Funded by
Federal Ministry for the Environment, Climate Action, Nature Conservation and Nuclear Safety
02E12112A

GeoPathSim

Research software documentation

Authors Duckstein[1], A.; Pospiech[1], S.; Tolosana-Delgado[2], R.

Affiliations [1]Helmholtz-Zentrum Dresden-Rossendorf, Institute for Resource Ecology (IRE), Dresden, Germany
[2]Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology (HIF), Freiberg, Germany

Software version 1.0.0

Release date Aug 2026

HZDR-RODARE release documentation

Contents

1 Introduction

GeoPathSim is a Python research-software workflow for deriving and stochastically simulating mineral structures exposed along potential transport pathways in heterogeneous geological materials.

Bulk mineral composition alone does not necessarily represent the accessible mineral assemblage, since mineral reactions relevant to solute and radionuclide transport are controlled by the phases that are actually exposed to connected pore space, fractures, and other potential transport pathways.

GeoPathSim therefore provides two complementary approaches for describing and simulating mineral exposure:

1. **1D Markov simulation** for visible cracks, fractures, or other explicit pathways can be traced in the input data. The workflow extracts mineral sequences along selected contours, estimates transition probabilities between successive mineral classes, and simulates additional one-dimensional mineral chains.
2. **2D multi-Gaussian simulation** when no distinct crack pathway is visible, when pathways cannot be traced reliably, or when complete spatial mineral fields are required. Indicator variograms are estimated from the classified mineral structure, Matérn models are fitted, and stochastic two-dimensional mineral realizations are generated with prescribed mineral proportions.

GeoPathSim starts from spatially resolved classified mineral rasters. It supports preprocessing, mineral-neighbour analysis, stochastic simulation, result aggregation, plotting, and optional post-processing add-ons.

GeoPathSim represents mineral structure and exposure statistics. It does not itself calculate reactive transport, prove hydraulic connectivity, or imply that every simulated pathway is physically connected. These interpretations require the surrounding geological, hydrological, and geochemical model.

This document contains the complete project-level documentation. An overview of the modular software workflow and all executable entry points is provided in Appendix A.

1.1 Release metadata

Field	Information
Software title	GeoPathSim
Software version	1.0.0
Release date	Aug 2026
Development status	Research software
Repository	HZDR-RODARE
DOI	10.14278/rodare.4933
License	Creative Commons Attribution 4.0 International

1.2 Typical workflows

Data situation	Recommended workflow
A visible crack or pathway can be traced in the classified image	Data preprocessing + 1D Markov simulation
No distinct crack is visible or a complete 2D field is needed	Data preprocessing + 2D multi-Gaussian simulation
Only standardized rasters, compositions, or neighbour statistics are needed	Data preprocessing only

The 1D and 2D routes are complementary. They are not intended to produce identical representations of the same structure. The relationship between preprocessing, both simulation routes, combined result plotting, and the optional add-ons is summarized in Appendix A.

1.3 Main features

- Conversion of classified mineral and mask images into compact integer rasters.
- Calculation of bulk and mask-area volume and weight fractions.
- Optional neighbour counting for the original two-dimensional training data.
- Separate treatment of configured resplit mineral classes during neighbour counting.
- Extraction and assembly of one-dimensional mineral pathways from visible structures.
- Estimation and diagnostic checking of Markov transition matrices.
- Reproducible simulation of one-dimensional mineral chains with optional TXT or NPZ export of every realization.
- Configurable boxplot center, whisker definition, and flier display for 1D and 2D result summaries.
- Balanced indicator sampling and empirical variogram estimation.
- Fitting of two-dimensional Matérn variogram models.
- Hierarchical multi-Gaussian simulation with exact integer target-pixel counts.
- Optional resplitting of configured joint-mineral classes.
- Combined weighted box plots for the mineralogical composition, summarizing the results of 1D Markov and 2D multi-Gaussian models across any number of samples.
- Export of effective configurations, manifests, numerical results, and plots.

1.4 Project structure

```
GeoPathSim/
├── data_preprocessing/
│   ├── data_output/
│   ├── rsc/
│   ├── 01_run_data_preprocessing.py
│   ├── config_preprocessing.json
│   └── USER_GUIDE_preprocessing.md
├── 1D_markov_simulation/
│   ├── data_output/
│   ├── rsc/
│   │   ├── markov_preprocessing/
│   │   └── markov_chain_simulation/
│   ├── 01_run_markov_preprocessing.py
│   ├── 02_run_markov_chain_simulation.py
│   ├── config_markov_01.json
│   ├── config_markov_02.json
│   └── USER_GUIDE_markov_simulation.md
├── 2D_multigaussian_simulation/
│   ├── data_output/
│   ├── rsc/
│   ├── 01_run_variogram_analysis.py
│   ├── 02_run_field_simulation.py
│   ├── 03_run_result_plotting.py
│   ├── config_multigaussian_01.json
│   ├── config_multigaussian_02.json
│   ├── config_multigaussian_03.json
│   └── USER_GUIDE_multigaussian_simulation.md
├── combined_result_plotting/
│   ├── data_output/
│   ├── rsc/
│   ├── 01_run_combined_result_plotting.py
│   ├── config_combined_results.json
│   └── USER_GUIDE_combined_result_plotting.md
├── add_ons/
│   ├── data_output/
│   ├── rsc/
│   ├── 01_run_highlight_minerals.py
│   ├── 02_run_bulk_vs_pathway.py
│   ├── 03_run_smoothing.py
│   ├── 04_run_crop_images.py
│   ├── 05_run_rve_size_analysis.py
│   ├── config_highlight_minerals.json
│   ├── config_...
│   └── USER_GUIDE_add_ons.md
├── data_input/
├── GEOPATHSIM_DOCUMENTATION.md
└── README.txt
```

2 Installation and software environment

GeoPathSim requires **Python 3.10 or newer**, while the archived environment records **Python 3.13.5**.

