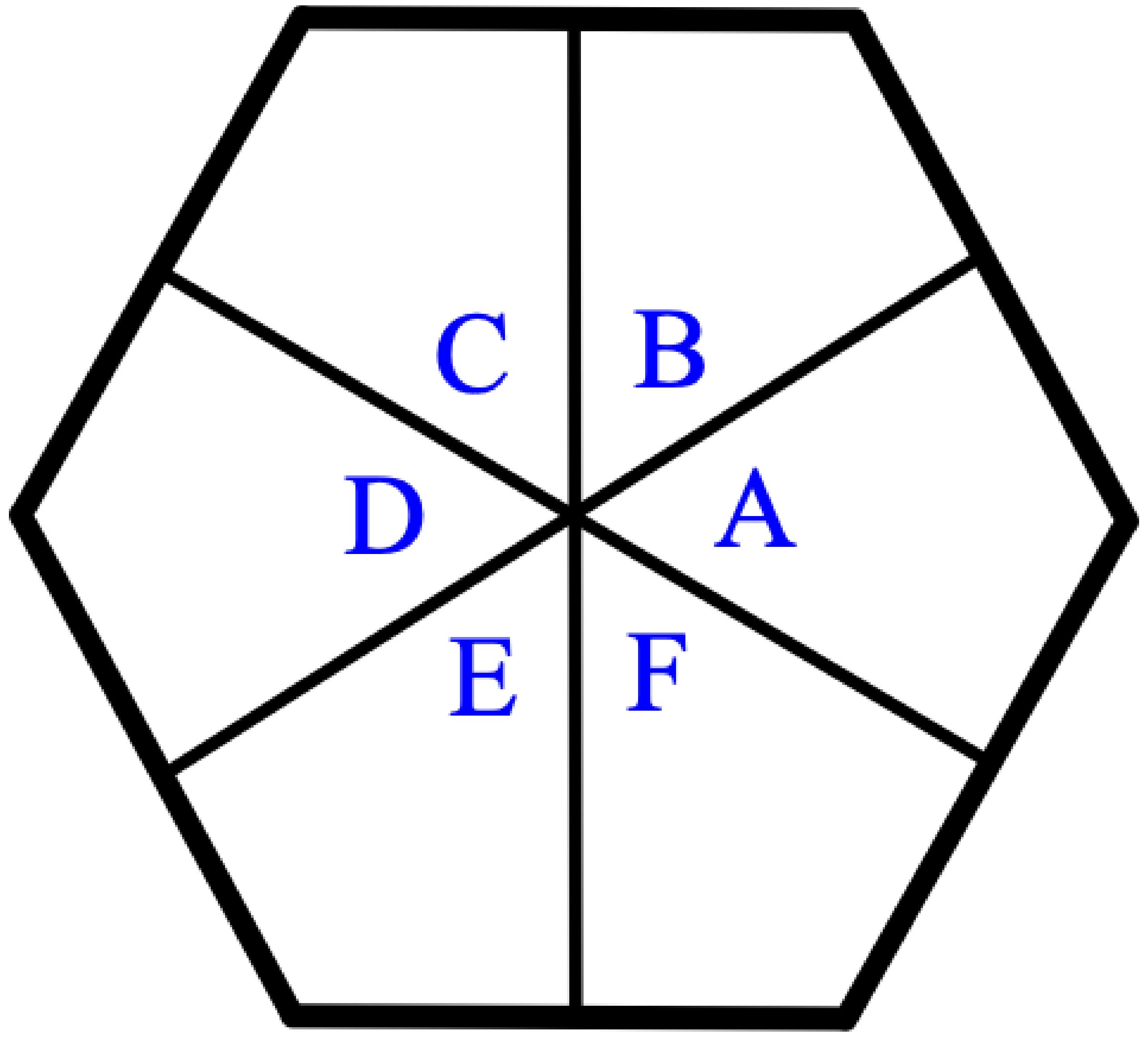
- PSA algorithm effect:
  - The clustering effects in the middle of the segment is enhanced by the algorithm:
  - The clustering regions corresponds to voxels where the transient signal area are almost null
  - In this regions, the PSA adaptive grid search is less efficient to find the good coarse grid voxel.
  - This effect is also seen in ADL basis

Idea to avoid this was proposed in the SePaGe ANR (not funded)

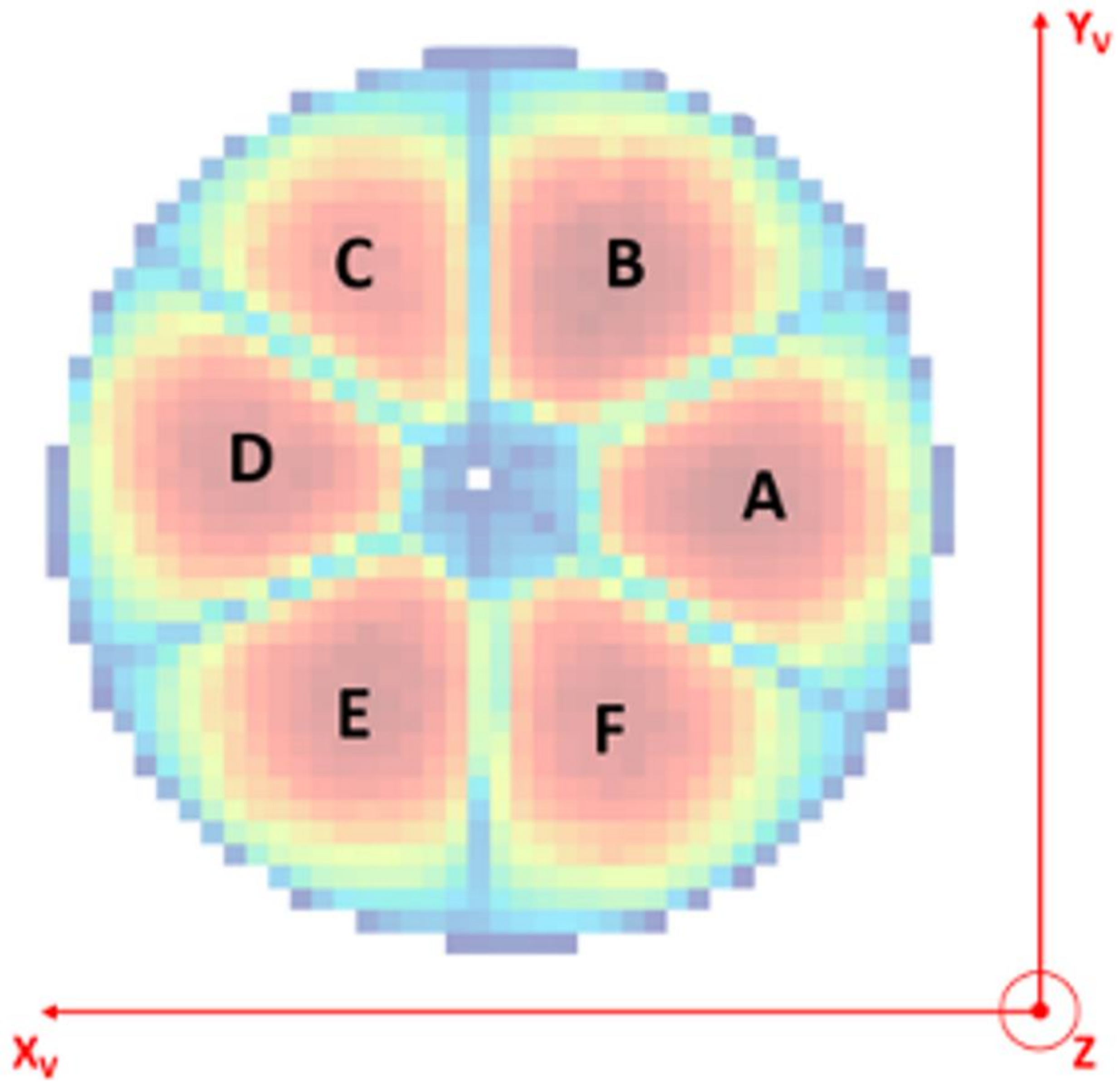


A005 scanning

- ➤ A005 crystal has been scanned previously in Liverpool and very recently in Strasbourg !
- ➤ We discovered that there was a mismatch in segments nomenclature conventions !

