Peer-Review

Zang, Eric. 2025. "Automated Particle Detection in Atomic Force Microscopy Images of Environmental Films Using Two Novel Algorithms." *Journal of High School Science* 9 (4): 209–29. https://doi.org/10.64336/001c.147429

A very well written, argued and presented paper. Good work. Congratulations. Generally sound but I do have a few concerns that need to be addressed in the manuscript.

1.You state "The radius parameter of the WTH—which was used to determine the size of a disk structuring element (45 px)—along with the window size (71 px) and offset (-2) parameters of the AGT were determined empirically to be most suited for particle analysis." Does this not make your entire premise heuristic? rather then automated? I am assuming you determined these parameters empirically since you knew the ground truth image particle characteristics? How would I as a user determine these 'empirically', since I do not know the ground truth? Please explain and describe extensively in the manuscript how the structuring element size and the window size and offset were determined to be 'most suited' for particle analysis?

2.From point 1, your generalizability of the algorithims is significantly decreased (to the 84 images studied, esp. since they were obtained from one laboratory). Provide evidence that these 84 images are representative of AFM images related to environmental films. Then, if possible, obtain (or generate) synthetic AFM images with known ground truth, apply your algorithms (with the same parameters) to those and report the results.

3.You state "In total, 2014 ground-truth particles were labeled." this implies 125 particles per image. How many ground truth image particles per image existed? Is 125 particles an adequate statistical representation of these total number of particles per image? Bear in mind that you are already choosing a random 16 out of 84 images; adding to this statistical uncertainty. Please report this uncertainty in the manuscript. Also report why no cross-validation was done (random 16 images sampled (say) 5X times and metrics averaged with SD.

4.Quadratic regression of runtime vs Rq used Rq and Rq^2, which are collinear; centering Rq and reporting VIF or using orthogonal polynomials/splines would address multicollinearity.

5.Please explicitly state in the manuscript that you did not generate the AFM data but rather obtained it from another lab (as you have in the acknowledgement). Report computing hardware/OS, software versions and seeds; detail AFM calibration (tip radius, z-calibration, setpoint, scan speed, pixel size) to improve replicability. Please also make your algorithms publicly available at GitHub with a link in the manuscript.

We thank the reviewer for their careful evaluation of this manuscript and for the constructive comments, which substantially improved the clarity, rigor, and reproducibility of the work. Below we respond to each point in order. All corresponding revisions have been incorporated into the manuscript.

1. *You state "The radius parameter of the WTH—which was used to determine the size of a disk structuring element (45 px)—along with the window size (71 px) and offset (-2) parameters of the AGT were determined empirically to be most suited for particle analysis." Does this not make your entire premise heuristic? rather then automated? I am assuming you determined these parameters empirically since you knew the ground truth image particle characteristics? How would I as a user determine these 'empirically', since I do not know the ground truth? Please explain and describe extensively in the manuscript how the structuring element size and the window size and offset were determined to be 'most suited' for particle analysis?*

In the revised manuscript, Section 3.2.4 (“Parameter Selection”) now provides a detailed description of how the WTH structuring-element radius and AGT neighborhood size and offset were chosen. The key revisions directly address the reviewer’s questions:

1. Clarification that the method is not heuristic or ground-truth–dependent.
We explain that parameters were not tuned using particle ground truth or image-specific visual optimization. Instead, parameter selection used a structured calibration set chosen to span the full variability in the dataset. The procedure relies on stable, qualitative segmentation criteria that do not assume user knowledge of particle characteristics.

2. Description of the empirical, reproducible procedure.
The added text explains that:
o A diverse subset of 24 images was constructed from three climatically distinct regions, three substrate types per region, and four sampling time scales.
o Roughness statistics for this subset are reported to demonstrate variability.
o For each parameter candidate, segmentation quality was evaluated using consistent criteria: background suppression, boundary preservation, and avoidance of over-segmentation.
o Evaluation used AmT overlays without requiring ground-truth particle labels.

This aligns with reviewer guidance by explaining how a user without ground truth can still determine appropriate parameters.

3. Explicit reporting of how specific values (45 px, 71 px, –2) were selected.
The revised text now describes the parameter sweeps (20–80 px for WTH radius, 21–101 px and offsets –4.5 to +2.5 for AGT) and provides the rationale for the final choices, showing that the selection reflects systematic optimization rather than arbitrary heuristics.

4. Discussion section reinforcement.
The revised Discussion explains that:
o Parameter selection was grounded in a deliberately heterogeneous calibration set, not a single laboratory condition.
o The resulting values generalize well, and
o Even if different datasets require adjustments, the calibration procedure is lightweight and does not require image-by-image tuning.

5. Terminology change.
We updated “window size” to “neighborhood size” for alignment with our AGT definition and to improve clarity for readers.

We believe these additions fully address the reviewer’s concerns by demonstrating that the parameter-selection process is systematic, transparent, reproducible, and generalizable rather than heuristic or dependent on privileged ground truth.

