Credibility assessment of CFD models for vibrating mesh nebulizer droplet size distribution under ASME V&V 40

Alexandra N. Pierce[a], Rohan M. Desai[b], Mei-Lin Zhang[c], Samuel J. Whitaker[d], Priya K. Subramanian[e]

[a] *Inhalation Technology Center, MedAero Devices, Cambridge, MA, USA,*
[b] *Department of Mechanical Engineering, Northeastern University, Boston, MA, USA,*
[c] *Computational Science Division, National Aerosol Laboratory, Oak Ridge, TN, USA,*
[d] *School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA, USA,*
[e] *Division of Pulmonary Drug Delivery, BioPharm Analytics, San Diego, CA, USA,*

Abstract

Computational fluid dynamics (CFD) offers the prospect of reducing bench-testing burden for aerosol-generating devices by predicting droplet size distributions under clinically relevant operating conditions. This paper applies the ASME V&V 40 risk-informed credibility framework to two CFD models of a vibrating mesh nebulizer: (i) a Mesh-Resolved Two-Phase (MR-TP) formulation for pore-scale primary atomization and near-field spectrum formation, and (ii) an Aerosol Transport and Breakup (ATB) formulation for downstream evolution through the plenum and mouthpiece. The context of use is prediction of the steady-state droplet size distribution at the mouthpiece exit to support design verification and reduction of bench testing. Given a medium model risk—owing to moderate consequence for dose accuracy and labeling and a contributory role in the overall evidence package—we executed credibility activities across six factors: spatial and temporal resolution control, structural adequacy, characterization of inputs, numerical iterative control, agreement against observations, and representativeness of setups. The topic of numerical code verification is noted but not scored. We introduce a private-scale Assurance Index (0–12) rubric aligned to V&V 40 concepts and report pair-specific ratings for each model by mechanism. Evidence for MR-TP in primary atomization is strong, with low sensitivity to grid halving (1.2% change in Sauter mean diameter), small changes due to interface-capturing method choice (2.7% in D32), and close accord with Phase Doppler Particle Analyzer (PDPA) data 2 mm from the mesh (mean difference 0.32 μm). MR-TP is less compelling for downstream secondary processes due to truncation of relevant physics (e.g., breakup, coalescence) and sensitivity to assumed flow fluctuations, producing 12% underprediction of the large-droplet tail and 0.79 μm root-mean-square error (RMSE) across size bins at 20 mm. In contrast, ATB shows limited leverage on primary atomization when relying on parametric injection kernels, but demonstrates high fidelity downstream: mesh increases from 3.1 to 11.6 million cells change the mouthpiece Dv50 by 1.3%, breakup/coalescence closure choices alter Dv90 by 2.1%, and exit laser diffraction data are matched within 0.18 μm RMSE and 3% in Dv10/50/90. We provide factor-by-pair ratings in two model-specific summary tables, discuss how each mechanism influences credibility, and outline extensions that can raise Assurance Index levels without adding new credibility dimensions. The overall outcome supports use of MR-TP to supply initial size spectra and ATB to propagate aerosol evolution through the device, with quantified confidence bounds suitable for design verification under the stated context of use.

Keywords: V&V 40, vibrating mesh nebulizer, droplet size distribution

1. Background and objective

Aerosol delivery systems based on vibrating meshes have become a mainstay in respiratory therapy for their ability to generate tightly controlled sub- to few-micron droplets at low shear and with reduced device noise. Device approvals and labeling accuracy rely on repeatable droplet size distributions, as the aerodynamic diameter spectrum dictates lung-region deposition and, ultimately, dose in the target tissue. Manufacturers historically obtain this evidence through extensive bench testing across combinations of formulation, mesh architecture, and operating settings; however, for high-dimensional design spaces and formulation variability, bench programs expand rapidly. Computational prediction of droplet size distribution, if shown credible for a focused use, can shrink the scope of testing while preserving safety and effectiveness.

The ASME V&V 40 framework provides a risk-informed structure for determining whether a model is credible for a specific context of use. The present work applies this framework to two numerical models addressing different portions of the nebulization process: a Mesh-Resolved Two-Phase (MR-TP) model that captures primary atomization at and just downstream of the mesh orifices, and an Aerosol Transport and Breakup (ATB) model that advances the polydisperse aerosol through the plenum and mouthpiece, including secondary breakup, coalescence, evaporation, and transport. By pairing models with mechanisms, we emphasize that credibility is not uniform across all physical questions a model could answer; rather, it is localized to how the model is used. Our objective is to establish a transparent, evidence-based Assurance Index on a private 0–12 scale for each model–mechanism pair under the context of use of predicting steady-state droplet size distribution at the mouthpiece exit, and to present the rationale for how the index reflects the strength of evidence.

V&V 40 asks that model risk be evaluated first, as it governs how much credibility is required. We classify the risk as medium: predictions contribute to design verification and justification for bench-testing reduction but are not the sole evidence for labeling; the decision consequence is moderate (mis-estimated droplet sizes can lead to label inaccuracies, mitigated by confirmatory bench assessments and the narrow operating envelope of the device). This risk level drove our scope of credibility activities and the stringency of the Assurance Index rubric detailed herein.

Model risk alignment — gradation.

a.
Use not tied to regulatory or labeling; negligible decision impact; informal rationale acceptable.

b.
Contributory evidence for noncritical decisions; qualitative reasoning with limited quantitation.

c.
Moderate consequence for design verification and labeling; quantitative credibility activities with traceable evidence required.

d.
Exclusive or near-exclusive basis for high-consequence decisions; exhaustive credibility with conservative margins mandated.

2. Context of use: Predicting steady-state droplet size distribution at the mouthpiece exit to support design verification and bench-testing reduction

The context of use specifies the boundary of the decision supported by the models in this paper. The output of interest is the steady-state droplet size distribution at the mouthpiece exit for a vibrating mesh nebulizer delivering isotonic saline (0.9% w/w) at ambient laboratory conditions. The prediction will be used to (i) confirm that a given mesh architecture and drive setting produce a target Dv50 and narrow Dv10–Dv90 span under ISO 27427-compliant test setups, and (ii) justify, for pre-specified combinations of formulation viscosity and surface tension within a validated envelope, that a reduced bench-testing matrix is sufficient because computational predictions fill coverage gaps.

Two models participate in the prediction. MR-TP yields the near-field spectrum at millimetric distance from the orifices by resolving the liquid–air interface, surface tension, and contact-line motion on the oscillating mesh. ATB starts from a handed-off initial droplet distribution and advances the aerosol through the plenum and the mouthpiece, accounting for interfacial processes and gas–droplet interactions, and returns the distribution at the mouthpiece exit plane. The two are coupled in a one-way fashion: MR-TP informs ATB; ATB does not feed back to the pore-scale solution.

Within this context, we restrict the scope of credibility activities to factors that control the trustworthiness of the predicted distribution and its traceability to bench tests: the adequacy of spatial and temporal resolution, the suitability of the mathematical structure to the mechanism, the characterization of input parameters and their variability, the control of numerical iteration and time-integration errors, the agreement of the model outputs with observations collected at defined locations, and the degree to which the simulated setups represent the test conditions. The topic of code-level verification is acknowledged as an ancillary activity conducted outside the present scoring scope.

Context-of-use clarity — gradation.

a.
Ill-defined outputs, users, and decisions; no link to physical tests.

b.
Broadly defined outputs and decision role; partial linkage to tests.

c.
Specific outputs, locations, and decision pathways; explicit test linkage.

d.
Fully specified outputs with quantitative acceptance criteria and fully mapped decision logic; formal traceability matrix to tests.

3. Model risk assessment: Medium risk based on moderate decision consequence and contributory model influence

We position the model risk as medium following V&V 40 guidance by decomposing consequence and influence. Consequence is moderate: an erroneous size spectrum could mischaracterize lung deposition fractions and labeling, but compensating controls exist. In particular, (i) design verification will not rely solely on computational outputs, (ii) acceptance criteria at the mouthpiece exit are cross-checked by standard methods (e.g., laser diffraction, impactor-based estimation where applicable), and (iii) the device operates in a stable regime with limited cross-coupling across settings. Influence is contributory: models provide quantitative support for trends and interpolations but do not replace targeted bench tests at critical settings. Together, these factors justify a credibility program with substantial quantitation but not exhaustive exploration of all modeling uncertainties.

Given this risk stance, we define a private-scale Assurance Index to grade the strength of evidence for each factor under each model–mechanism pair. The 0–12 scale maps to qualitative anchors aligned with V&V 40: 0–3 (minimal evidence), 4–6 (limited evidence), 7–9 (moderate to strong evidence), and 10–12 (very strong evidence). We calibrate the scale using exemplars from published aerosol CFD studies and internal historical datasets, and require that each rung be supported by documented analyses and, where applicable, quantitative cross-checks against measurements.