The software does not require access to high-performance computing (HPC) infrastructure and is designed to run on a conventional desktop computer or laptop. The reference calculations and simulations reported for the archived release were performed on a system with an Intel Core i5-1135G7 processor, 4 CPU cores / 8 threads, 16 GB RAM, running Windows 10 Enterprise LTSC. No dedicated GPU is required. Computational time and memory requirements depend primarily on the dimensions of the input rasters, the number and size of simulated realizations, and the selected analysis settings.

The source code imports the following external packages, listed, including their areas of application and the package version of the recorded release environment:

Functional group	Package	Version
Numerical and scientific computing	`numpy`	2.3.5
Numerical and scientific computing	`scipy`	1.17.0
Numerical and scientific computing	`pandas`	2.2.3
Plotting and image output	`matplotlib`	3.10.8
Plotting and image output	`seaborn`	0.13.2
Plotting and image output	`Pillow`	12.2.0
Plotting and image output	`matplotlib-scalebar`	0.9.0
Image and raster processing	`opencv-python`	4.13.0.92
Image and raster processing	`scikit-image`	0.26.0
Geostatistical modelling and simulation	`gstools`	1.7.0

Create an isolated virtual environment before installing the dependencies. For example:

```
python -m venv .venv

# Linux/macOS
source .venv/bin/activate

# Windows PowerShell
.venv\Scripts\Activate.ps1
```

Install the exact release dependencies from the provided `requirements.txt`:

```
python -m pip install --upgrade pip
python -m pip install -r requirements.txt
```

3 Quick start

3.1 Prepare the input data

Place the external input files in:

```
data_input/
```

Adapt the paths and processing settings in:

```
data_preprocessing/config_preprocessing.json
```

The included configuration files are templates and must be checked against the local filenames, class definitions, and intended output settings.

3.2 Run data preprocessing

From the GeoPathSim project root:

```
python data_preprocessing/01_run_data_preprocessing.py
```

Review the generated manifest and output files before continuing.

3.3 Run the 1D Markov workflow

Adapt:

```
1D_markov_simulation/config_markov_01.json
1D_markov_simulation/config_markov_02.json
```

Then run:

```
python 1D_markov_simulation/01_run_markov_preprocessing.py
python 1D_markov_simulation/02_run_markov_chain_simulation.py
```

The corresponding result directories are named consistently with the multi-Gaussian workflow:

```
step01_<timestamp>_<run_name>/
step02_<timestamp>_<run_name>/
```

3.4 Run the 2D multi-Gaussian workflow

Adapt:

```
2D_multigaussian_simulation/config_multigaussian_01.json
2D_multigaussian_simulation/config_multigaussian_02.json
2D_multigaussian_simulation/config_multigaussian_03.json
```

Then run the required steps in numerical order:

```
python 2D_multigaussian_simulation/01_run_variogram_analysis.py
python 2D_multigaussian_simulation/02_run_field_simulation.py
python 2D_multigaussian_simulation/03_run_result_plotting.py
```

Step 03 is used for result aggregation and plotting and may be omitted when these outputs are not required.

3.5 Combine any number of simulation results

After saving all Markov chains and the multi-Gaussian Step 02 neighbour matrices, adapt:

```
combined_result_plotting/config_combined_results.json
```

Then run:

```
python combined_result_plotting/01_run_combined_result_plotting.py
```

The combined workflow accepts arbitrary numbers of result directories for both methods and supports weighting by simulation count, equal weighting per sample, or custom sample weights. It converts the mineral composition from the Markov simulations and the mineral compositions along the grain boundaries from the Multi-Gauss simulations into relative frequencies for each simulation, and aggregates these into a single distribution for each mineral. The plotted statistics are exported both as TXT and as `plot_data_summary.csv`. Optional mineral groups can be defined by listing the source mineral names and a higher-level output name. Grouping is applied before frequencies are calculated.

4 Workflow details

4.1 Data preprocessing

The preprocessing workflow converts classified mineral and mask images into integer rasters, calculates bulk and optional mask-area compositions, and writes metadata and a manifest. Source classes may share the same mineral name. In that case, GeoPathSim maps all corresponding colors and indices to the smallest source index carrying that name and uses the representative color and specific gravity from this index. The output therefore contains one retained name and index per mineral, together with traceability tables documenting the original-to-canonical mapping. The workflow can also save a standalone mineral-color legend under `metadata/mineral_legend.<format>`. This legend contains only the canonical mineral classes that actually occur in at least one processed sample; unused metadata entries are omitted. One or more file formats and a separate raster resolution can be configured. Optional neighbour counting can be enabled for the original training data. Mineral pairs configured for resplitting are retained as separate classes during neighbour counting, while joint pairs not marked for resplitting remain merged according to the configured joint-mineral definition. The associated program flow is shown in Appendix B.