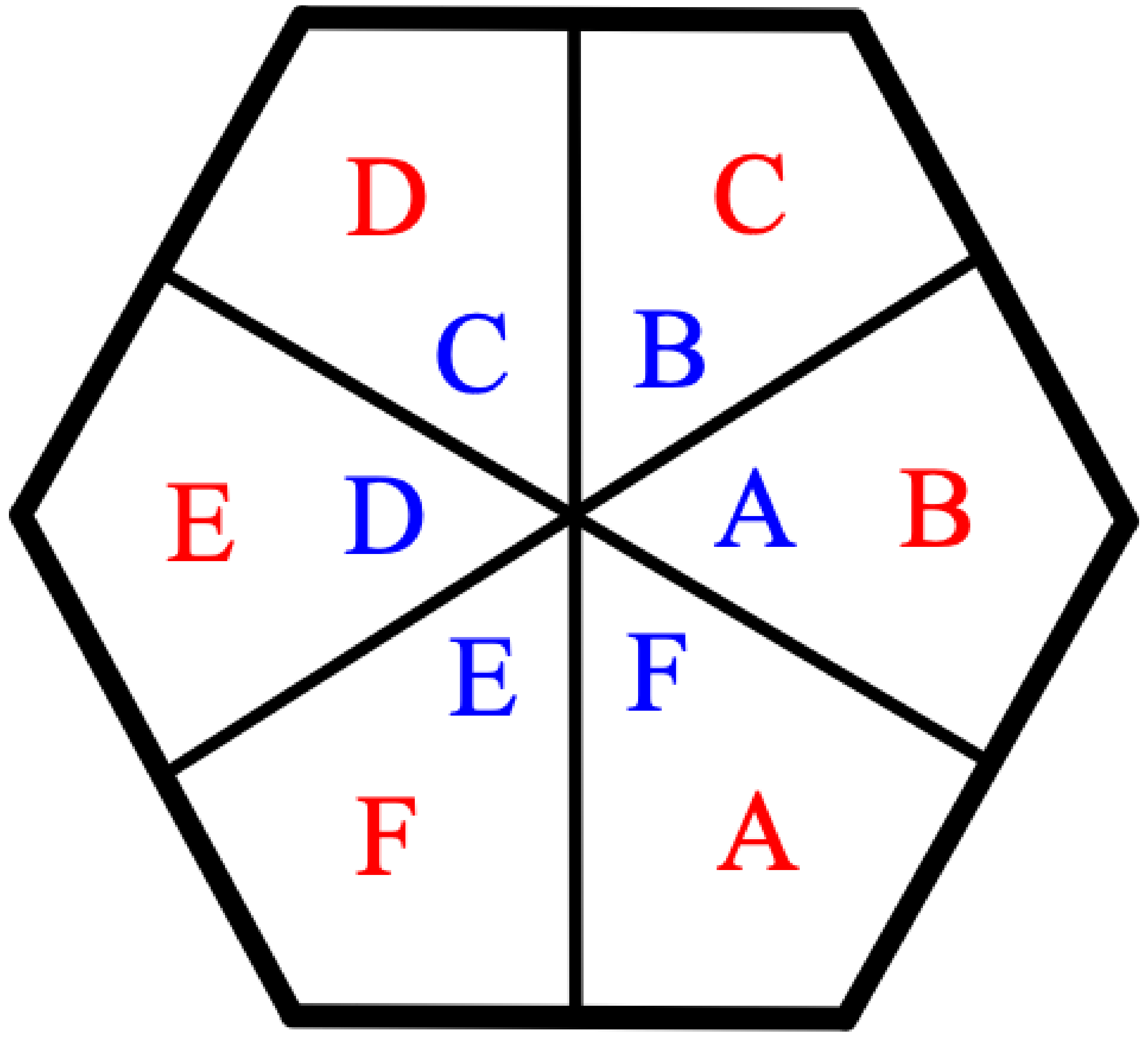


IKP (IPHC) convention

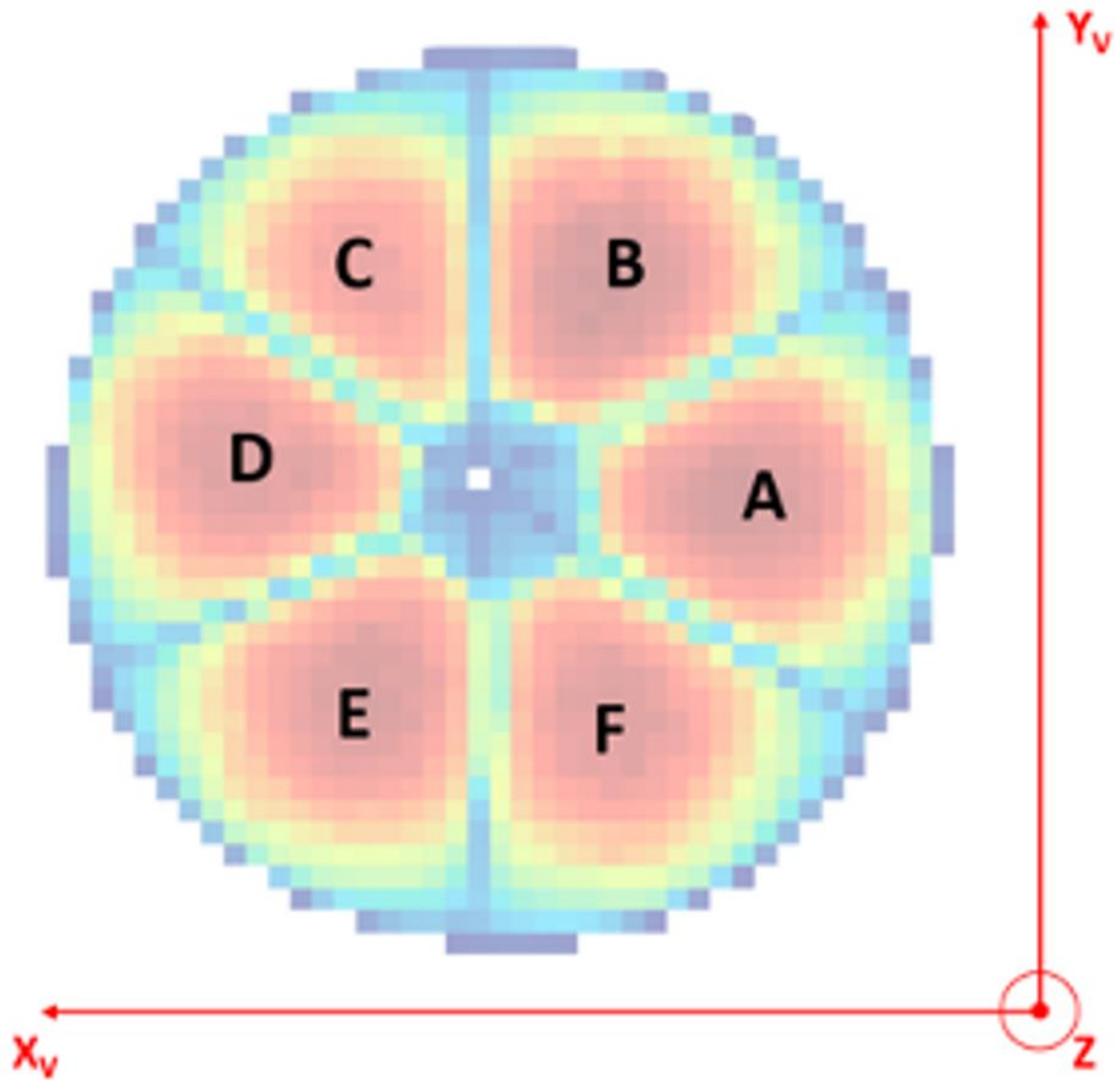


A005 scanning

- ➤ A005 crystal has been scanned previously in Liverpool and very recently in Strasbourg !
- ➤ We discovered that there was a mismatch in segments nomenclature conventions !
  - ➥ Segments A in the scanning corresponds to segments B in AGATA (ADL) nomenclature