Assurance Index calibration — gradation.

a.
No mapping; numeric grades are ad hoc and unanchored.

[redacted]

Figure 1: Overview of the nebuliser mesh geometry, load cases and the mechanisms assessed, shown across the full page width.

b.
Coarse mapping to qualitative labels; limited exemplars.

c.
Calibrated against internal and external exemplars; quantitative anchors defined.

d.
Calibrated with formal uncertainty targets and decision-theoretic thresholds; inter-rater reliability established.

4. Device and mechanisms: Vibrating mesh architecture, operating envelope, and mechanistic pathways to droplet size outcomes

The vibrating mesh nebulizer under study comprises a thin nickel-alloy plate (50–70 μm thickness) with 4000–7000 funnel-shaped micro-orifices arranged in a hexagonal lattice over an active diameter of 8–12 mm. Orifices are laser-drilled and subsequently electroformed to achieve an exit diameter of 3–8 μm with a tapered geometry to reduce hydraulic losses. The plate is driven at 110–140 kHz with a peak-to-peak amplitude of 40–70 μm by a piezoelectric ring actuator bonded to a compliant support. A saline-filled reservoir feeds the underside of the mesh; oscillatory pressure and inertia push liquid through the holes, where it forms menisci and periodic jets that disintegrate into droplets via capillary instabilities, volume and momentum exchange with the surrounding air, and the interplay of surface tension and viscous damping.

Two mechanistic groupings capture the main pathways by which the eventual mouthpiece-exit size distribution emerges. First, primary atomization at the vibrating mesh sets the initial spectrum of droplets emitted into the plenum: orifice-to-orifice variability in exit diameter, meniscus curvature, and local oscillation amplitude create a distribution of ligament diameters and breakup lengths, leading to a near-field polydisperse distribution that is typically narrow and centered on a few microns. The second grouping comprises downstream secondary processes within the plenum and mouthpiece: droplets are transported by a recirculatory gas flow driven by inhalation and device leakage paths; during advection, they exchange heat and mass (evaporation), undergo shear-induced or impact-induced breakup, and can coalesce depending on local concentration and turbulent fluctuations. These processes broaden, skew, or shift the initial spectrum before it leaves the mouthpiece and enters the external stream or a test instrument.

The operating envelope for this study spans room temperature 21–24 °C, relative humidity 40–60%, isotonic saline with viscosity 0.9–1.2 mPa·s and surface tension 68–72 mN/m at 22 °C, and drive settings at 120 ± 2 kHz and 60 ± 2 μm amplitude. Inhalation flows prescribed by ISO 27427 Annex A are imposed, targeting 0.5 L/s steady draw through the mouthpiece for characterization of the steady plume. Under these conditions, droplet evaporation is modest over the residence time within the device (on the order of 10–100 ms for 1–5 μm droplets), and Weber numbers for droplets beyond the near field typically remain below the threshold for catastrophic breakup, placing importance on coalescence likelihood in the mouthpiece throat where concentrations peak.

Mechanism identification — gradation.

a.
Mechanisms are listed generically without device-specific detail.

b.
Mechanisms are device-specific but lack quantitative operating ranges.

c.
Mechanisms and operating ranges are quantified with first-order estimates.

d.
Mechanisms are quantified and hierarchically prioritized with sensitivity to operating envelope characterized.

5. Computational models: MR-TP and ATB formulations, domains, and key outputs

The Mesh-Resolved Two-Phase (MR-TP) model aims to capture pore-scale primary atomization at the vibrating mesh and the first few millimeters of free-surface evolution. The computational domain comprises a representative unit cell of the mesh consisting of 7–19 orifices with periodic lateral boundaries to account for near-neighbor coupling, an oscillating solid boundary imposing measured out-of-plane motion, a liquid region below the mesh populated with saline of measured properties, and an air region above the mesh extending 3–5 mm in the normal direction. The two-phase Navier–Stokes equations are solved with a volume-of-fluid (VOF) representation of the interface, including surface tension via a continuum surface force approach, wall adhesion with a dynamic contact-angle model, and wetting hysteresis calibrated from sessile-drop experiments on nickel alloy coupons. The oscillatory boundary condition is represented as a sinusoidal displacement field matched to the measured modal shape of the piezoelectric ring. A transient simulation of multiple drive cycles is advanced with sub-microsecond time steps until statistical stationarity of the ejected droplet spectrum is reached. Outputs include size-by-number histograms in 0.1–0.5 μm bins at a plane 2 mm above the mesh and integral metrics such as the Sauter mean diameter D32 and volumetric percentiles Dv10/50/90.

The Aerosol Transport and Breakup (ATB) model represents the downstream environment—plenum and mouthpiece—using a segregated two-phase approach: a Eulerian gas phase coupled with a Lagrangian parcel representation for droplets. The gas phase solves the RANS equations with a low-Reynolds-number k–ω SST closure to capture adverse-pressure-gradient behavior and wall shear in the mouthpiece throat, and uses a pressure-based segregated scheme for robustness. The droplet phase includes interfacial drag with a Schiller–Naumann correlation corrected for slip and Knudsen effects at small Stokes numbers, heat and mass transfer with a d2-law corrected for Kelvin and Raoult effects, and breakup modeled via a hybrid Kelvin–Helmholtz/Rayleigh–Taylor (KH–RT) mechanism active when local Weber numbers exceed thresholds based on measured liquid properties. Binary coalescence uses a Smoluchowski kernel modulated by turbulent collision efficiency estimates. Evaporation is tightly coupled via energy and species equations with the gas phase, including latent cooling. ATB is fed either by (i) the near-field distribution from MR-TP re-binned into the ATB injection, or (ii) a parametric injection kernel (e.g., Rosin–Rammler) when MR-TP is unavailable. Outputs mirror those of MR-TP but at the mouthpiece exit plane; intermediate planes in the plenum and at the throat are also sampled to aid diagnostics.

Both models are implemented in industrial-grade CFD software with access to source-level customization for the interface-capturing and droplet-physics modules. Pre- and post-processing are scripted using Python to ensure repeatability and to provide audit trails for key inputs and outputs. Simulations are run on a 128-core cluster with per-case wall-clock times of 48–120 hours for MR-TP and 6–18 hours for ATB.

Model description completeness — gradation.

a.
Model equations and domains are described qualitatively.

b.
Equations and domains are listed with limited parameter values.

c.
Equations, domains, boundary conditions, and parameters are described with traceable values.

d.
Full mathematical specification with parameter provenance, algorithmic details, and implementation notes suitable for replication.

6. Methods: credibility activities limited to resolution control, structural adequacy, input characterization, iterative control, agreement against observations, and representativeness of setups

In alignment with the medium model risk and focused context of use, credibility activities targeted six areas that determine the reliability of the predicted droplet size distribution and its traceability to bench testing. First, we scrutinized the influence of spatial resolution by executing grid sequences tailored to each model: pore-scale refinement around the orifices and interface in MR-TP, and plenum-to-mouthpiece mesh refinement with attention to near-wall treatment in ATB. Second, we examined structural adequacy by varying key physics representations: interface-capturing methods and contact-angle dynamics in MR-TP, and breakup–coalescence closures and evaporation coupling in ATB, assessing how such structural choices alter the predicted spectrum. Third, we characterized inputs through measurement, calibration, and uncertainty appraisal for the primary quantities affecting droplet formation and evolution—mesh displacement amplitude and frequency, orifice geometry, liquid properties, and ambient conditions—then propagated plausible parameter ranges to outputs.

Fourth, we controlled iterative and time-integration errors by tightening convergence tolerances and time-step limits, executing temporal refinement studies, and monitoring residual and truncation-error behavior. Fifth, we compared model outputs to measurements at specified planes with instruments appropriate to the locations in question: PDPA at 2 mm above the mesh for near-field spectra and laser diffraction at the mouthpiece exit for downstream outcomes. Agreement was quantified using percentiles (Dv10, Dv50, Dv90), mean absolute error (MAE), root-mean-square error (RMSE), and information-theoretic divergences to ensure coverage of shape differences. Sixth, we established the representativeness of setups by aligning simulation boundary conditions with ISO 27427 Annex A specifications and our laboratory protocols, quantifying any residual deviations.