Detailed formats and configuration settings are documented in:

```
data_preprocessing/USER_GUIDE_preprocessing.md
```

4.2 One-dimensional Markov simulation

The first Markov step extracts contour-based pathway candidates, removes configured straight runs (since these can be considered data errors), and assembles the selected one-dimensional mineral chain. Its program flow is shown in Appendix C.

The second step estimates the transition matrix, performs consistency and connectivity checks (density, spectral norm, irreducibility and row sums of the transition matrix), and generates repeated Markov-chain realizations and composition statistics; see Appendix D.

Detailed instructions are provided in:

```
1D_markov_simulation/USER_GUIDE_markov_simulation.md
```

4.3 Two-dimensional multi-Gaussian simulation

Step 01 preprocesses selected mask areas, samples balanced indicator classes, estimates empirical variograms, and fits Matérn models. The corresponding analysis sequence is shown in Appendix E. Step 02 converts mineral fractions into exact integer target-pixel counts based on the frame size to be simulated and generates hierarchical two-dimensional mineral fields; see Appendix F. The internal subprocess for Gaussian-field generation and exact masked-pixel assignment is detailed in Appendix G. Optional joint-mineral fields separate configured resplit pairs. Step 03 aggregates outputs and can compare simulated neighbour statistics with those calculated from the preprocessed training data; its program flow is shown in Appendix H. For directory-based Step 03 inputs, each configured simulation directory may specify one shared `values_file` that is applied to all neighbour matrices matching the configured filename pattern. Smoothed reference counts and neighbour data can be supplied either as three explicit files or through an `add_ons` smoothing manifest.

Detailed instructions are provided in:

```
2D_multigaussian_simulation/USER_GUIDE_multigaussian_simulation.md
```

4.4 Combined result plotting

This optional combined workflow counts mineral occurrences in every Markov chain and the neighboring minerals along grain boundaries in multi-Gauss simulations. The values from both methods are summarized in a single boxplot distribution by mineral. Samples can contribute according to their number of simulations, with equal total weight per result directory, or with user-defined weights. Numerical boxplot values and effective sample contributions are exported as TXT files. The complete program flow is provided in Appendix I.

Detailed instructions are provided in:

```
combined_result_plotting/USER_GUIDE_combined_result_plotting.md
```

4.5 Optional add-ons

The add_ons workflows are optional tools that complement the main GeoPathSim workflows. Most add-ons use outputs from data preprocessing, such as compact mineral rasters, mask rasters, BSE rasters, compositions, or neighbour statistics. Depending on the analysis, however, results from later workflow steps may also serve as inputs. In particular, the bulk-versus-pathway workflow can use training-pathway data produced by Markov Step 01 to compare the minerals represented in the extracted training data with the bulk mineral composition.

The add-ons can highlight selected minerals with metadata or custom colors, compare bulk and pathway-related mineral compositions, smooth categorical mask-area classifications while producing files directly usable by Multi-Gaussian Step 03, crop mineral and BSE rasters, and determine the frame size of representative volume elements using configurable sliding windows. The corresponding program flows are shown for mineral highlighting in Appendix J, bulk-versus-pathway comparison in Appendix K, categorical smoothing in Appendix L, and image cropping in Appendix M. The optimized RVE workflow performs the sliding-window analysis and the subsequent RVE-size evaluation in one run using a common configuration; its overall program flow is shown in Appendix N. The computational sliding-window analysis and RVE-size evaluation remain separate internal modules, documented in Appendices O and P, but no manual transfer of an analysis output path is required. Each add-on uses its own JSON config and timestamped output directory.

Detailed instructions are provided in:

```
add_ons/USER_GUIDE_add_ons.md
```

5 Some key concepts and definitions

5.1 Representative volume element (RVE)

We define a representative volume element (RVE) as the smallest image or material domain that is sufficiently large for a selected property to become statistically stable with respect to further increases in the observation-window size. The RVE size therefore depends on the property under consideration and on the convergence criterion used to define an acceptable variation.

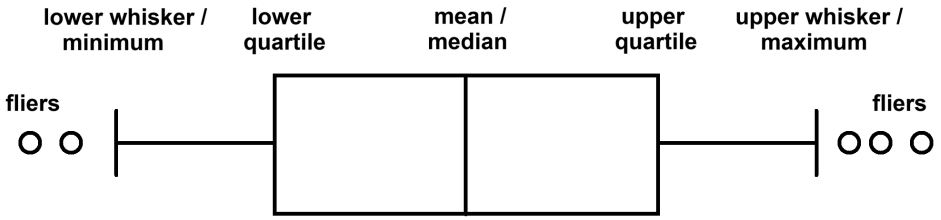
In GeoPathSim, the RVE analysis focuses primarily on the bulk mineral composition. A window is considered representative when its mineral fractions approximate the bulk composition of the complete analysed image within the

configured tolerance and remain sufficiently stable as the window size increases. This emphasis is chosen because the bulk mineral fractions are also used as target proportions in the two-dimensional multi-Gaussian simulation.

The resulting RVE should therefore be interpreted specifically as representative with respect to mineral composition. It does not necessarily guarantee that other structural properties, such as grain size and shape, mineral-contact frequencies, spatial continuity, variogram characteristics, or pathway connectivity, are represented at the same window size.