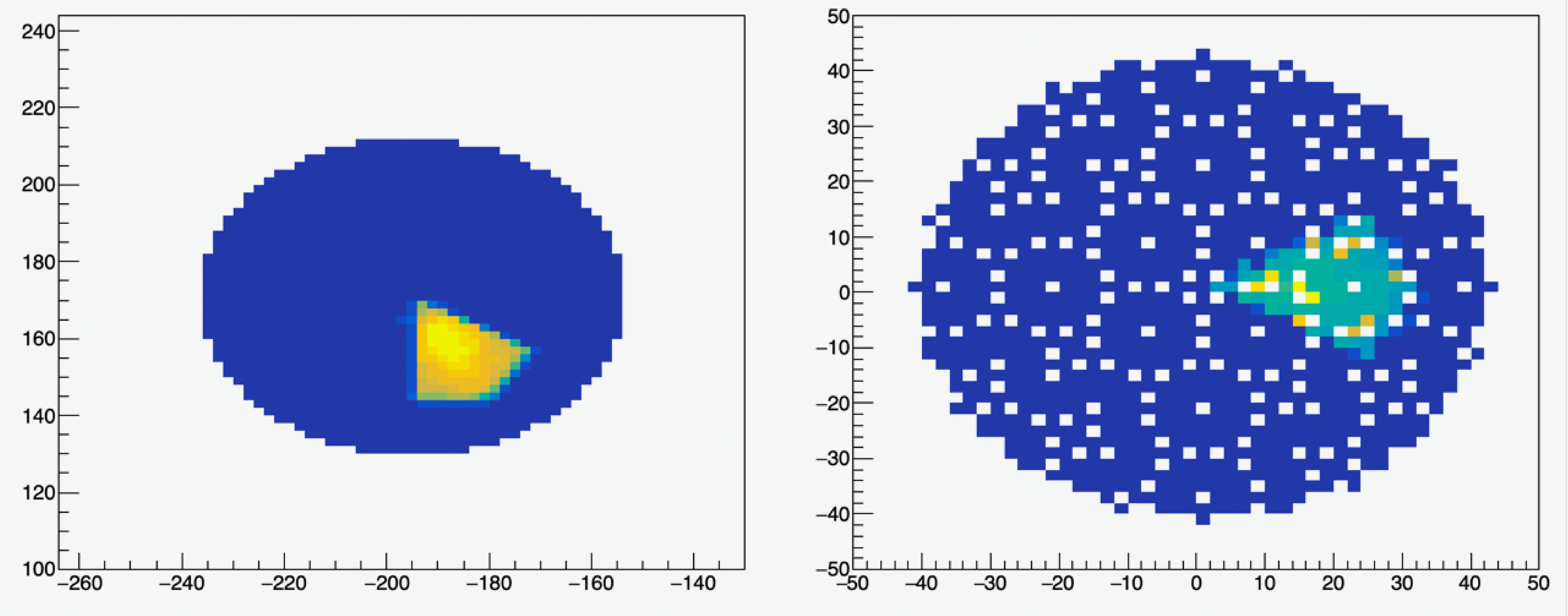


IKP (IPHC) convention
AGATA convention



A005 scanning

- How to fix this nomenclature problem to apply PSA on ADL data ?
  - ➥ Simply rotation the ADL by 60° is not possible (2mm cartesian grid basis issues)
  - ➥ Producing a new ADL basis in IKP nomenclature (who can do it ?)



A005 scanning

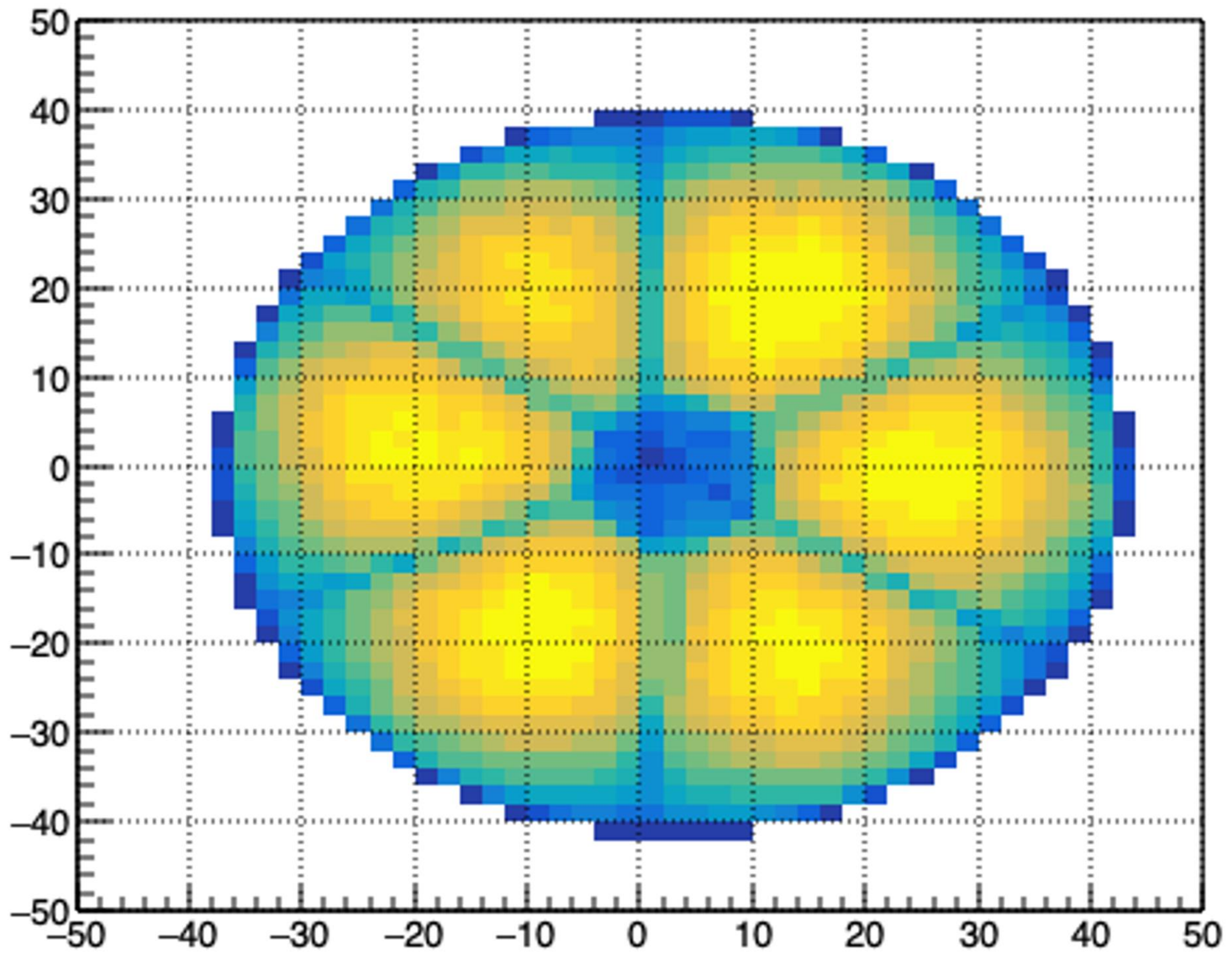
- Three steps solution !
  - Before PSA: Segments nomenclature in the data flow is switched to AGATA convention
  - The standard PSA with ADL basis in AGATA convention is performed
  - Results are rotated by ~60° to match with the scanning positions
- This will allow to analyze the scanned data with ADL, but also to analyze data taken online with A005

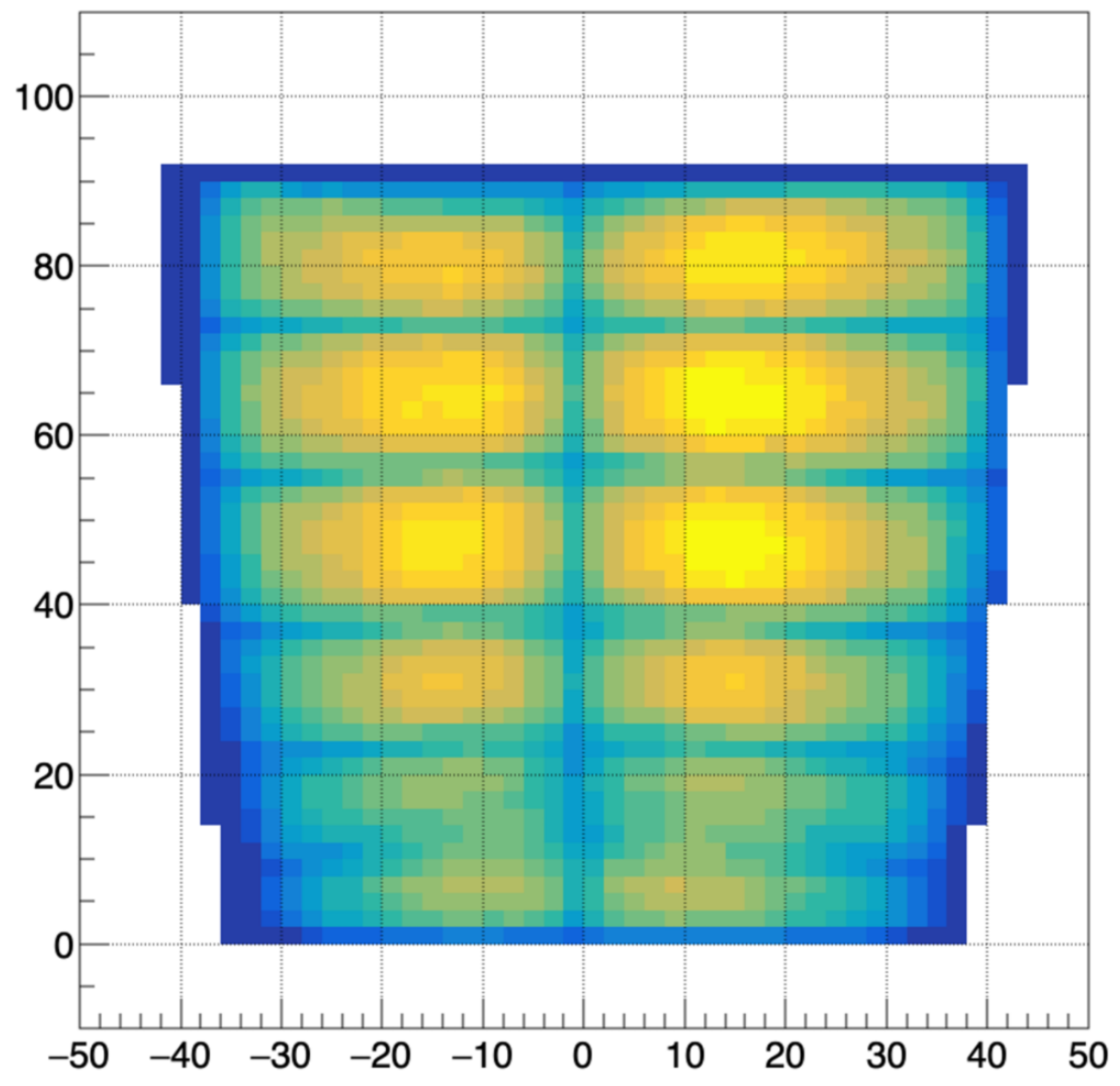
New data flow developed

TNT2_to_ADF Producer → Preprocesing Filter → IKP_to_AGATA Filter → PSA Filter → PSA rotation Filter

A005 scanning issues….