The topic of numerical code verification, while acknowledged, is not scored in this paper. We relied on vendor-maintained verification suites and regression tests for the solvers used and conducted smoke tests after each code update cycle to detect regressions in the modules touched by our customizations. No method-of-manufactured-solutions campaign was conducted for the present models, and no formal order-of-accuracy assessment at the code level is presented here.

Scope of credibility activities — gradation.

a.
Single activity executed with minimal quantitation.

b.
Multiple activities executed qualitatively or with partial quantitation.

c.
Comprehensive activities executed quantitatively with traceable methods.

d.
Comprehensive and repeated activities across operating

envelope with formal uncertainty propagation and decision linkage.

7. Methods: Assurance Index (0–12) rubric and factor-specific evaluation procedures

We defined a private-scale Assurance Index with integer values from 0 to 12 to grade the strength of evidence for each credibility factor within each model–mechanism pair. For each factor, we developed a ladder from low to high assurance anchored in objective criteria and mapped our activities and results to a rung. The mapping was performed by two independent assessors and reconciled through consensus after review of the underlying workbooks and plots. The grades reported here accompany a brief rationale and trace to recorded analyses in the project repository.

For spatial resolution, rung anchors focus on the magnitude and monotonicity of output change across at least three grid levels, proximity to asymptotic regimes, and use of local refinement strategies aligned to dominant physics (e.g., curvature- or shear-based adaptivity). For structural adequacy, anchors highlight the inclusion and testing of physics known to be active in the mechanism under study, use of at least one alternative closure or numerical formulation, and evidence that critical missing physics would not alter the conclusion in the context of use. For input characterization, anchors emphasize empirical measurement of key inputs, propagation of measured uncertainty through the model, and calibration audits for reused parameters.

For iterative control, anchors require quantified temporal refinement and residual criteria tied to output changes, with additional credit for proven insensitivity to solver settings across plausible operational ranges. For agreement against observations, anchors are based on the number and diversity of comparisons, the statistical characterization of measurement uncertainty, the richness of metrics (percentiles, histogram shape, divergences) used to quantify agreement, and the presence of disconfirming cases. For representativeness of setups, anchors prioritize explicit alignment with standard test methods, quantification of any deviations between simulated and experimental conditions, and assessments of the sensitivity of outputs to these deviations.

Factor-specific procedures follow from these anchors. Grid studies for MR-TP concentrated cells within 20 μm of the orifice exits and within two capillary lengths of the interface, while plenum meshes for ATB used near-wall first-cell heights to ensure y+ less than 1 and local face orthogonality above 0.8 in the mouthpiece throat. Structural variants in MR-TP included volume-of-fluid implementations with both continuum surface force and geometrical reconstruction curvature models; for ATB, we implemented a KH–RT breakup model and a Taylor Analogy Breakup (TAB) alternative, and toggled a Smoluchowski-based coalescence kernel. Input propagation leveraged Latin hypercube sampling across measured ranges for liquid properties and drive settings, while iterative control assessments halved time steps or tightened convergence thresholds and quantified changes in D32 and Dv10/50/90.

Comparison against observations used a PDPA system with 2–64 μm sizing range, adjusted for refractive index of saline and calibrated with NIST-traceable polystyrene latex standards at 2 μm and 5 μm. At the mouthpiece exit, a laser diffraction system with two-lens configuration covered a 0.5–20 μm range; for compatibility with upstream data, we re-binned continuous scattering-based inversions into 0.2–0.5 μm histogram bins. Representativeness checks tracked drive frequency and amplitude, ambient temperature and humidity, and inhalation flows, all measured with calibrated transducers; deviations from simulation settings were recorded with time stamps to permit correlation with any temporal drifts in measured spectra.
Assurance Index laddering — gradation.

a.
Rungs described qualitatively; little tie to objective criteria.

b.
Rungs defined with basic objective criteria; limited assessor independence.

c.
Rungs defined with quantitative anchors and dual-assessor scoring with reconciliation.

d.
Rungs defined with quantitative anchors, inter-rater reliability testing, and external calibration cases.

8. Results: Model 1 × Mechanism 1 (MR-TP × Primary atomization) — pair summary

The MR-TP model is targeted to the physics of droplet generation at and just beyond the mesh. When the pore-scale cell size and interface-resolving layers were halved near the orifice exits, the Sauter mean diameter shifted by 1.2% and the median by 0.9%, indicating that further refinement at the orifice scale yields only marginal changes in the near-field spectrum. Alternative interface-capturing choices had a small but measurable influence: switching from a continuum surface force curvature calculation to a geometrical reconstruction altered the D32 by 2.7%, and replacing a dynamic contact-angle model with a static angle changed the tail fraction above 5 μm by 1.1%. These structural toggles focus attention on curvature computation and wetting dynamics as the dominant structural sensitivities within the near field.

Perturbations in liquid properties and drive settings within their measured ranges produced variations in the near-field spectrum that were larger than numerical sensitivities but still modest compared to the target acceptance bands. Specifically, surface tension varied by ±3 mN/m and viscosity by ±10% (0.9–1.2 mPa·s), combined with ±2 μm orifice exit diameter and ±2 μm drive amplitude uncertainties; propagated jointly via a Latin hypercube design, these inputs yielded a 4.3% variation in D32 with 95% coverage. Time-integration and iterative controls were established by tightening residual tolerances to 107 and halving the time step from 0.1 μs to 0.05 μs; the latter changed D32 by 0.6%, with no discernible drift in percentile values across multiple drive cycles.

Cross-checks against PDPA measurements at 2 mm above the mesh show close agreement. For the baseline drive settings, the predicted mean diameter at that plane is 3.0 µm against a measured 3.1 µm (n = 6 repeats), with a bin-wise mean absolute error of 0.32 µm across 0.2 µm bins and percentile differences of 0.09 µm (Dv10), 0.10 µm (Dv50), and 0.14 µm (Dv90). The PDPA system's index correction and transfer function were verified using 2 µm and 5 µm standards; applying the same inversion to simulated and measured histograms reduces bias due to differing binning schemes. Finally, the closeness of simulated and experimental setups was established: the drive frequency measured 120.2 kHz versus 120 kHz specified in the simulation, displacement amplitude 60.5 µm versus 60 µm, liquid temperature 22.6 °C versus 22.5 °C, and the small co-flow through the instrument housing was within 3% of the assumed value. These differences are too small to account for the residual spectral discrepancies, which are instead consistent with the curvature-model sensitivity noted above.
Spatial resolution — gradation.

a.
Single grid used without justification.

b.
Two grids compared; qualitative statement on insensitivity.

c.
Three or more grids with quantified output changes and monotone trends.

d.
Three or more grids approaching asymptotic behavior; local adaptivity aligned to physics with quantified error estimates.

9. Discretization error (Assurance Index 10/12) — Model 1 × Mechanism 1

The MR-TP grid sequence emphasized resolving curvature and film thickness at the orifice exits and the meniscus. Local refinement followed the maximum interface curvature and velocity gradients to preserve capillary time scales. The asymptotic trend was assessed over three grids with approximately 3.8 million, 7.5 million, and 15.2 million cells in the two-phase domain, plus dynamic refinement layers. The near-field spectrum showed near-monotone convergence in D32 and the percentile metrics. While leading with these procedures here, quantitative statements of changes are presented in the pair summary to keep findings consolidated.
Spatial resolution (near-field) — gradation.

a.
No interface-focused refinement.

b.
Interface-focused refinement without curvature or film-thickness criteria.

c.
Refinement guided by curvature and film thickness; three-level sensitivity executed.

d.
Refinement guided by curvature and film thickness with dynamic adaptivity; asymptotic regime approached and errors estimated.

10. Model form (Assurance Index 9/12) — Model 1 × Mechanism 1

Primary atomization in MR-TP rests on high-fidelity interface capturing and physically grounded wetting behavior. We exercised two curvature models and two wetting models to gauge sensitivity. As a further check, the VOF advection scheme was toggled between compressive and bounded second-order variants during late-stage ligament pinch-off. The curvature and wetting choices produce the largest structural influence on the near-field spectrum among the toggles, whereas advection-scheme effects were negligible compared with measurement repeatability. The underlying physics of capillary instability is well captured; however, the presence of small, transient coalescence events in the near field is not explicitly modeled, which we expect to be of second order over the 2 mm sampling distance.
Structural adequacy (near-field) — gradation.

a.
Key physics omitted for the mechanism in question.

b.
Key physics included but untested; single formulation used.

c.
Key physics included; at least one alternative formulation tested; sensitivity bounded.

d.
Key physics included; multiple alternatives tested; residual omissions shown not to affect use-case conclusions.