5.2 Details on boxplot statistics

Several GeoPathSim workflows export mineral-wise boxplots and corresponding numerical boxplot statistics. The following sketch illustrates the meaning of the reported terms.



The terms are interpreted as follows:

- `lower_whisker / minimum`: the lower end of the whisker, that is, the smallest value still considered part of the main distribution according to the configured whisker rule (”2sigma” for whisker=2.275% of all data points or ”1.5_iqr” for whisker=Q1-1.5*IQR with IQR=Q3-Q1 interquartile range).
- `lower_quartile`: the 25th percentile (Q1). About 25% of the values lie below this level.
- `median`: the 50th percentile. Half of the values lie below it and half above it.
- `mean`: the arithmetic average of all values in the distribution.
- `upper_quartile`: the 75th percentile (Q3). About 75% of the values lie below this level.
- `upper_whisker / maximum`: the upper end of the whisker, that is, the largest value still considered part of the main distribution according to the configured whisker rule (”2sigma” for whisker=97.725% of all data points or ”1.5_iqr” for whisker=Q3+1.5*IQR with IQR=Q3-Q1 interquartile range).
- `fliers`: values outside the whisker range. Depending on the selected plotting settings, they may be shown as individual points or suppressed in the figure, but they are still part of the underlying data.

6 Input data

6.1 Analytical dataset and provenance

The project includes harmonized class metadata for SEM-based automated mineralogy data from **2 granodiorite samples from the Kindisch Quarry in Upper Lusatia, Germany**. Polished thin sections were analysed with a Mineral Liberation Analyzer combining backscattered-electron imaging and EDX spectroscopy for mineral identification. Measurements were conducted at a resolution of **3 × 3 µm per pixel**.

The corresponding source dataset is:

> Bachmann, Kai, Renno, Axel D., Pospiech, Solveig, & Duckstein, Alexandra. (2026). *SEM-based automated mineralogy analysis of various granodiorite samples from the Kindisch quarry (Upper Lusatia, Germany)* (Version 1.0) [Data set]. RODARE. DOI: `10.14278/rodare.4896`.

The project-specific files `minerals_joint.txt`, `minerals_specgrav.txt`, and `residual_mask.txt` additionally contain geological interpretation and modelling choices made for the GeoPathSim workflows. They must not be interpreted as unmodified fields copied directly from the RODARE dataset.

6.2 Raster image inputs

The raster input data consist of paired mineral-classification and backscattered-electron images:

1. `*_MINERAL_SAMPLE.bmp` contains the spatially assigned mineral information for the corresponding sample. Each pixel stores a class colour representing the mineral or non-mineral class assigned by the SEM-based automated mineralogy analysis. The original colour legend is provided in `Granite_Legend.png`. GeoPathSim converts the colour-coded image into an integer mineral-index raster using the aligned entries in `minerals_color.txt` and `minerals_name.txt`.
2. `*_BSE_SAMPLE.bmp` contains the backscattered-electron contrast data as grayscale values. Brighter and darker pixels represent relative differences in backscattered-electron intensity. The BSE image is an additional image product and may have different pixel dimensions from the corresponding mineral-classification raster.

The wildcard character * denotes the sample-specific file prefix. Mineral and BSE files belonging to one sample must be associated through their common sample designation rather than through identical image dimensions.

Depending on the selected workflow, `data_input/` may additionally contain mask images or integer mask rasters, residual-class definitions, joint and resplit mineral definitions, and manifests or selected output files from earlier workflow steps.

6.3 Legend harmonization and class corrections

The original class information in `Granite_Legend.png` was adapted to provide a consistent GeoPathSim classification:

- `EmptyPixel` was added as a dedicated technical class for pixels without assigned mineral information.
- `Residuals` was added as a target class for minerals and auxiliary classes merged according to `residual_mask.txt` before subsequent geochemical modelling.
- `Magnetite_Ti` was simplified to `Magnetite`.
- `Ilmenite with Mg` was simplified to `Ilmenite`.
- `Sphalerite with medium Fe` was simplified to `Sphalerite`.

- `Chromite` and `Enstatite` were reassigned to `NoData` because their interpretation was considered geologically questionable for the investigated samples.
- `Unknown`, `No_XRay`, and `LowCounts` were combined under the common class name `NoData`.

These are classification and modelling decisions made for GeoPathSim. The harmonized class names and colours are stored in `minerals_name.txt` and `minerals_color.txt`; the original `Granite_Legend.png` remains the source reference for the analytical class colours.

6.4 Index-aligned metadata files

The following files are strictly aligned by line number:

```
minerals_name.txt
minerals_color.txt
minerals_specgrav.txt
residual_mask.txt
```

For every index i, line i in all four files must describe the same mineral or class. They must contain the same number of entries and must not be reordered independently.

- `minerals_name.txt` contains the mineral or class name assigned to each integer index, beginning with index 0.
- `minerals_color.txt` contains one RGB triplet in the range 0–255 for every mineral index and preserves the analytical class colours.
- `minerals_specgrav.txt` contains one dimensionless specific-gravity value for every mineral index.
- `residual_mask.txt` contains one binary value for every mineral index: `0` retains the mineral as an individual class and `1` merges it into `Residuals`.

If duplicate names occur in the source metadata, the canonical mapping produced by data preprocessing is authoritative: every duplicate source index is represented by the smallest index carrying that name.