- Scanning of A005:
  - ➥ Vertical scan processed in August 2023 → Full calibration processed
  - ➥ Horizontal scan processed in February 2024 → Full recalibration that shown some scanning issues
  - ➥ New Horizontal scan processed in July 2024 → Full recalibration

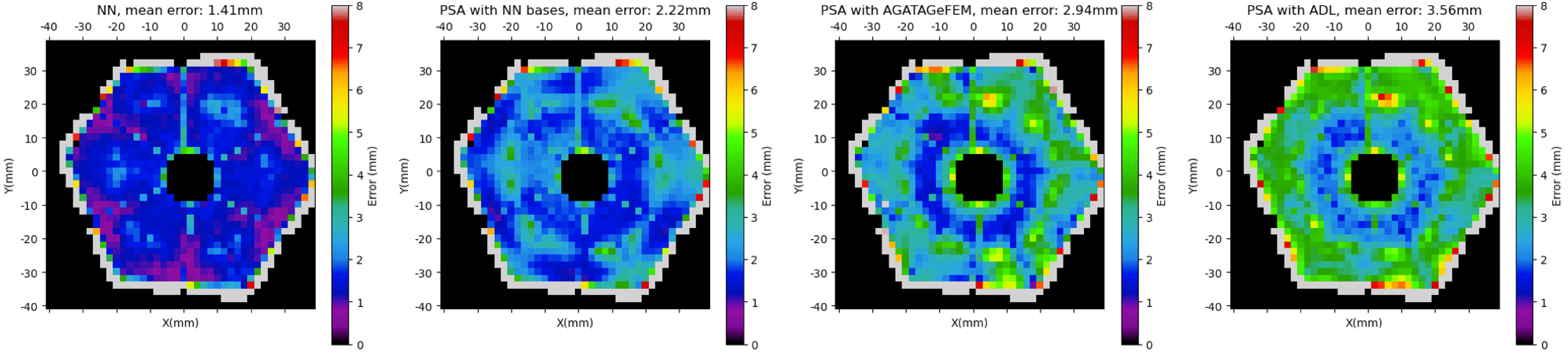




A005: Results

- Comparison of:
  - NN results
  - PSA with NN basis
  - PSA with AGATAGeFEM, rotated to IKP convention (thanks Joa)
  - PSA with ADL, using IKP to AGATA filter and PSA rotation filter

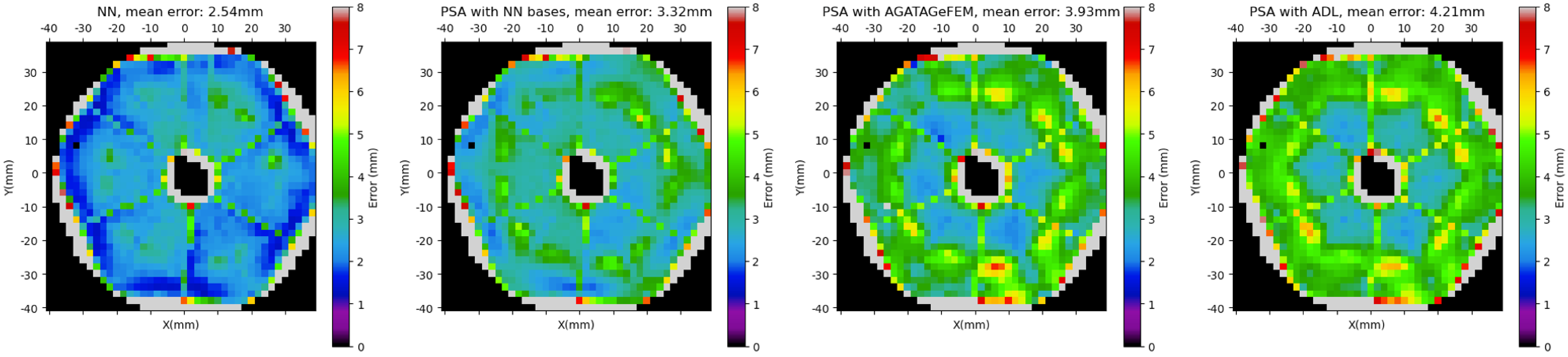
Layer 1



A005: Results

- Comparison of:
  - NN results
  - PSA with NN basis
  - PSA with AGATAGeFEM, rotated to IKP convention (thanks Joa)
  - PSA with ADL, using IKP to AGATA filter and PSA rotation filter

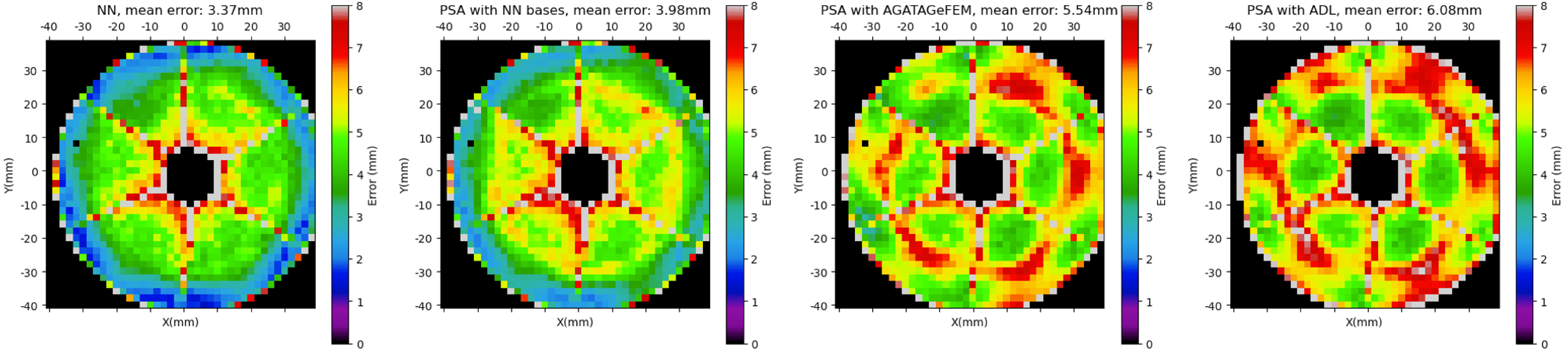
Layer 3



A005: Results

- Comparison of:
  - NN results
  - PSA with NN basis
  - PSA with AGATAGeFEM, rotated to IKP convention (thanks Joa)
  - PSA with ADL, using IKP to AGATA filter and PSA rotation filter

Layer 5



On average, the results comply with AGATA specifications, but there is room for improvement in the regions where the hot spots are located.

