Formative User Study & System Design Requirements

FZ-VIS: A Visual Analytics Framework for Quantities-of-Interest-Aware Scientific Lossy Compression.

Part I: Formative User Study

Demographics of the Participants

Our formative investigation involved semi-structured interviews/discussions with 10 targeted participants spanning three core user groups:

ID	User Group Role	Current Role	Experience	Primary Scientific Domain
P1	Novice User (U)	Master Student	6 Months	Data Visualization
P2	Novice User (U)	Fresh PhD Student	1 Year	Geological Oceanography
P3	Novice User (U)	Senior PhD Student	3 Years	Environmental Science
P4	Compressor Developer (D)	Research Scientist	8 Years	High-Performance Computing (HPC)
P5	Compressor Developer (D)	Senior Research Scientist	12 Years	Scientific Data Compression
P6	Compressor Developer (D)	Assistant Professor	5 Years	Parallel Algorithms & Solvers
P7	QoI Domain Scientist (S)	Principal Investigator	8 Years	Combustion & Turbulent Flows
P8	QoI Domain Scientist (S)	Senior PhD Student	3 Years	Topological Data Analysis
P9	QoI Domain Scientist (S)	Associate Professor	10 Years	Materials Science
P10	QoI Domain Scientist (S)	Staff Scientist	6 Years	Cosmological Simulation

General Interview Questionnaire

The following questions were presented to all 10 participants to help establish system design requirements:

Q1. Baseline & Dataset Context: What specific scientific datasets do you generate / work with, and how does lossy compression currently fit into your downstream analytics pipeline?

Q2. Composition & Setup Challenges: How do you configure and assemble pipeline modules, and what are your main bottlenecks in setting up your compression workflow?

Q3. Parameter Space Navigation: What strategies do you use to navigate parameter spaces, and how do you identify compressor configurations that fit your need?

Q4. Visualization Tool Usage: What visualization software, tools, or libraries do you often use, and which specific features are essential to your compression workflow?

Q5. QoI & Feature Preservation: Which domain-specific features or Quantities-of-Interest (QoI) need to be preserved for downstream analysis during the lossy compression?

Q6. Telemetry & Visual Diagnostics: What intermediate views, error maps, or comparison plots would assist you most in identifying optimal compression configurations?

Novice Users: Visual Needs

User Profile (P1, P2, P3)

Typically graduate students, trainees, or scientific domain engineers who treat compressors as a black box. They understand the overall goals of data reduction but lack knowledge of internal data transform, prediction, or quantization stages.

Key Research Inquiries

"What visualization features will be helpful?"

Interactive compressor composition: Can we have a visual interface to compose different compressors?

Interpreting error bounds: How does changing the tolerance directly affect the spatial reconstruction error?

Side-by-side comparisons: Can we see a comparative preview of 2 or 3 settings at once?

RD curves: How can we customize the axes of a Rate-Distortion curve to show different plots?

Compressor Developers: Diagnostic Needs

</> User Profile (P4, P5, P6)

Computer science specialists focused on designing, tuning, and benchmarking compression algorithms. They construct pipelines containing predictors (e.g., Lorenzo, Interpolation), quantizers, and entropy encoders to improve the performance of lossy compressors.

? Key Research Inquiries

"What diagnostic views will be helpful?"

Stage isolation: Which stages in the pipeline will be affected by the prediction distortion?

Quantization histograms: Can we inspect the density distribution of unpredicted bin values?

Residual profiling: Can we visualize high-dimensional residuals relative to spatial step constraints?

QoI-Driven Scientists: Fidelity Needs

User Profile (P7, P8, P9, P10)

Domain experts who care about preserving specific Quantities of Interest (QoIs) that are essential for downstream analysis. They focus primarily on scientific features besides pointwise errors, e.g., critical points, merge trees, and Morse-Smale (MS) segmentation / complexes in the topological domain, or power spectra in the frequency domain.

Key Research Inquiries

"What scientific plots will be helpful?"

Topological features: Can we visualize and compare topological descriptors (e.g., critical points, merge trees, or MS complexes) in the decompressed data?

Spectral errors: Can we generate the plot to visualize the power spectrum and spectral errors after lossy compression?

Preservation: How can we assist users in finding the optimal settings for feature preservation and correction when using lossy compression?

Part II: Co-Designed Requirements

R1 & R2: Workflows & Parameters

R1: Interactively Composing Workflows

Expose available modules and parameters of lossy compressors. All user groups need an intuitive graphical interface to construct compression pipelines, modify configurations on the fly, and inspect results without writing or building code.

R2: Exploring Parameter Spaces

A key requirement for novice users is to learn how parameters interact. Rather than forcing blind adjustment, the system must help identify optimal candidates, highlighting spatial errors, and supporting iterative refinement.

R3 & R4: Debugging & Profiling

R3: Inspecting Intermediate Results

Compressor developers express the need to separate different compression stages, and the system needs to expose stage-wise predictions, quantization errors, residuals, and statistics, allowing users to diagnose modular design performance.

R4: Profiling through Linked Views

Link high-level analytical metrics (PSNR, SSIM, compression ratio) to direct visual side-by-side reconstruction comparisons. Help users move from abstract summaries to region-specific inspection.

R5: QoI Analysis and Preservation

R5

Scientific Fidelity

Domain-Specific Scientific Features

Pointwise error indices (like RMSE or L2-norm) do not guarantee that derived quantities remain scientifically valid. Scientists often need dedicated tools to compute quality metrics, visualize topological descriptors, and plot frequency power spectra.

Provide a unified environment for researchers to extract critical domain-specific features and evaluate their fidelity under lossy compression, while guiding the preservation / correction of QoIs through optimized compressor settings.

Development Cycle

Phase 1

Collaborative Discussions

Conducted structured discussions with 10 participants to understand baseline workflows and gather their essential needs for a visualization system.

Phase 2

Requirements Design

Synthesized five primary technical requirements and iteratively refined them in collaboration with domain researchers.

Phase 3

System Prototyping

Developed interactive workflow composers, linked views, and QoI analysis modules.

Phase 4

Case Validations

Reviewed core design choices across diverse practical case studies provided by target user groups.