6.5 Joint-mineral definitions

`minerals_joint.txt` defines mineral classes treated as genetically or spatially related, primarily in the 2D multi-Gaussian workflow. The current entries are:

```
Biotite + Chlorite | resplit
Albite + Plagioclase | stick
```

Chlorite occurs only in association with biotite and is interpreted as biotite alteration, especially chloritisation. The keyword `resplit` marks that combined treatment is followed by separation into the original mineral classes where required. Albite occurs predominantly along plagioclase grain boundaries; `stick` indicates that this association is retained in the corresponding processing step.

6.6 Specific gravity and residual classes

GeoPathSim uses the specific gravity (dimensionless density ratio) as relative-density factors to convert volume frequencies into normalized weight frequencies. For generic mineral groups, the selected value is representative within a composition-dependent range rather than a unique constant.

A specific gravity of 3.2 is assigned to the classes `Residuals` and `NoData`. This value is intentionally higher than the frequently cited average value of approximately 2.65 for granite. The bulk-granite average is dominated by comparatively light major rock-forming minerals, particularly quartz and feldspars. These minerals are typically treated as separate classes during the modeling process and therefore contribute only to a limited extent to the `Residuals` class. The value of 3.2 is therefore used as a representative modelling estimate for this mineral assemblage rather than as an average specific gravity of the entire granite.

The `EmptyPixel` value 0.01 is a technical placeholder and not a physical mineral property. Empty pixels should preferably be excluded from volume-to-weight conversion.

The current residual mask assigns Ilmenite, Titanite, Actinolite, Hornblende, Apatite, Monazite, Sphalerite, and the `NoData` classes at indices 21–26 to `Residuals`. The existing `Residuals` class at index 27 is the target class and is not itself marked for merging.

6.7 Current class table

Index	Mineral/Class	RGB	Specific gravity (dimensionless)	Residual mask
0	EmptyPixel	255 255 255	0.01	0
1	Albite	186 216 252	2.62	0
2	Orthoclase	255 192 128	2.59	0
3	Plagioclase	156 199 250	2.65	0
4	Quartz	254 203 230	2.65	0
5	Hematite	153 51 255	5.26	0
6	Magnetite	157 0 157	5.18	0
7	Biotite	233 150 122	3.00	0
8	Muscovite	139 58 58	2.82	0
9	Chlorite	126 150 101	2.95	0
10	Calcite	0 0 255	2.71	0
11	Pyrrhotite	255 0 0	4.61	0
12	Epidote	73 186 130	3.44	0
13	Chamosite	99 118 80	3.20	0
14	Ilmenite	0 236 0	4.72	1
15	Titanite	255 136 255	3.54	1
16	Actinolite	0 166 0	3.18	1
17	Hornblende	0 128 0	3.15	1
18	Apatite	230 80 193	3.15	1
19	Monazite	168 136 193	5.25	1
20	Sphalerite	255 72 72	4.00	1
21	NoData	81 81 81	3.20	1
22	NoData	0 85 0	3.20	1
23	NoData	211 211 211	3.20	1
24	NoData	128 128 128	3.20	1
25	NoData	169 169 169	3.20	1
26	NoData	0 0 0	3.20	1
27	Residuals	204 204 204	3.20	0

7 Output data

GeoPathSim writes outputs separately for each workflow. Depending on the enabled options, these may include:

- compact mineral and mask rasters with duplicate names consolidated to canonical indices;
- an optional standalone mineral-color legend;
- bulk and mask-area composition tables;
- training-data neighbour matrices;
- pathway candidates and selected one-dimensional mineral chains;
- transition matrices and transition diagnostics;
- Markov-chain realizations in optional TXT or NPZ format and composition statistics;
- empirical variograms and fitted Matérn parameters;
- two-dimensional stochastic mineral fields;
- realization compositions and neighbour statistics;
- combined 1D/2D mineral-count boxplots, boxplot-value tables, and sample-weight tables;
- plots, effective configuration files, manifests, and log files.

The effective configuration and manifest written during each run should be archived together with the corresponding scientific results.

8 Figure formats

Every workflow configuration contains a `plotting` section with `dpi` and `formats`:

```
"plotting": {
  "dpi": 300,
  "formats": ["png", "pdf"]
}
```

Supported figure formats are `png`, `pdf`, `svg`, and `eps`. When multiple formats are listed, every enabled figure is generated in each selected format. PNG is a raster format whose resolution is controlled directly by `dpi`; PDF, SVG, and EPS are vector formats, although rasterized image content within them can still depend on the configured DPI. The same `dpi` and `formats` settings apply to every enabled figure within each workflow.

9 Reproducibility

Random operations use explicit seeds:

- the variogram seed controls balanced pixel subsampling;
- the Markov seed controls simulated one-dimensional chains;
- the 2D seed and seed strategy control deterministic Gaussian random fields.

Using the same software version, input data, configuration, dependency versions, and random seeds should reproduce the same numerical workflow. Differences in library versions may influence random-number generation, image processing, variogram fitting, or plotting. Therefore, exact dependency versions should be preserved for archived releases.