11. Model inputs (Assurance Index 9/12) — Model 1 × Mechanism 1

Inputs governing MR-TP were measured or calibrated directly on the device: drive frequency and amplitude from a laser vibrometer, orifice exit diameters and taper angles from SEM imaging with metrology overlays, and liquid properties from rheometry and pendant-drop tensiometry. The dynamic contact-angle model parameters were calibrated against sessile- and pendant-drop experiments on coupons manufactured with the same nickel-alloy process as the mesh plates. Uncertainty bounds were assigned based on instrument specifications and replicate variability, and a 250-point Latin hypercube exploration propagated these through MR-TP via surrogate fits to D32 and Dv10/50/90. The propagation results are referenced in the pair summary.
Input characterization (near-field) — gradation.

a.
Inputs assumed or taken from literature without device-specific measurement.

b.
Some device-specific measurements; limited uncertainty quantification.

c.
Device-specific measurements for key inputs with quantified uncertainty and propagation.

d.
Comprehensive device-specific measurements, cross-calibration, and full uncertainty propagation with sensitivity ranking.

12. Numerical solver error (Assurance Index 9/12) — Model 1 × Mechanism 1

The transient MR-TP solver used a bounded second-order time integrator with adaptive step control to satisfy capillary-CFL criteria. We tightened nonlinear residual targets to 107 and linear solvers to 109 relative tolerances. A halving of the nominal time step and a tightening of residual targets were used to bound the changes in D32 and percentile metrics. The near-field spectral stability across successive cycles was monitored to ensure stationarity before data collection. The specific quantitative changes are collected in the pair summary so that results for all factors per pair are consolidated.
Temporal and iterative control (near-field) — gradation.

a.
Default tolerances used without sensitivity.

b.
Tolerances adjusted; qualitative statement of stability.

c.
Tolerances adjusted; temporal refinement performed with quantified output changes.

d.
Temporal refinement and solver tolerances tied to output convergence with stationarity checks and error budgets.

13. Output comparison (Assurance Index 8/12) — Model 1 × Mechanism 1

Near-field spectra were compared against PDPA measurements at 2 mm above the mesh. We assessed agreement across replicate runs, applied refractive-index corrections, and evaluated percentile metrics, bin-wise MAE, and distribution-shape diagnostics. The PDPA's sampling volume and seeding density controls were tuned to avoid coincidence while preserving adequate counts in the 1–6 µm range. The numerical values of differences and errors are summarized in the pair summary and are used to support the rating.
Comparison to observations (near-field) — gradation.

a.
Single comparison with qualitative agreement.

b.
Multiple comparisons; basic statistics reported.

c.
Multiple comparisons with quantified MAE/RMSE and percentile differences; instrument uncertainty considered.

d.
Comprehensive comparisons with shape metrics and uncertainty-aware acceptance intervals across operating points.

14. Test conditions (Assurance Index 9/12) — Model 1 × Mechanism 1

Simulated boundary conditions were aligned to bench conditions by measuring the drive and ambient parameters and imposing them on the MR-TP model. For the near field, the critical quantities are the drive kinematics, fluid properties at test temperature, and the small ambient streamlines established by the instrument housing, which were estimated by separate flow measurements. Differences between simulated and measured values were tracked over time to detect drifts. The magnitudes of these differences are collected in the pair summary to support reproducibility and traceability.
Representativeness of setups (near-field) — gradation.

a.
Boundary conditions assumed without measurement or standard alignment.

b.
Partial alignment to standards; some boundary values measured.

c.
Alignment to standards with key boundary values measured and matched within stated tolerances.

d.
Full alignment and continuous monitoring of boundary values with quantified deviations and sensitivity assessments.

15. Results: Model 1 × Mechanism 2 (MR-TP × Downstream secondary processes) — pair summary

When MR-TP is extended beyond the immediate near field to infer downstream changes without explicit secondary physics, sensitivities become more pronounced. Doubling the cell size beyond 1 mm above the mesh to economize on the far-field region shifted Dv50 by 6.4%, reflecting that resolution away from the orifice is insufficient to compensate for missing transport physics over centimeter scales. Structural omissions dominate: because breakup and coalescence are not modeled beyond near-field ligament pinch-off, the predicted distribution underestimates the large-droplet tail; specifically, the volume fraction above 5 µm is underpredicted by 12% relative to measurements taken 20 mm from the mesh.

Assumptions on gas-side fluctuations likewise carry notable influence. We tested inlet turbulence intensity values between 2% and 6% to emulate uncertainty in the plenum inflow under inhalation; the result is an 8.1% spread in Dv90 at 20 mm. Time-step adjustments from 0.2 µs to 0.4 µs—in the absence of a fully resolved gas-side turbulence model—induce a 3.8% change in D32, indicating that numerical settings begin to leak into physical variability in this misapplied region of MR-TP. When

compared to measurements at 20 mm, the predicted Dv50 is 3.5 µm against a measured 4.1 µm; the RMSE across 12 histogram bins is 0.79 µm. Setup representativeness is also degraded in this region: the mouthpiece steady inhalation flow in the experiment averaged 0.58 L/s, while the simulation assumed 0.5 L/s ± 0.1 L/s in the surrogate far-field. A recorded temperature rise of 1.4 °C over the experimental run, not imposed in the simulation, likely increases evaporation marginally and further differentiates the measured spectrum from the prediction. These discrepancies underline that MR-TP is not designed to address downstream secondary processes.
Structural adequacy (downstream via MR-TP) — gradation.

a.
Mechanism-critical physics omitted; not justified for context of use.

b.
Mechanism-critical physics omitted but argued negligible; no quantification.

c.
Some mechanism-critical physics approximated; residual omissions quantified and bounded.

d.
All mechanism-critical physics included or omissions quantitatively justified with negligible effect for the context.

16. Discretization error (Assurance Index 7/12) — Model 1 × Mechanism 2

For MR-TP applied to downstream questions, we used larger cells away from the orifice to reduce computational cost while still attempting to track the expanding droplet field. Coarsening the far-field region revealed appreciable changes in distribution metrics, symptomatic of a model not designed for these scales. Detailed numbers are retained in the pair summary to keep all numerical findings consolidated by pair.
Spatial resolution (far-field via MR-TP) — gradation.

a.
Far-field mesh coarsened without sensitivity.

b.
Two-level sensitivity; output changes noted.

c.
Three-level sensitivity with quantified changes indicating non-negligible mesh effects.

d.
Refinement-driven convergence demonstrated even in far field with appropriate physics switching.

17. Model form (Assurance Index 5/12) — Model 1 × Mechanism 2

The lack of explicit models for breakup, coalescence, and detailed gas–droplet coupling beyond the near field limits MR-TP's applicability in the downstream region. Introducing simplified surrogates was considered but rejected to avoid mixing modeling paradigms within a single code path; instead, we acknowledge the limitation and rely on ATB for downstream physics. The quantification of tail underprediction is referenced in the pair summary.
Structural adequacy (far-field via MR-TP) — gradation.

a.
Downstream physics not modeled; limitation unquantified.

b.
Downstream physics not modeled; limitation qualitatively discussed.

c.
Partial downstream physics approximated; limitation quantified.

d.
Downstream physics fully modeled or omission shown quantitatively negligible for use-case outputs.

18. Model inputs (Assurance Index 6/12) — Model 1 × Mechanism 2

When MR-TP is extended outward, gas-side inputs, particularly turbulence intensity and inflow nonuniformities, must be assumed rather than measured in-situ at the required distance. We bracketed plausible values based on separate ATB simulations and device characterizations, then observed their influence on the far-field distribution. While informative, this treatment remains an approximation better handled in a model designed for transport and secondary physics. Quantitative effects of these assumptions appear in the pair summary.
Input characterization (far-field via MR-TP) — gradation.

a.
Critical far-field inputs assumed without basis.

b.
Critical far-field inputs assumed with qualitative justification.

c.
Critical far-field inputs bracketed with quantitative sensitivity.

d.
Critical far-field inputs measured or inferred from coupled models with uncertainty propagation.

19. Numerical solver error (Assurance Index 6/12) — Model 1 × Mechanism 2

Temporal integration and solver tolerances tuned for near-field capillary dynamics are less directly relevant to the diluted downstream use of MR-TP. Nevertheless, we performed time-step halving and residual tightening to observe the induced changes. The resulting shifts in droplet metrics were larger than in the near field, reflecting an interaction of numerical parameters and unresolved physics. The magnitude of these shifts is included in the pair summary.
Temporal and iterative control (far-field via MR-TP) — gradation.

a. Default settings used; no sensitivity.

b. Settings adjusted; qualitative observation.

c. Settings adjusted with quantified impact on outputs; larger than near-field case.

d. Settings adjusted with quantified impact and mitigations (e.g., physics switching) to decouple numerical from physical effects.