A005: Results

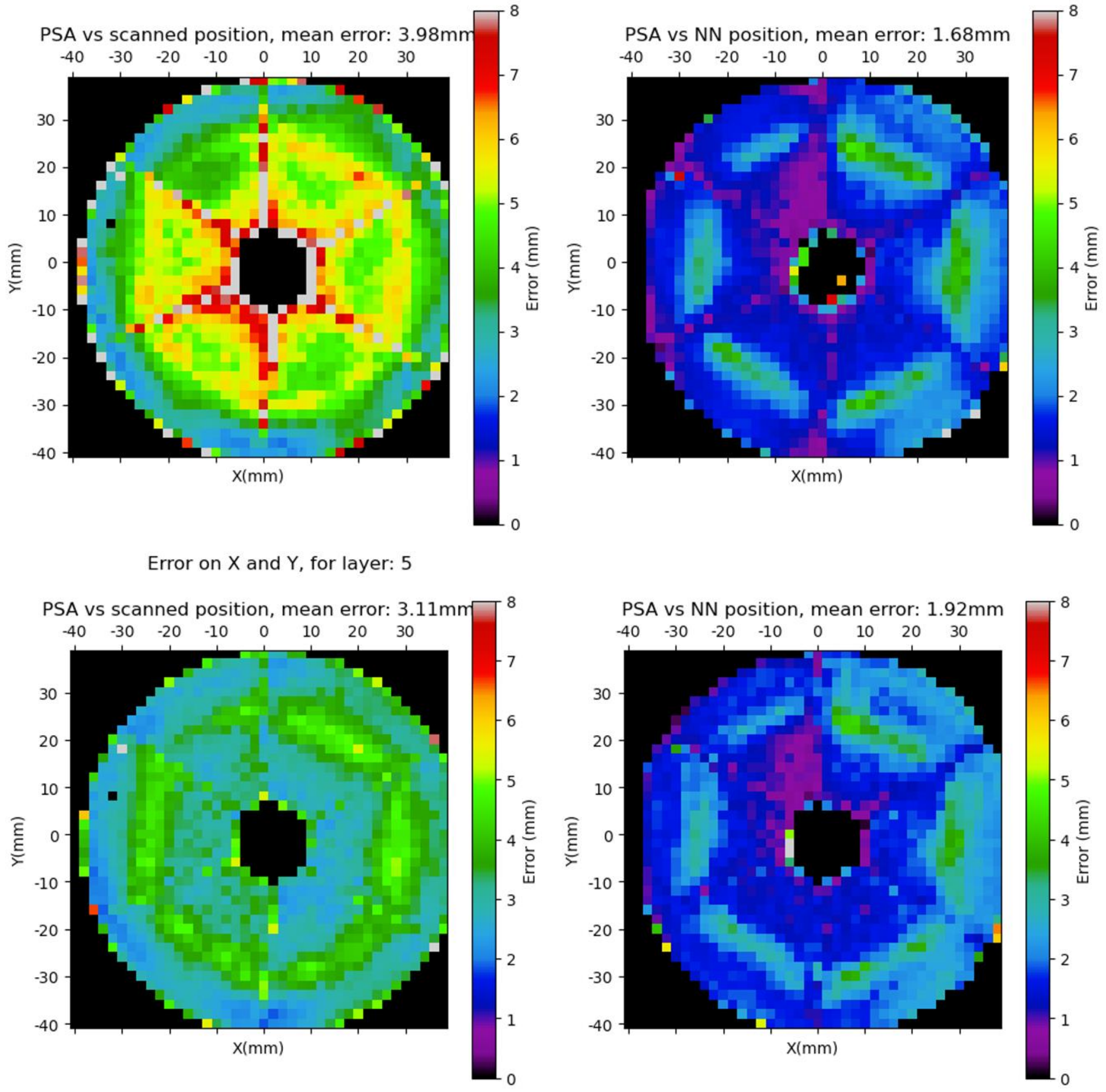
- Up to which level can we trust the scanned positions ?
  - ➥Given values are relative to the scan positions, but what about **random Compton** or **multiple hits in one segment ?**
- To test the robustness of the NN, we trained it with 50% of the data with bad random labels
  - ➥ after training, the network was still able to predict the good positions

A005: Results

- ➤ Up to which level can we trust the scanned positions ?
  - ➥Given values are relative to the scan positions, but what about **random Compton** or **multiple hits in one segment ?**
- ➤ To test the robustness of the NN, we trained it with 50% of the data with bad random labels
  - ➥ after training, the network was still able to predict the good positions
- ➤ Selecting only photopeak events (limiting random Compton events), the results are much better in the last layer

All energies

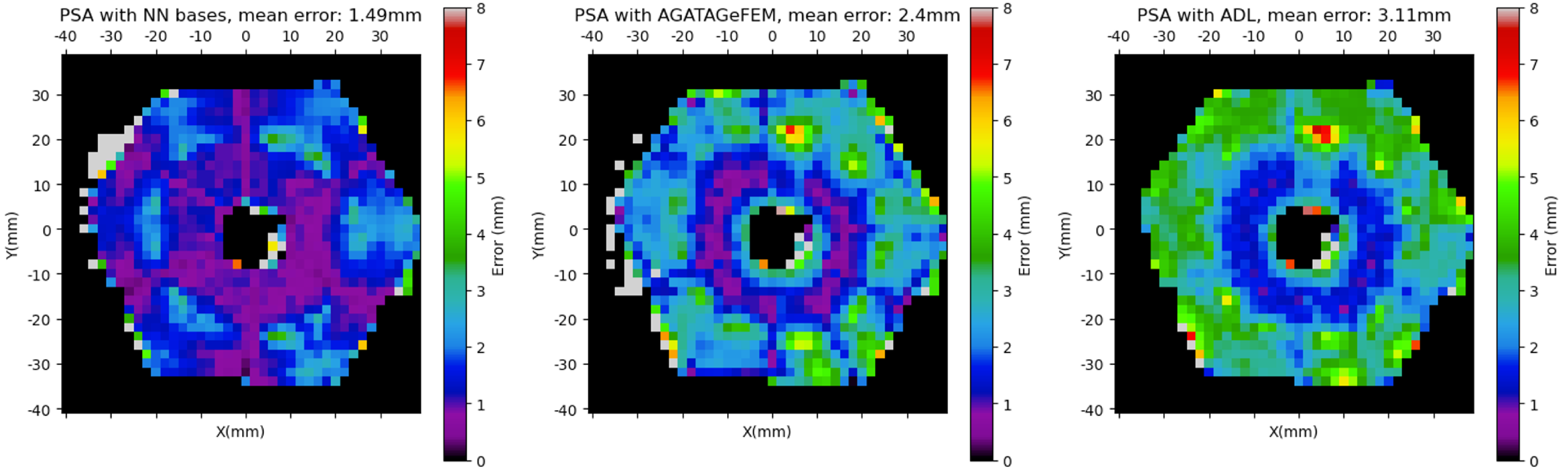
Photopeak only



A005: Results

- Comparison between PSA basis compared to NN predictions (assumed to be more precise than scanning):
  - ➥ NN experimental basis
  - ➥ AGATAGeFEM basis
  - ➥ ADL basis

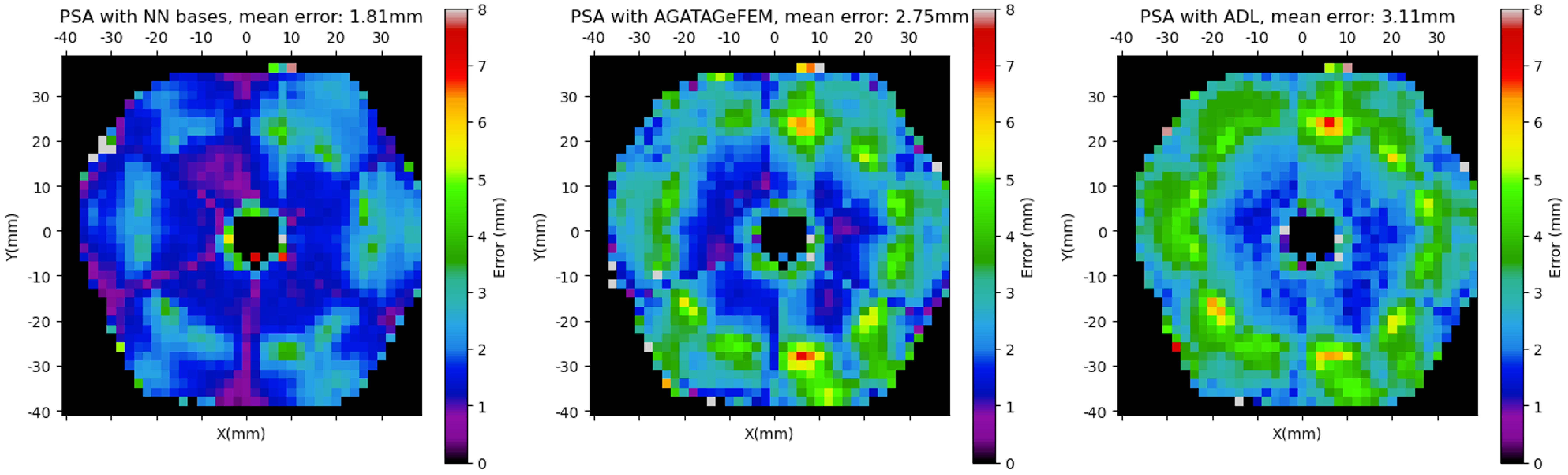
Layer 1



A005: Results

- ➤ Comparison between PSA basis compared to NN predictions (assumed to be more precise than scanning):
  - ➡ NN experimental basis
  - ➡ AGATAGeFEM basis
  - ➡ ADL basis

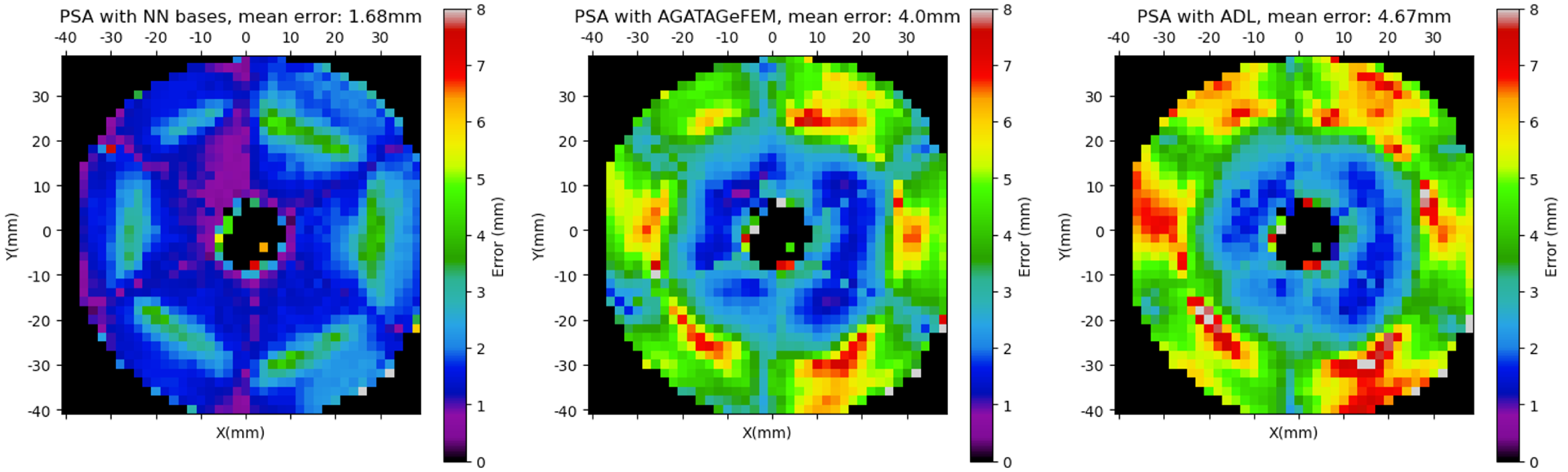
Layer 3



A005: Results

- ➤ Comparison between PSA basis compared to NN predictions (assumed to be more precise than scanning):
  - ➥ NN experimental basis
  - ➥ AGATAGeFEM basis
  - ➥ ADL basis