10 Scientific limitations

- The input classification determines which mineral phases can be represented.
- Residual thresholds and joint/resplit definitions alter the effective class system.
- The 1D workflow represents mineral sequences along selected image-derived contours; it does not prove hydraulic connectivity.
- The 2D workflow generates stochastic realizations that reproduce selected proportions and spatial statistics; it does not uniquely reconstruct the original microstructure.
- Neighbour statistics describe raster contacts and depend on resolution and neighbourhood definition.
- GeoPathSim does not calculate flow, transport, sorption, mineral reactions, or reactive transport.
- Model outputs must be interpreted together with geological, mineralogical, and hydrological plausibility.

11 Funding

GeoPathSim was developed in the context of the **SANGUR** project (Systematic Sensitivity Analysis for Mechanistic Geochemical Models using Field Data from Crystalline Rock), funded by BMUKN (Federal Ministry for the Environment, Climate Action, Nature Conservation and Nuclear Safety), ref. no. 02E12112A.

12 CRediT authorship contribution statement

Alexandra Duckstein: Software, Formal analysis, Validation, Visualization, Writing – original draft

Solveig Pospiech: Data curation, Conceptualization, Supervision, Writing – review & editing

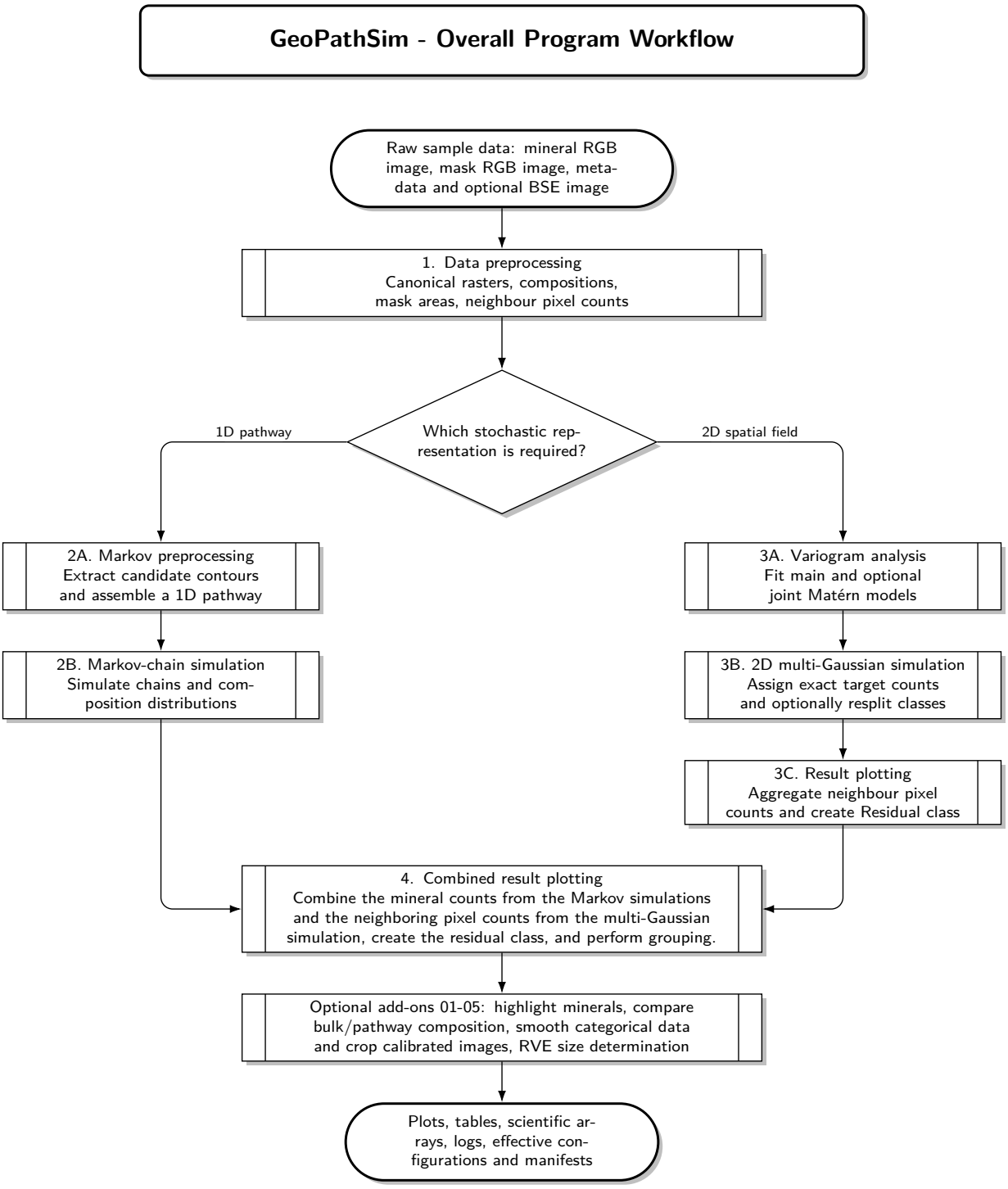
Raimon Tolosana-Delgado: Conceptualization, Methodology, Supervision