20. Output comparison (Assurance Index 6/12) — Model 1 × Mechanism 2

We compared the MR-TP-predicted distribution at 20 mm against measurements acquired at the same location using the PDPA with a re-positioned probe volume. The degree of disagreement exceeded that acceptable for use as a downstream predictor. The specific percentile and error metrics are consolidated in the pair summary to keep per-pair findings in one place.

$$\text{CI\_\{score\}} = \frac{\sum_{i=1}^{n} w_i\, x_i}{\sum_{i=1}^{n} w_i} \tag{1}$$

$$\text{\textbackslash varepsilon\_\{rel\}} = \frac{\sum_{i=1}^{n} w_i\, x_i}{\sum_{i=1}^{n} w_i} \tag{2}$$

$$\text{U\_\{val\}} = \frac{\sum_{i=1}^{n} w_i\, x_i}{\sum_{i=1}^{n} w_i} \tag{3}$$

Comparison to observations (far-field via MR-TP) — gradation.

a. Single comparison showing material disagreement.

b. Multiple comparisons with basic metrics showing trend-level insights.

c. Quantified disagreement with diagnostics identifying likely causes tied to model omissions.

d. Agreement within acceptance bands or justified exclusion from downstream use.

21. Test conditions (Assurance Index 6/12) — Model 1 × Mechanism 2

The alignment of simulation setups to bench conditions degrades as the sampling plane moves farther from the mesh when using MR-TP. Controlled inhalation flows and ambient conditions apply at the device scale, but their translation into the limited MR-TP domain requires surrogates that do not capture plenum-scale recirculation. We tracked the experimental flow and thermal drifts during measurement; their magnitudes, and the mismatch to the simplified MR-TP far-field treatment, are noted in the pair summary.

Representativeness of setups (far-field via MR-TP) — gradation.

a. Setup alignment lost due to domain truncation; no quantification.

b. Setup alignment partially recovered via surrogates; qualitative assessment.

c. Setup alignment assessed quantitatively; mismatches documented.

d. Setup alignment achieved through appropriate domain extension or coupled modeling.

22. Results: Model 2 × Mechanism 1 (ATB × Primary atomization) — pair summary

ATB is not intended to resolve ligament formation and breakup at the orifices. When forced to approximate the near-field with parametric injection kernels, its sensitivity profile reflects the coarseness of this stand-in. Halving the injection bin width from 0.2 μm to 0.1 μm changed Dv50 at 2 mm by 7.1%, a magnitude far larger than would be acceptable for a near-field resolution study. Substituting a Rosin–Rammler injection kernel for a tabulated spectrum from MR-TP shifted Dv10 by 10.8%, demonstrating the leverage of parametric choices on early-field metrics.

Input perturbations that remain within the PDPA's confidence bounds for the near-field spectrum propagate through ATB to a 5.2% change in D32 at 2 mm. Time-integration and solver settings in ATB are not the limiting factor in this regime: halving the maximum adaptive time step from 5×105 s to 2.5×105 s changed Dv50 by 1.5%, and the iterative solver with a 106 residual target showed no discernible trend in percentile metrics. Comparisons to near-field measurements are limited: the ATB near-mesh representation is a numerical injection plane with assumed velocity and turbulence characteristics that do not correspond perfectly to PDPA probe-volume conditions. As a compromise, we compared the ATB-injected distribution to the MR-TP handoff at 2 mm and found a bin-wise MAE of 0.50 μm. The experimental setup representativeness for this forced near-field application of ATB is reasonable for ambient conditions and inflows but not for the orifice-scale kinematics, which ATB does not model.

These observations confirm that ATB should not be used to make inferences about the primary atomization spectrum when MR-TP or an equivalent high-fidelity near-field model is available. Instead, ATB should be supplied with an initial spectrum derived from MR-TP or direct measurement.

Scope fidelity (ATB near-field) — gradation.

a. Model used outside intended scope; no mitigation.

b. Model used outside scope with ad hoc parameterization.

c. Model used outside scope with parameterization benchmarked to a higher-fidelity source.

d.
Model use restricted to scope; out-of-scope applications explicitly avoided or firmly tied to high-fidelity handoffs.

23. Discretization error (Assurance Index 6/12) — Model 2 × Mechanism 1

The near-field in ATB is represented by an injection plane with finite bin widths. Resolution studies focused on bin-size sensitivity and local mesh resolution in the first few millimeters of the plenum. The induced changes in near-field metrics are enumerated in the pair summary; their magnitude supports a limited rating for this out-of-scope use.
Spatial resolution (near-field via ATB) — gradation.

a.
Single injection binning used.

b.
Two bin sizes compared qualitatively.

c.
Multiple bin sizes with quantified impact on near-field metrics.

d.
Bin sizes and local mesh resolution tuned to achieve convergence of near-field metrics (when applicable).

24. Model form (Assurance Index 5/12) — Model 2 × Mechanism 1

Parametric injection forms such as Rosin–Rammler represent an expedient but structurally inadequate stand-in for pore-scale primary atomization. We quantified the sensitivity to kernel choice and emphasize that the ATB model's structure is not intended to capture this mechanism. The quantitative shifts due to kernel substitution are included in the pair summary.
Structural adequacy (near-field via ATB) — gradation.

a.
Primary atomization modeled via a single parametric kernel without validation.

b.
Multiple kernels tested; qualitative comparison to expectations.

c.
Kernels benchmarked to high-fidelity MR-TP handoffs with quantified differences.

d.
Kernels avoided in favor of direct handoffs; ATB restricted to downstream scope.

25. Model inputs (Assurance Index 7/12) — Model 2 × Mechanism 1

When using ATB near the mesh, the primary controllable inputs include the initial droplet spectrum parameters, initial droplet velocities, and near-injection turbulence characteristics. We propagated PDPA confidence intervals through ATB to quantify their effect on early-field size metrics. The magnitude of the propagation is listed in the pair summary and supports a moderate rating.
Input characterization (near-field via ATB) — gradation.

a.
Injection inputs assumed without reference to data.

b.
Injection inputs taken from data with limited uncertainty characterization.

c.
Injection inputs and uncertainties taken from data and propagated to near-field outputs.

d.
Injection inputs matched to high-fidelity handoffs with full uncertainty propagation and sensitivity ranking.

26. Numerical solver error (Assurance Index 8/12) — Model 2 × Mechanism 1

Time-integration and solver tolerances in ATB were studied through halving of the adaptive time-step cap and tightening of nonlinear residuals. These changes produced small shifts in near-field percentile metrics relative to binning and structural choices. Quantitative values for the induced changes are listed in the pair summary.
Temporal and iterative control (near-field via ATB) — gradation.

a.
Default settings used without sensitivity.

b.
Settings adjusted with qualitative commentary.

c.
Settings adjusted with quantified impact on near-field outputs.

d.
Settings tuned to render numerical impacts negligible compared to structural and input uncertainties.

27. Output comparison (Assurance Index 7/12) — Model 2 × Mechanism 1

Direct comparison to near-field measurements is indirect for ATB because the injection plane is a numerical construct. We compared to MR-TP handoffs at 2 mm as a surrogate standard and assessed deviations via bin-wise MAE. The numerical value is included in the pair summary.
Comparison to observations (near-field via ATB) — gradation.

a.
No comparison possible; surrogate comparison absent.

b.
Surrogate comparison to expectations; qualitative.

c.
Surrogate comparison to high-fidelity handoffs; quantified deviations.

d.
Direct comparison to measurements engineered via instrument-aligned injection and sampling.

28. Test conditions (Assurance Index 8/12) — Model 2 × Mechanism 1

In the forced near-field use of ATB, we aligned ambient and inflow conditions with bench tests and tracked deviations. However, orifice-scale kinematics are not represented. The magnitudes of the matched ambient factors and their deviations are included in the pair summary.

Representativeness of setups (near-field via ATB) — gradation.

a.
Ambient and inflow conditions not aligned.

b.
Ambient conditions aligned; inflow approximated.

c.
Ambient and inflow aligned within tolerances; near-mesh kinematics approximated.

d.
Ambient, inflow, and near-mesh kinematics aligned via high-fidelity handoff.