Layer 5

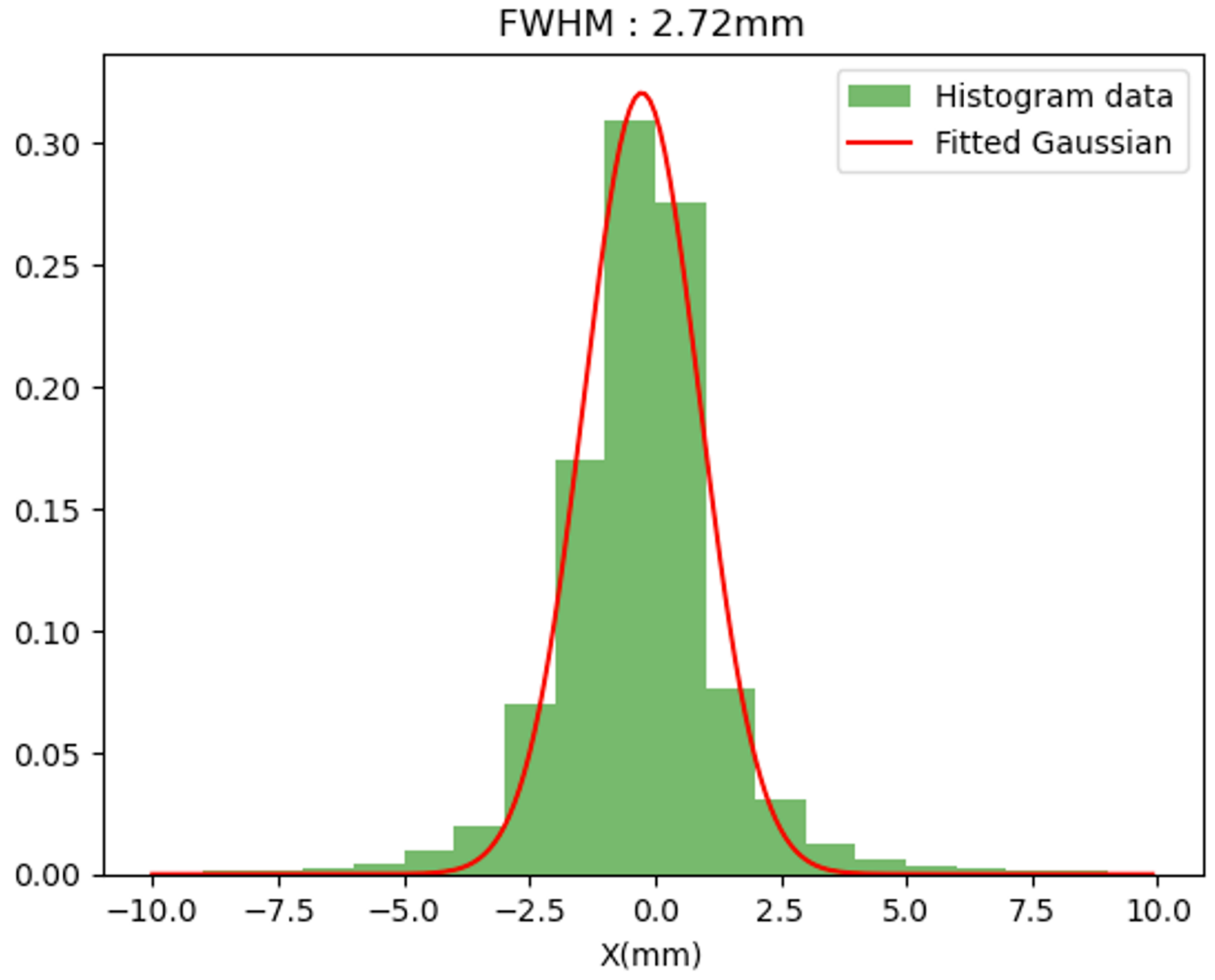


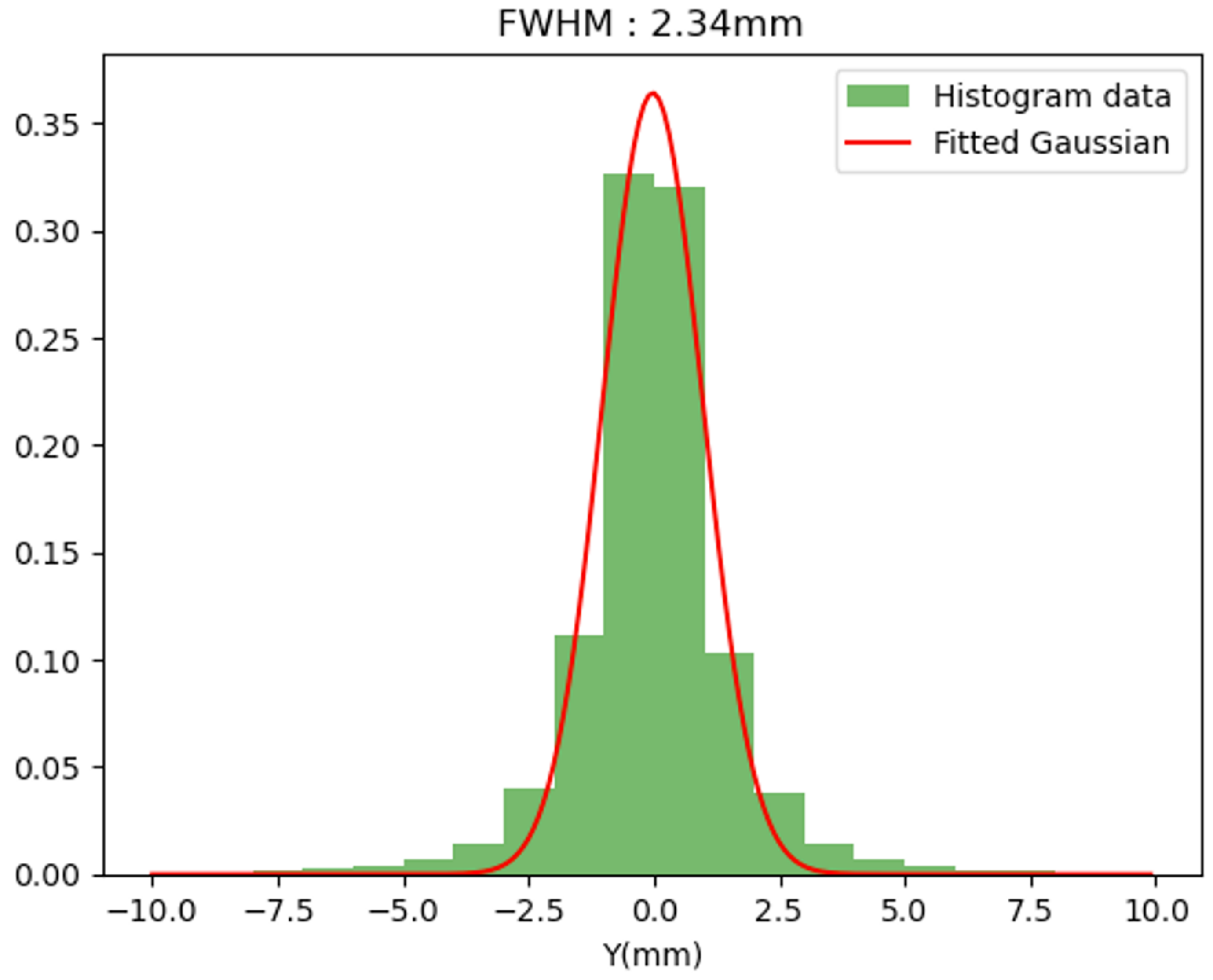
➔ Effects inherent to the PSA algorithm

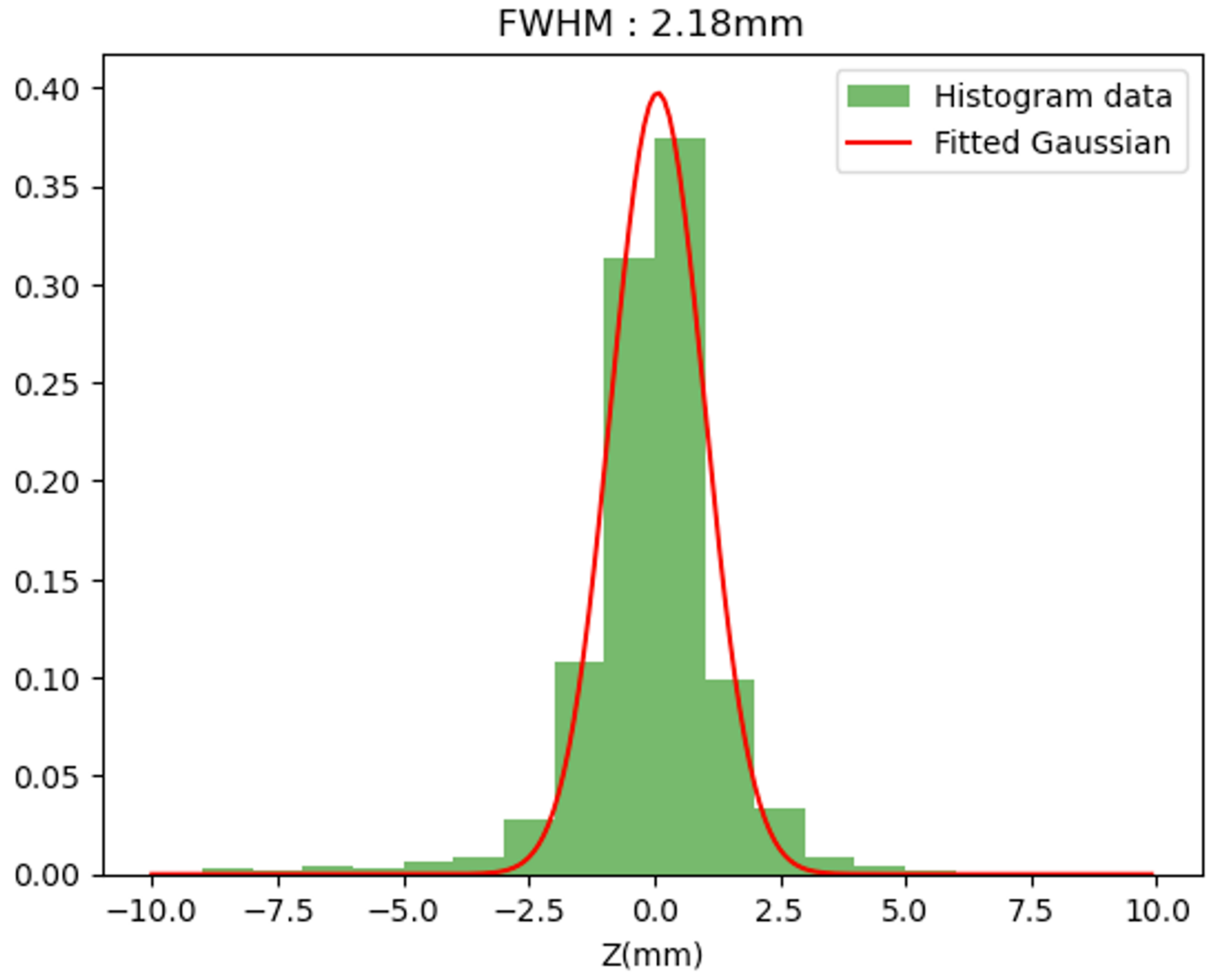
➔ Effects coming from the basis simulations (ADL ~ AGATAGeFEM)