13 How to cite GeoPathSim

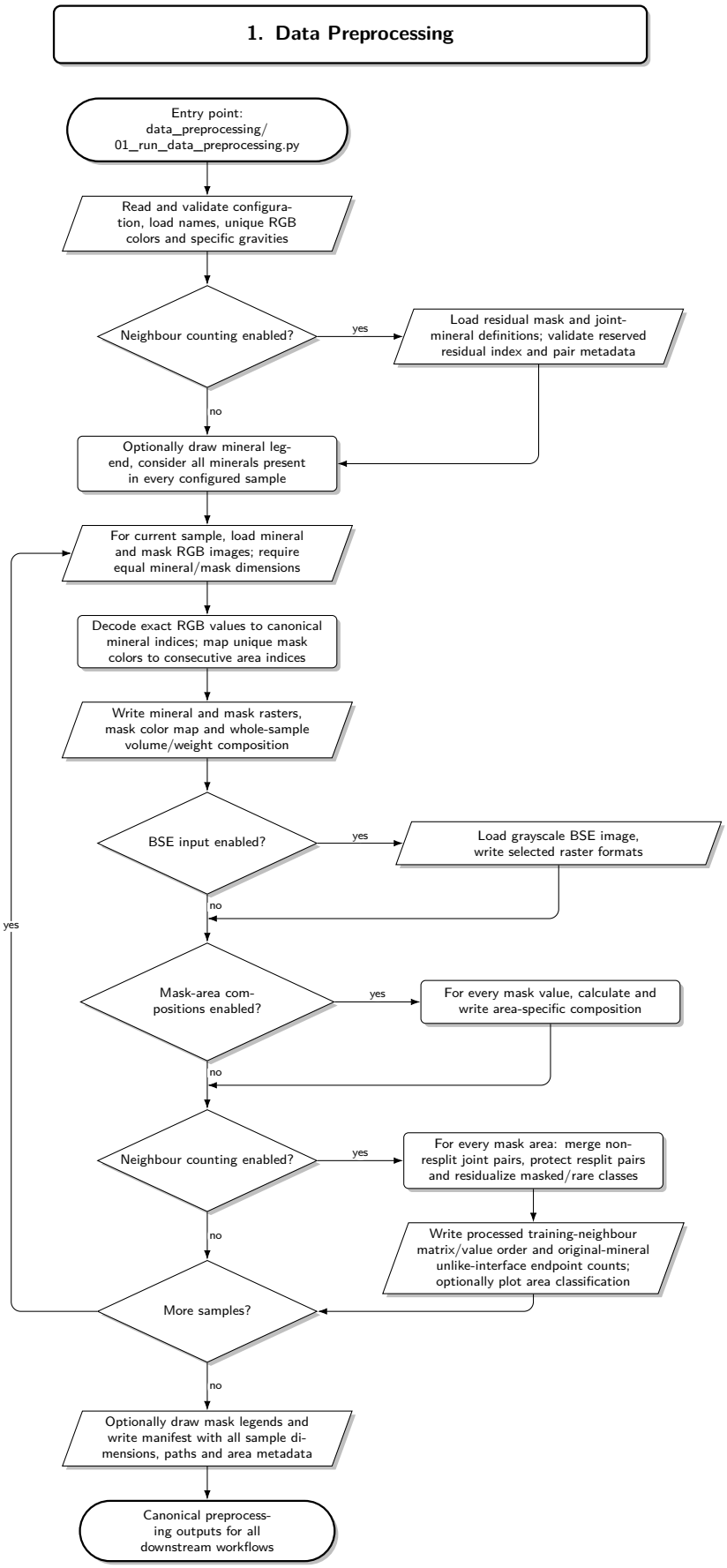
Duckstein, Alexandra; Pospiech, Solveig; Tolosana-Delgado, Raimon (2026). *GeoPathSim: A Python workflow for geostatistical simulation of mineral structures along potential transport pathways based on classified mineral images* (Version 1.0.0). RODARE. DOI: `10.14278/rodare.4933`.