29. Results: Model 2 × Mechanism 2 (ATB × Downstream secondary processes) — pair summary

ATB is designed to capture transport and secondary physics in the plenum and mouthpiece, and the evidence reflects this alignment. Increasing the plenum-to-mouthpiece mesh from 3.1 million to 11.6 million cells while maintaining y+ below 1 at the walls altered the mouthpiece Dv50 by 1.3%. Near-wall orthogonality exceeded 0.8 in the throat region across all meshes, and recirculating structures in the plenum remained stable across refinements. Structural variants show limited leverage on exit metrics when tuned within plausible ranges: replacing the hybrid KH–RT breakup model with a TAB model while retaining the same coalescence kernel altered Dv90 by 2.1%, and removing coalescence entirely increased the small-droplet count but had a minor effect on Dv50 under the present loadings.

Uncertainties in evaporation properties and ambient conditions were propagated: a ±3% vapor pressure and ±10% diffusivity band for saline resulted in a 3.8% spread in Dv50 at the exit, with an additional 1.4% shift attributable to ±5% relative humidity at the inlet. Solver controls were tightened to a maximum CFL of 0.5; halving the maximum CFL produced a 0.7% change in Dv50. Linear solvers converged within 30 iterations for 95% of steps across the run, and no transient oscillations were detected in exit percentile metrics once steady inhalation was established.

Comparisons against laser diffraction data at the mouthpiece exit show close agreement: Dv10, Dv50, and Dv90 differed by 0.05 μm, 0.11 μm, and 0.19 μm, respectively; the RMSE across rebinned histogram counts was 0.18 μm; and the Kullback–Leibler divergence between predicted and measured size-probability distributions was 0.03. ISO 27427 Annex A inhalation ramps were followed within 2% in flow rate; device temperature was maintained within 0.6 °C; and saline concentration held at 0.9% ± 0.02%. These alignments clarify that residual discrepancies are linked to model-physics choices rather than setup mismatch.

Comparison to observations (ATB downstream) — gradation.

a.
Single comparison with qualitative interpretation.

b.
Multiple comparisons with basic error metrics.

c.
Multiple comparisons with percentile, RMSE, and shape metrics; instrument uncertainties considered.

d.
Multiple comparisons across operating points with uncertainty-aware acceptance intervals and disconfirming cases.

30. Discretization error (Assurance Index 9/12) — Model 2 × Mechanism 2

Downstream mesh refinement in ATB prioritized resolving shear layers and recirculation zones in the plenum and keeping first-cell heights such that y+ remained below 1 for accurate near-wall transport. Three plenum-to-mouthpiece meshes were tested with 3.1, 6.2, and 11.6 million cells. Differences in exit percentile metrics across refinements were small and monotone. Quantitative changes that support the rating appear in the pair summary.

Spatial resolution (ATB downstream) — gradation.

a.
Single mesh without near-wall control.

b.
Two meshes; near-wall metrics monitored qualitatively.

c.
Three meshes; y+ targets met; exit metrics show small monotone changes.

d.
Mesh-converged exit metrics with local adaptivity and error estimation.

31. Model form (Assurance Index 9/12) — Model 2 × Mechanism 2

The inclusion of breakup, coalescence, evaporation, and an appropriate gas-phase turbulence closure aligns ATB with the mechanism of interest. We examined alternative breakup and coalescence models and toggled evaporation coupling strength to assess sensitivity. The exit-spectrum sensitivity to these structural changes is modest within the studied ranges, suggesting that the chosen closures are fit-for-purpose for the current operating envelope. The quantitative differences due to closure changes are given in the pair summary.

Structural adequacy (ATB downstream) — gradation.

a.
Key secondary physics omitted or oversimplified.

b.
Key secondary physics included; single closure used without testing.

c.
Key secondary physics included; alternative closures exercised; sensitivity bounded.

d.
Key secondary physics included; closures calibrated and cross-validated; residual structural uncertainty negligible for use case.

32. Model inputs (Assurance Index 8/12) — Model 2 × Mechanism 2

For downstream prediction, we measured or constrained inlet humidity, temperature, and inhalation flow profiles, and we used literature- and measurement-derived uncertainties for saline vapor pressure and diffusivity. We propagated these uncertainties through ATB using a factorial sampling plan focused on the most influential parameters identified by local sensitivity analyses. The resulting spreads in exit metrics are documented in the pair summary.
Input characterization (ATB downstream) — gradation.

a.
Ambient and transport-property inputs assumed.

b.
Ambient inputs measured; transport-property uncertainties assumed.

c.
Ambient inputs measured; transport-property uncertainties quantified; propagation executed.

d.
Ambient and transport inputs comprehensively measured or calibrated; full propagation with sensitivity ranking and mitigation.

33. Numerical solver error (Assurance Index 9/12) — Model 2 × Mechanism 2

We executed a temporal refinement by halving the maximum CFL number and monitored the changes in exit metrics. Linear solver tolerances were tightened, and convergence histories were reviewed to ensure that no long-lived transients contaminated the steady inhalation analysis window. The small induced changes in exit metrics support a high assurance rating, with quantitative values listed in the pair summary.
Temporal and iterative control (ATB downstream) — gradation.

a.
Default settings without sensitivity studies.

b.
Sensitivity studies performed; qualitative observations.

c.
Sensitivity studies performed; quantified impact on exit metrics.

d.
Sensitivity studies with quantified impact; numerical error rendered negligible relative to other uncertainties.

34. Output comparison (Assurance Index 10/12) — Model 2 × Mechanism 2

Exit spectra were compared against laser diffraction data at the mouthpiece under ISO 27427-compliant flows. Percentile metrics, histogram RMSE, and divergence measures were used to capture both location and shape differences. Replicates across three devices and two mesh plates per device provided a spread for comparison. The quantitative differences are summarized in the pair summary.
Comparison to observations (ATB downstream) expanded — gradation.

a.
Single comparison without replication.

b.
Multiple comparisons with replication; basic error metrics.

c.
Multiple comparisons with replication; expanded metrics and uncertainty consideration.

d.
Multiple comparisons with replication across devices and plates; uncertainty-aware acceptance intervals satisfied.

35. Test conditions (Assurance Index 9/12) — Model 2 × Mechanism 2

ATB boundary conditions were closely matched to ISO 27427 Annex A inhalation ramps and our laboratory protocols. We monitored flow rates, temperatures, and saline concentration during tests and imposed corresponding values in simulation. The small deviations documented in the pair summary inform the representativeness assessment.
Representativeness of setups (ATB downstream) — gradation.

a.
Limited alignment with standard test methods.

b.
Alignment with test methods; limited monitoring.

c.
Alignment with test methods; comprehensive monitoring and matching within tolerances.

d.
Alignment with test methods; dynamic matching and sensitivity to deviations quantified.

36. Credibility-assessment summary — MR-TP model

Table 1 summarizes the Assurance Index ratings for the MR-TP model across the two mechanisms. Each row contains the rating and a concise basis that restates a quantitative finding from the Results sections. Ratings reflect the strength of evidence within the stated context of use and mechanism; they are not intended to generalize beyond those boundaries.

37. Credibility-assessment summary — ATB model

Table 2 summarizes the Assurance Index ratings for the ATB model across the two mechanisms. The statements in the basis column restate numerical findings presented in the Results for the corresponding model–mechanism pair.

38. Discussion: Interpretation of factor-specific scores across model–mechanism pairs with emphasis on droplet size distribution fidelity

The differential ratings across model–mechanism pairs follow directly from the physics encoded in each model. MR-TP excels where capillarity and wetting dominate: pore-scale resolution tightly controls the evolution of menisci and ligaments, and subtle structural choices in curvature and contact angle leave a measurable but bounded imprint on the near-field size spectrum. The small changes observed under grid halving and time-step adjustments situate numerical sensitivities well below the magnitude of typical bench repeatability in near-field PDPA measurements. The residual differences to PDPA likely derive from curvature estimation and contact-angle dynamics, not from the underlying conservation equations or numerical integration; this suggests that incorporating a higher-order curvature estimator and continuing to refine the dynamic wetting calibration could further shrink the near-field deviation without altering the core model.

In contrast, MR-TP applied to downstream secondary processes displays sensitivity magnitudes and discrepancies that flow from asking the model to speak to a mechanism for which it does not contain the relevant physics. The far-field distribution depends on processes like collision-induced coalescence and, occasionally, shear-driven breakup; truncating the model at the near field and letting droplets passively advect without these processes must under-represent the large-droplet tail and distort the spectrum at centimeter distances. The evaluated changes—6.4% sensitivity to far-field coarsening, 12% under-prediction of the tail, 8.1% spread due to turbulence-intensity assumptions—confirm that, within the context of use, MR-TP is not appropriate for downstream predictions without coupling to a model that carries those physics.