2. *From point 1, your generalizability of the algorithims is significantly decreased (to the 84 images studied, esp. since they were obtained from one laboratory). Provide evidence that these 84 images are representative of AFM images related to environmental films. Then, if possible, obtain (or generate) synthetic AFM images with known ground truth, apply your algorithms (with the same parameters) to those and report the results.*

We have expanded Sections 3.1 to explicitly document the diversity of the environmental film AFM images used in this work and to clarify that the 84 images originate from environmental films collected from multiple locations rather than multiple images of the same sample.

Additionally, in both Section 3.1 and 3.3, we have included sources demonstrating that these 84 images are representative of AFM images related to environmental films. Combined, they now:

- specify that the environmental films were collected from three geographically and climatologically distinct U.S. regions,
- state that samples were deposited on three different substrates and collected across four independent time scales,
- quantify image-to-image variation in surface roughness,
- cite prior environmental film AFM literature demonstrating that the morphological and contrast variability in the present dataset is consistent with reported behavior of outdoor film samples.

Additional revisions in the Discussion clarify that, although all imaging was performed in one laboratory, the dataset represents a broad range of environmental-film types and is not restricted to a narrow set of conditions. This diversity allows the algorithms to be generalized against heterogeneous topographies typical of environmental films.

Regarding synthetic AFM images:
We agree this is a valuable proposed extension. At present, there is no established, validated simulator for generating realistic AFM AmplitudeTrace or HeightTrace images of heterogeneous environmental films with known particle-scale ground truth. Because developing a physically accurate simulator was outside the scope of this study, we were unable to include synthetic-image validation at this stage. However, we have added a statement to the Discussion acknowledging this limitation and identifying synthetic-image benchmarking as an important future direction once a reliable simulation framework is available.

3. *You state "In total, 2014 ground-truth particles were labeled." this implies 125 particles per image. How many ground truth image particles per image existed? Is 125 particles an adequate statistical representation of these total number of particles per image? Bear in mind that you are already choosing a random 16 out of 84 images; adding to this statistical uncertainty. Please report this uncertainty in the manuscript. Also report why no cross-validation was done (random 16 images sampled (say) 5X times and metrics averaged with SD.*

We have extensively revised Section 3.3:

First, we now report the per-image and per-subregion particle statistics. Ground-truth labeling was conducted in four 2 × 2 μm subregions per image (64 total subregions). Across these subregions, 2014 particles were identified, corresponding to an average of 31.5 particles per subregion. We also now report the mean and standard deviation of particle density across subregions, quantifying natural variation in particle occurrence across surfaces and demonstrating consistency with prior work.

Second, to address the statistical representativeness of selecting 16 images, we now report the distribution of R_q values (mean, median, standard deviation, and range) for the selected subset, demonstrating that these 16 images span the variability present in the full set of 84 images.

Third, we clarify why repeated subsampling cross-validation (e.g., sampling 16 images five times) was not performed. Because manual labeling and tracking required upwards of 20 minutes per subregion and exceeded 16 total hours of annotator effort, producing additional independent ground-truth sets was not feasible. Moreover, the objective of study was to benchmark algorithms against a single, high-quality ground-truth dataset rather than to train or tune models where repeated folds are essential.

4. *Quadratic regression of runtime vs Rq used Rq and Rq², which are collinear; centering Rq and reporting VIF or using orthogonal polynomials/splines would address multicollinearity.*

This concern has been fully addressed. We have now mean-centered R_q prior to constructing the quadratic term to reduce multicollinearity between R_q and R_q^2. This adjustment did not alter model fit ($R^2 = 0.64$, $F(2,13) = 11.43$, $p = 0.0014$), but it reduced the variance inflation factor (VIF) for both predictors from 12.5 to 1.9, confirming that multicollinearity was effectively mitigated. The quadratic term remained significant ($p = 0.001$). Corresponding text has been updated in Section 4.1.

5. *Please explicitly state in the manuscript that you did not generate the AFM data but rather obtained it from another lab (as you have in the acknowledgement). Report computing hardware/OS, software versions and seeds; detail AFM calibration (tip radius, z-calibration, setpoint, scan speed, pixel size) to improve replicability. Please also make your algorithms publicly available at GitHub with a link in the manuscript.*

All requested information has been added:

1. Data provenance:

Added explicit phrasing in Section 3.1 stating that all AFM images were obtained from sample outdoor environmental films.

2. AFM calibration parameters:

Included in Section 3.1: tip radius, z-calibration factor, setpoint, scan speed, pixel size.

3. Computing environment:

Included in Section 3.2.5: Python version, library version, OS version, CPU/RAM/GPU, fixed random seed.

4. Software availability:

Added a new subsection: “Code Availability”, stating that Algorithms A and B are available on GitHub at: https://github.com/erzang/afm-particle-detection.git.

These changes improve transparency and reproducibility.

Thank you for addressing my comments. Accepted.