A005: Results

- Average PSA Position resolution (FWHM):
  - ➥ NN: **2.4 mm**

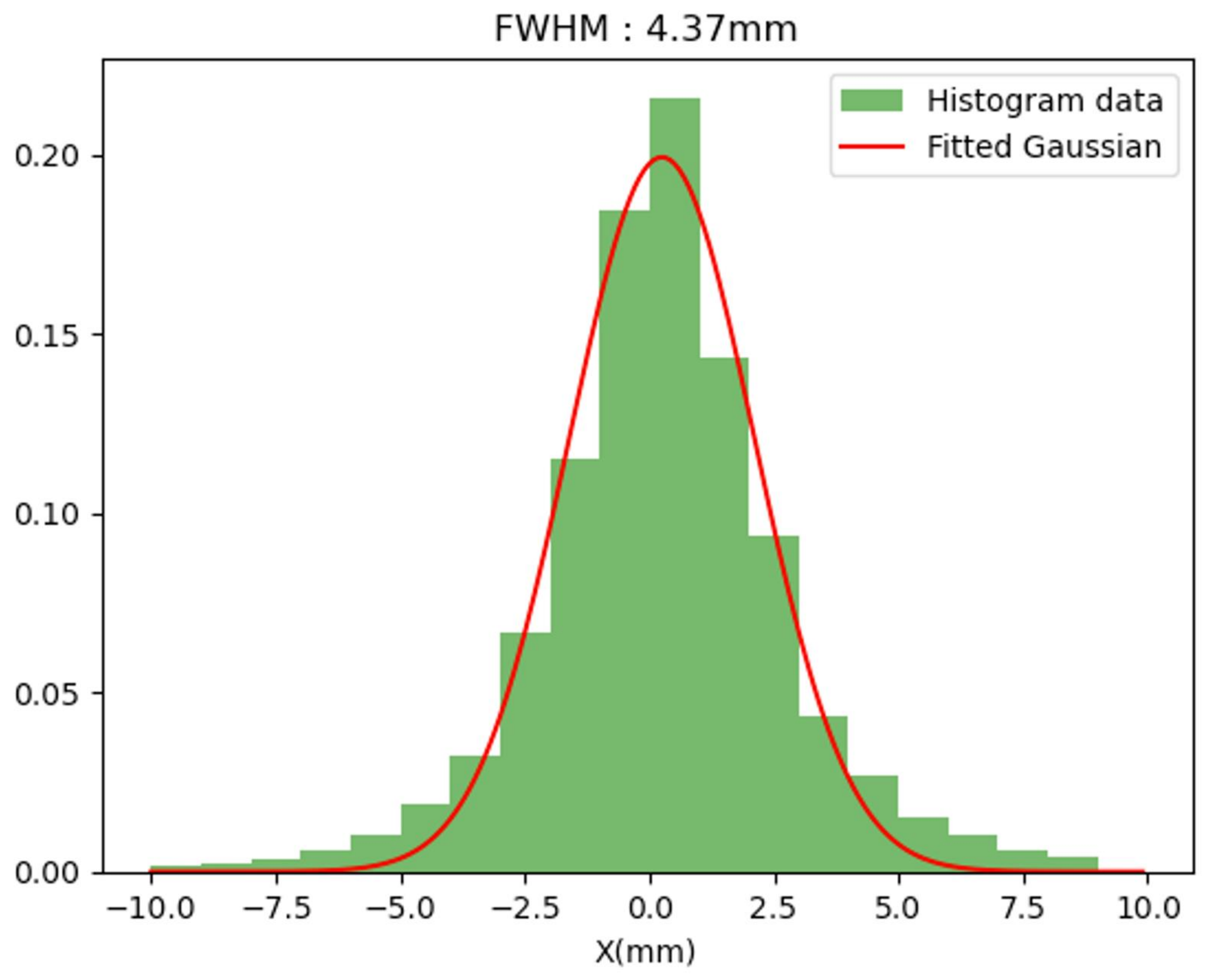


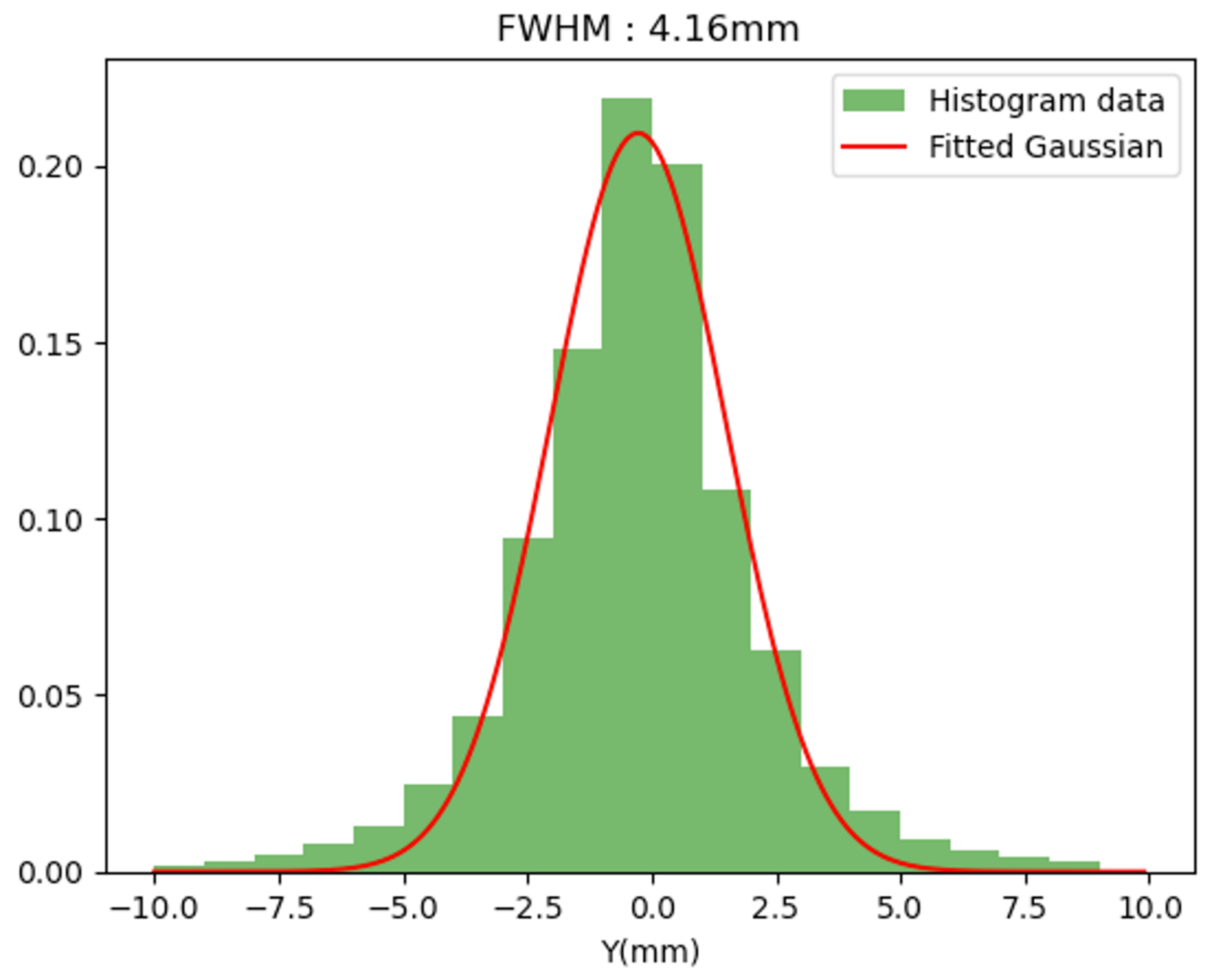


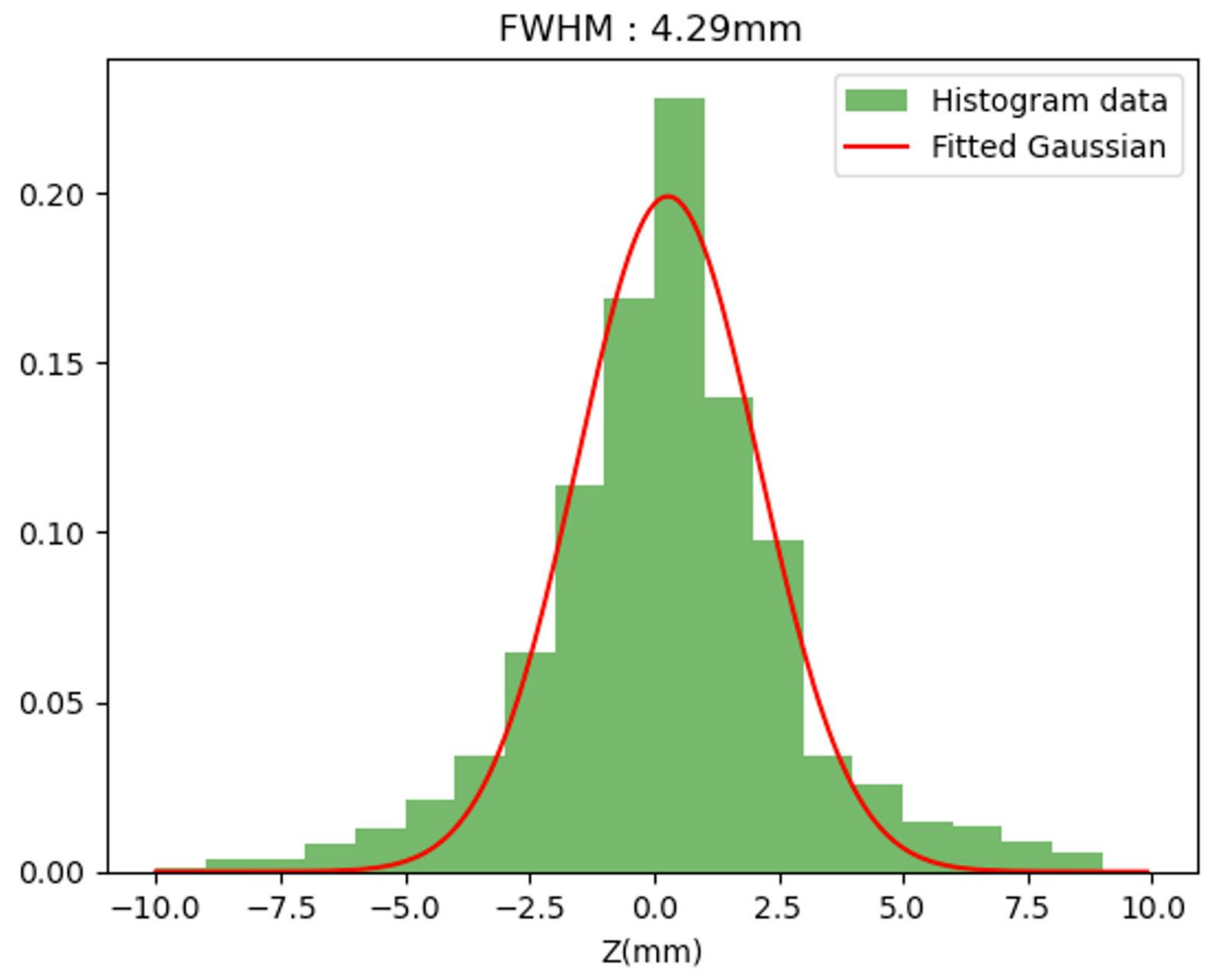


A005: Results

- Average PSA Position resolution (FWHM):
  - ➥ NN: **2.4 mm**
  - ➥ AGATAGeFEM **4.3 mm**

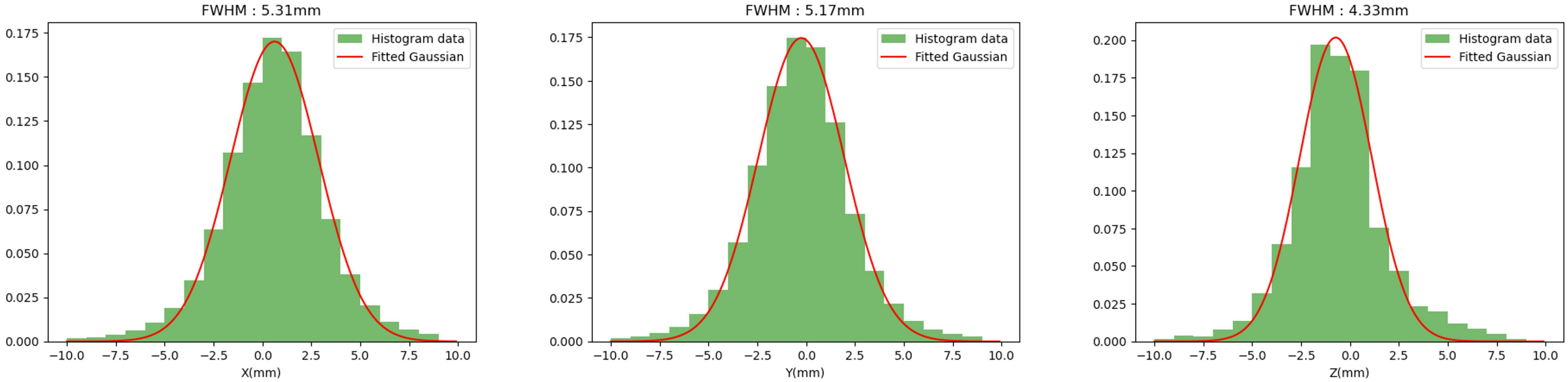






A005: Results

- Average PSA Position resolution (FWHM):
  - NN: **2.4 mm**
  - AGATAGeFEM **4.3 mm**
  - ADL **4.9 mm**



Conclusions and perspectives

Conclusions:

- ➤ Still preliminary results (new horizontal scan analysis not yet included in the NN analysis)
- ➤ But already an impressive quantitative and qualitative work performed by Mojahed !
- ➤ New results already very useful for a better understanding of the PSA performances

Perspectives:

- ➤ Finalize A005 analysis
- ➤ Explore AGATA data taken online with A005 and compare NN basis with simulated ones
- ➤ Raw and calibrated data (ADF) will be uploaded on the AGATA iRODS to be accessible by the collaboration.

Personal comments:

- ➤ We should push toward a standard use of AGATAGeFEM basis, because the full expertise is in the collaboration (Joa)
- ➤ We should push to have an AGATA electronic on the scanning tables to really have comparable results with online data

Merci !