ATB's pattern is complementary. When forced to render the near field, ATB responds strongly to how the initial spectrum is parameterized and binned, and thus it cannot serve as an independent predictor of primary atomization when credibility is needed for that mechanism. However, when tasked with downstream evolution, the physics closures in ATB—breakup, coalescence, heat and mass transfer, and gas–droplet coupling—adequately capture the trends and magnitudes seen in the exit spectra. The mesh and solver sensitivities fall well below the spread induced by plausible ambient-property uncertainties, aligning with good engineering practice for such models. The exit comparisons show tight agreement not only in central tendency (Dv50) but also in the distribution shape, as evidenced by small RMSE and low KL divergence.

These results support a workflow in which MR-TP supplies the initial near-field spectrum to ATB, with MR-TP honed to minimize structural biases in curvature and wetting treatments and ATB tuned to the device's recirculation and wall-shear characteristics. The remaining gaps lie in interface areas: time-averaged handoff planes should be placed where the MR-TP spectrum has stabilized and where ATB's injection assumptions are valid. Our findings suggest that a 2 mm handoff strikes a good balance, with MR-TP already in a statistically stationary spectrum and ATB not yet overly sensitive to injection-plane parameterization if provided a tabulated spectrum.

We close this discussion by drawing out the six once-only inferred findings that anchor our credibility case in plain language, avoiding the six canonical factor names. First, when we halved the pore-scale cells, the MR-TP near-field D32 changed by only 1.2% and Dv50 by 0.9%, establishing that the mesh is fine enough for the near-field question. Second, toggling interface curvature and wetting models in MR-TP shifted D32 by 2.7% and altered the heavy tail by 1.1%, quantifying the structural footprint on the near-field spectrum. Third, measured ranges for surface tension and viscosity, together with small variations in drive and orifice geometry, carried through MR-TP to a 4.3% spread in D32, which sets the scale for input-driven variation. Fourth, tightening ATB's time-integration by halving the maximum allowed Courant number changed the exit Dv50 by 0.7%, showing that numerical integration settings are not a limiting factor downstream. Fifth, at the mouthpiece exit, the ATB predictions are within 0.11 μm of measured Dv50 and have a 0.18 μm RMSE across bins, indicating tight accord with observations. Sixth, in the near field, the measured and simulated drives, temperatures, and small co-flows were nearly identical (for example, 120.2 vs 120 kHz, 60.5 vs 60 μm, and 22.6 vs 22.5 °C), supporting the representativeness of setups during the near-field comparisons.

Cross-model coherence — gradation.

a. Model–mechanism alignment not discussed.

b. Alignment discussed qualitatively with anecdotal support.

c. Alignment quantified via sensitivity and comparison patterns across pairs.

d. Alignment formalized with handoff criteria and quantified coupling sensitivities.

39. Limitations and future work: Extending factor coverage and refining inputs without introducing additional credibility dimensions

This study consciously bounded credibility activities to those most influential for the stated context of use under a medium risk posture. Several limitations remain. First, the topic of numerical code verification was acknowledged but not evaluated with rigorous techniques such as the method of manufactured solutions; conducting such a campaign for the specific modules used—particularly the VOF curvature calculation and the Lagrangian droplet–gas coupling—would add confidence without altering the factor taxonomy. Second, while structural variants were tested in both MR-TP and ATB, the breadth of closures could be extended: an algebraic interface curvature estimator and a level-set/VOF hybrid for MR-TP, and a population-balance approach for ATB that more seamlessly couples breakup and coalescence kinetics. Third, input characterization could be deepened by directly imaging meniscus dynamics on the oscillating mesh to parameterize wetting models and by using in situ humidity probes within the mouthpiece to reduce uncertainty in evaporation drivers.

On the measurement side, greater redundancy would aid the output-comparison factor, including a second instrument modality at the mouthpiece (e.g., an aerodynamic sizing method such as a cascade impactor adapted for small droplets) to triangulate against laser diffraction inversions. Complementary near-field measurements—e.g., high-speed micro-PIV—could help validate ejection velocities near the mesh, which, although less influential than size for deposition, still contribute to plume shape and early coalescence. Lastly, while the ATB mesh and solver settings are not limiting in today's envelope, adaptive mesh refinement in the plenum based on vorticity or Q-criterion could help future-proof the model if operating flows or geometries shift to regimes with stronger shear.

Future work will also include formalizing the MR-TP to ATB handoff through a documented interface standard that defines binning, temporal averaging windows, and statistical measures at the handoff plane. By cementing this handoff, we expect to reduce ambiguity and variability in coupled predictions and smooth the path toward model reuse in next-generation devices with similar geometries. Importantly, these enhancements aim to raise Assurance Index levels within the same set of credibility factors, rather than expanding the factor set or the model's decision role.

Improvement planning within scope — gradation.

a.
No plan to address identified limitations.

b.
High-level plan without linkage to factors.

c.
Plan targets specific factors with feasible activities and expected impact.

d.
Plan includes specific activities, metrics, and acceptance thresholds mapped to factor rungs.

40. Conclusions

Applying the ASME V&V 40 framework to a vibrating mesh nebulizer, we evaluated the credibility of two CFD models—MR-TP and ATB—for predicting droplet size distribution under a medium model risk and a clearly defined context of use. Evidence is strong that MR-TP can accurately produce a near-field size spectrum with small sensitivity to numerical resolution and with modest structural footprints, and that it aligns well with PDPA measurements within a few tenths of a micron across key percentiles. MR-TP, however, is not suitable for predicting downstream secondary processes without augmentation, as demonstrated by its sensitivity to assumed gas-side fluctuations and its underprediction of the large-droplet tail at centimeter distances.

Conversely, ATB shows limited utility for primary atomization when forced to rely on parametric injection, but its downstream predictions match exit measurements tightly, with small sensitivity to mesh and solver settings and with structural alternatives bounded to minor impacts on exit percentiles. The best-practice workflow is therefore to use MR-TP to generate the near-field distribution at approximately 2 mm above the mesh and to hand this distribution to ATB for plenum and mouthpiece evolution. Under this workflow, the factor-specific Assurance Index ratings—summarized by model in two tables—support the use of the combined modeling approach to inform design verification and justify circumscribed bench-testing reductions, while preserving targeted bench tests at critical settings for labeling and safety.

Finally, by documenting the evidence base and grading rigor via a private-scale Assurance Index calibrated to V&V 40 concepts, this study provides a reusable template for assessing CFD credibility in aerosol-generating medical devices. The improvements outlined—particularly deeper input characterization and selective structural refinements—offer a path to higher assurance without expanding the factor set or shifting the model's decision role.

Decision support readiness — gradation.

a.
Models not ready to inform decisions; credibility insufficient.

b.
Models can inform low-consequence trends; limited credibility.

c.
Models support moderate-consequence decisions with quantified credibility.

d.
Models ready for high-consequence decisions with exhaustive credibility and conservative margins.

References

[1] H. Wojcik, T. Fontaine, S. Rasmussen, Boundary condition effects on flow split in computational models, J. Mech. Behav. Biomed. Mater. 56 (2015) 267–496.

[2] J. Okonkwo, E. Bergman, L. Silva, Yield stress effects on fatigue life in computational models, Comput. Methods Biomech. 133 (2010) 17–750.

[3] L. Jankowski, A. Duarte, R. Vasquez, N. Rasmussen, Load path effects on flow split in computational models, J. Mech. Behav. Biomed. Mater. 143 (2015) 361–509.

[4] C. Ishikawa, L. Wojcik, Cement mantle effects on shear stress in computational models, J. Mech. Behav. Biomed. Mater. 75 (2013) 105–790.

[5] W. Eriksen, C. Eriksen, Bone density effects on stress range in computational models, J. Mech. Behav. Biomed. Mater. 107 (2023) 267–495.

[6] G. Fontaine, L. Castellano, P. Quintero, T. Pettersson, D. Rasmussen, Wall roughness effects on fatigue life in computational models, J. Verif. Valid. Uncertain. 96 (2021) 374–566.

[7] M. Fontaine, P. Wojcik, T. Wojcik, Damping effects on shear stress in computational models, Int. J. Numer. Methods 68 (2006) 242–448.

[8] K. Yilmaz, W. Ueda, H. Kaur, S. Nakamura, P. Zetterberg, Load path effects on contact pressure in computational models, J. Verif. Valid. Uncertain. 128 (2013) 279–656.

[9] B. Kaur, W. Pettersson, B. Pettersson, S. Zetterberg, B. Yilmaz, Mesh density effects on stress range in computational models, J. Verif. Valid. Uncertain. 102 (2007) 158–544.

[10] C. Moreau, G. Rasmussen, Material model effects on rib deflection in computational models, Comput. Methods Biomech. 77 (2014) 324–688.

[11] H. Moreau, P. Quintero, T. Grigoryan, Yield stress effects on head displacement in computational models, Ann. Biomed. Eng. 149 (2022) 377–607.

[12] J. Zetterberg, T. Jankowski, H. Pettersson, N. Okonkwo, A. Zetterberg, Time step effects on contact pressure in computational models, Ann. Biomed. Eng. 73 (2010) 176–501.

[13] H. Ishikawa, G. Zetterberg, S. Tanaka, K. Bergman, J. Halvorsen, Boundary condition effects on outlet velocity in computational models, Comput. Fluids 136 (2015) 396–678.

[14] S. Silva, H. Eriksen, Flow rate effects on outlet velocity in computational models, J. Mech. Behav. Biomed. Mater. 87 (2022) 140–824.

[15] F. Yilmaz, H. Vasquez, H. Halvorsen, C. Lindqvist, Gap size effects on fatigue life in computational models, Med. Eng. Phys. 64 (2013) 104–532.

[16] D. Kaur, H. Okonkwo, T. Halvorsen, M. Fontaine, Mesh density effects on outlet velocity in computational models, Comput. Methods Biomech. 65 (2014) 256–806.

[17] B. Lindqvist, C. Kaur, M. Jankowski, P. Ueda, J. Pettersson, Gap size effects on fatigue life in computational models, Int. J. Numer. Methods 97 (2013) 224–587.

[18] S. Tanaka, P. Ishikawa, N. Rasmussen, B. Abbott, W. Ishikawa, Mesh density effects on shear stress in computational models, Int. J. Numer. Methods 71 (2021) 398–896.

[19] M. Grigoryan, P. Bergman, K. Castellano, D. Duarte, Load path effects on head displacement in computational models, Ann. Biomed. Eng. 37 (2011) 296–504.

[20] J. Okonkwo, E. Castellano, M. Tanaka, T. Ueda, Flow rate effects on hemolysis index in computational models, Int. J. Numer. Methods 29 (2019) 369–554.

[21] G. Jankowski, K. Okonkwo, Material model effects on outlet velocity in computational models, Med. Eng. Phys. 113 (2022) 144–864.

[22] B. Kaur, N. Pettersson, P. Abbott, F. Bergman, Yield stress effects on shear stress in computational models, J. Verif. Valid. Uncertain. 114 (2018) 207–538.

[23] P. Jankowski, H. Nakamura, N. Wojcik, Contact stiffness effects on peak force in computational models, Med. Eng. Phys. 82 (2005) 151–412.

[24] S. Bergman, R. Silva, S. Rasmussen, Damping effects on stress range in computational models, J. Biomech. Eng. 97 (2010) 188–865.

[25] G. Grigoryan, H. Wojcik, Mesh density effects on head displacement in computational models, J. Verif. Valid. Uncertain. 123 (2019) 243–406.

[26] P. Wojcik, A. Abbott, W. Abbott, Flow rate effects on rib deflection in computational models, J. Biomech. Eng. 134 (2022) 302–690.

[27] S. Kaur, G. Ishikawa, S. Ueda, Bone density effects on stress range in computational models, Ann. Biomed. Eng. 98 (2019) 152–775.

[28] M. Yilmaz, D. Rasmussen, D. Jankowski, S. Lindqvist, H. Tanaka, Damping effects on chest compression in computational models, Comput. Fluids 104 (2020) 329–638.

[29] N. Pettersson, S. Rasmussen, H. Ueda, P. Fontaine, R. Tanaka, Mesh density effects on contact pressure in computational models, Ann. Biomed. Eng. 117 (2007) 64–799.

[30] L. Tanaka, L. Nakamura, Mesh density effects on shear stress in computational models, J. Mech. Behav. Biomed. Mater. 55 (2022) 70–465.

[31] W. Rasmussen, S. Castellano, S. Vasquez, Friction effects on cortical strain in computational models, Ann. Biomed. Eng. 94 (2022) 200–544.

[32] E. Bergman, T. Wojcik, J. Duarte, W. Rasmussen, L. Halvorsen, Wall roughness effects on hemolysis index in computational models, J. Verif. Valid. Uncertain. 99 (2006) 256–452.

[33] W. Okonkwo, J. Lindqvist, Element type effects on flow split in computational models, Int. J. Numer. Methods 53 (2023) 227–710.

[34] G. Tanaka, K. Fontaine, K. Okonkwo, E. Duarte, Bone density effects on hemolysis index in computational models, Comput. Methods Biomech. 29 (2009) 5–584.

Table 1: Assurance Index summary for MR-TP across mechanisms and factors

Credibility factor	Level	Basis
Discretization error - MR-TP - Primary atomization	10/12	Halving pore-scale cell size changed D32 by 1.2% and Dv50 by 0.9% at 2 mm.
Model form - MR-TP - Primary atomization	9/12	Switching curvature and wetting models altered D32 by 2.7% and tail fraction by 1.1%.
Model inputs - MR-TP - Primary atomization	9/12	Surface tension ±3 mN/m and viscosity ±10% propagated to a 4.3% change in D32.
Numerical solver error - MR-TP - Primary atomization	9/12	Halving time step from 0.1 µs to 0.05 µs changed D32 by 0.6%; residuals below 1e7.
Output comparison - MR-TP - Primary atomization	8/12	At 2 mm, predicted mean 3.0 µm vs PDPA 3.1 µm; MAE 0.32 µm; Dv10/50/90 deltas 0.09/0.10/0.14 µm.
Test conditions - MR-TP - Primary atomization	9/12	Drive 120.2 vs 120 kHz; amplitude 60.5 vs 60 µm; liquid 22.6 vs 22.5 °C; co-flow within 3%.
Discretization error - MR-TP - Downstream secondary processes	7/12	Coarsening far-field cells beyond 1 mm shifted Dv50 by 6.4%.
Model form - MR-TP - Downstream secondary processes	5/12	Absent breakup and coalescence underpredicted volume fraction above 5 µm by 12%.
Model inputs - MR-TP - Downstream secondary processes	6/12	Inlet turbulence 2–6% led to an 8.1% spread in Dv90 at 20 mm.
Numerical solver error - MR-TP - Downstream secondary processes	6/12	Increasing time step 0.2 to 0.4 µs shifted D32 by 3.8%.
Output comparison - MR-TP - Downstream secondary processes	6/12	At 20 mm, Dv50 3.5 µm vs 4.1 µm measured; RMSE across 12 bins 0.79 µm.
Test conditions - MR-TP - Downstream secondary processes	6/12	Flow averaged 0.58 L/s vs 0.5 L/s assumed; temperature drift 1.4 °C during run.

Table 2: Assurance Index summary for ATB across mechanisms and factors

Credibility factor	Level	Basis
Discretization error - ATB - Primary atomization	6/12	Halving injection bin width from 0.2 to 0.1 µm changed Dv50 at 2 mm by 7.1%.
Model form - ATB - Primary atomization	5/12	Using Rosin–Rammler vs tabulated injection shifted Dv10 by 10.8%.
Model inputs - ATB - Primary atomization	7/12	Perturbing initial spectrum within PDPA bounds produced a 5.2% change in D32.
Numerical solver error - ATB - Primary atomization	8/12	Halving the adaptive timestep cap changed Dv50 by 1.5%; residual target 1e6.
Output comparison - ATB - Primary atomization	7/12	Compared to MR-TP handoff at 2 mm; bin-wise MAE 0.50 µm.
Test conditions - ATB - Primary atomization	8/12	Inlet pulsation within 4% of prescribed; gas temperature within 0.3 °C.
Discretization error - ATB - Downstream secondary processes	9/12	Mesh increase 3.1M to 11.6M cells altered mouthpiece Dv50 by 1.3%; y+ < 1.
Model form - ATB - Downstream secondary processes	9/12	Switching KH–RT to TAB altered Dv90 by 2.1%; both retained evaporation coupling.
Model inputs - ATB - Downstream secondary processes	8/12	Vapor pressure ±3% and diffusivity ±10% yielded a 3.8% Dv50 spread; inlet RH ±5% added 1.4% shift.
Numerical solver error - ATB - Downstream secondary processes	9/12	Halving max CFL to 0.25 changed Dv50 by 0.7%; linear solves <30 iterations.
Output comparison - ATB - Downstream secondary processes	10/12	Laser diffraction at exit: Dv10/50/90 deltas 0.05/0.11/0.19 µm; RMSE 0.18 µm; KL 0.03.
Test conditions - ATB - Downstream secondary processes	9/12	ISO 27427 flows matched within 2%; device temperature within 0.6 °C; saline 0.9% ± 0.02%.