Appendix

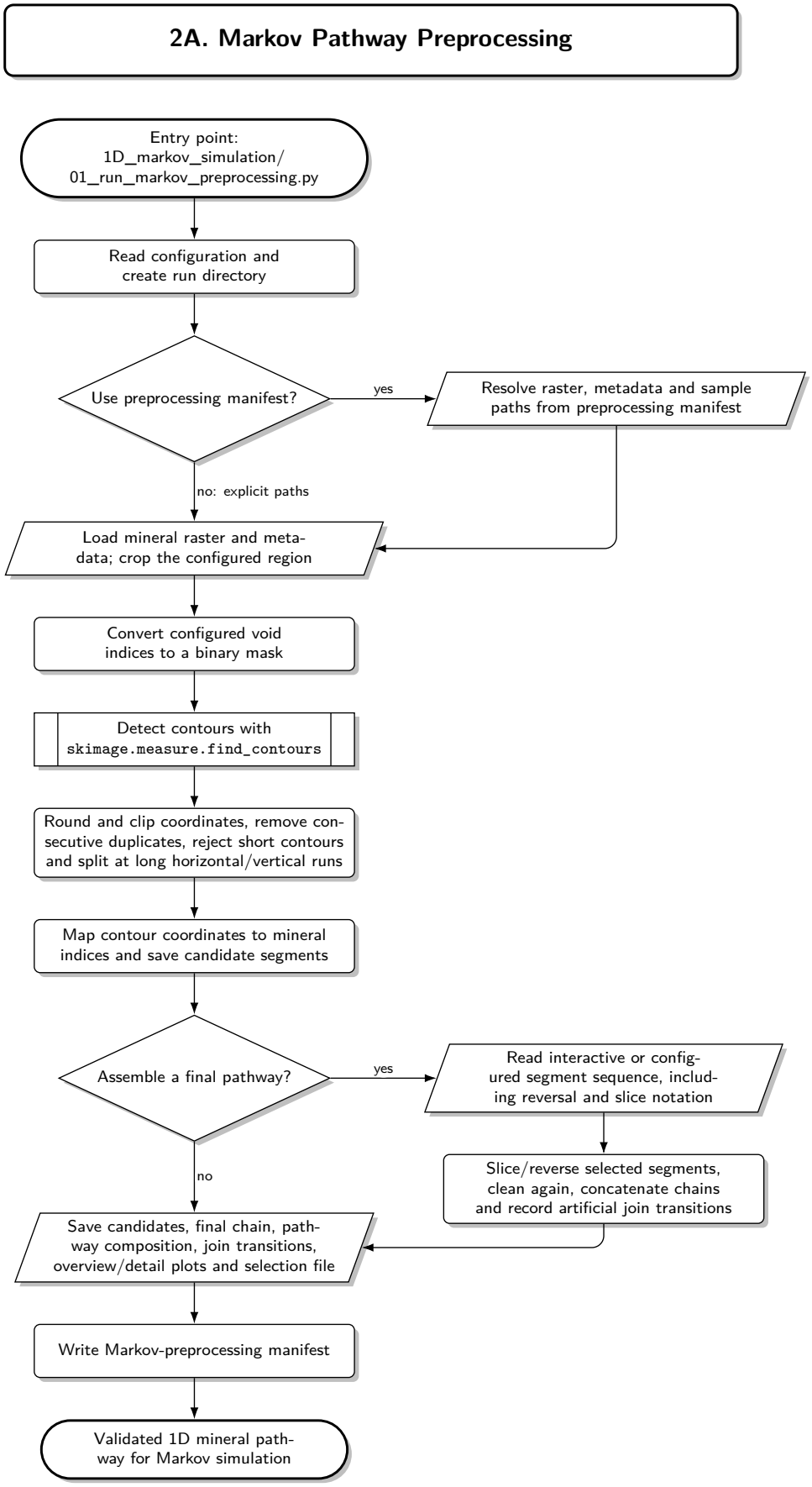
A Overall GeoPathSim workflow



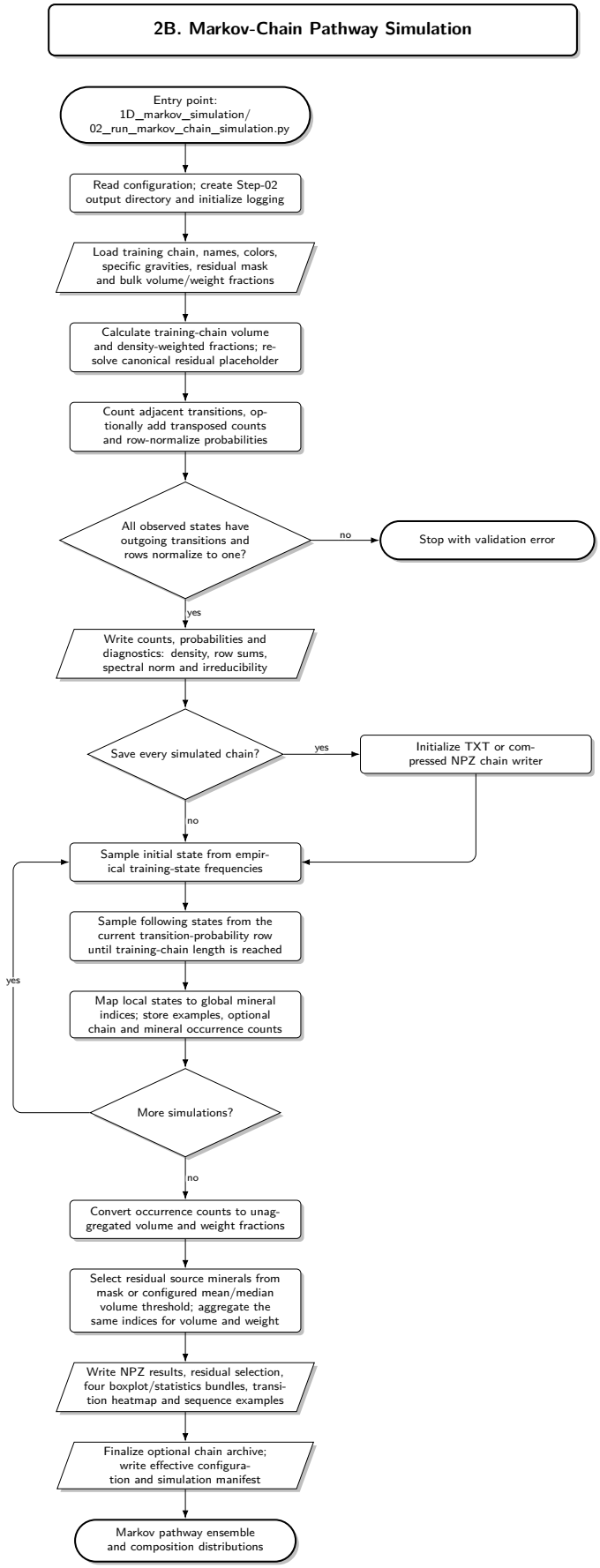
B Data preprocessing



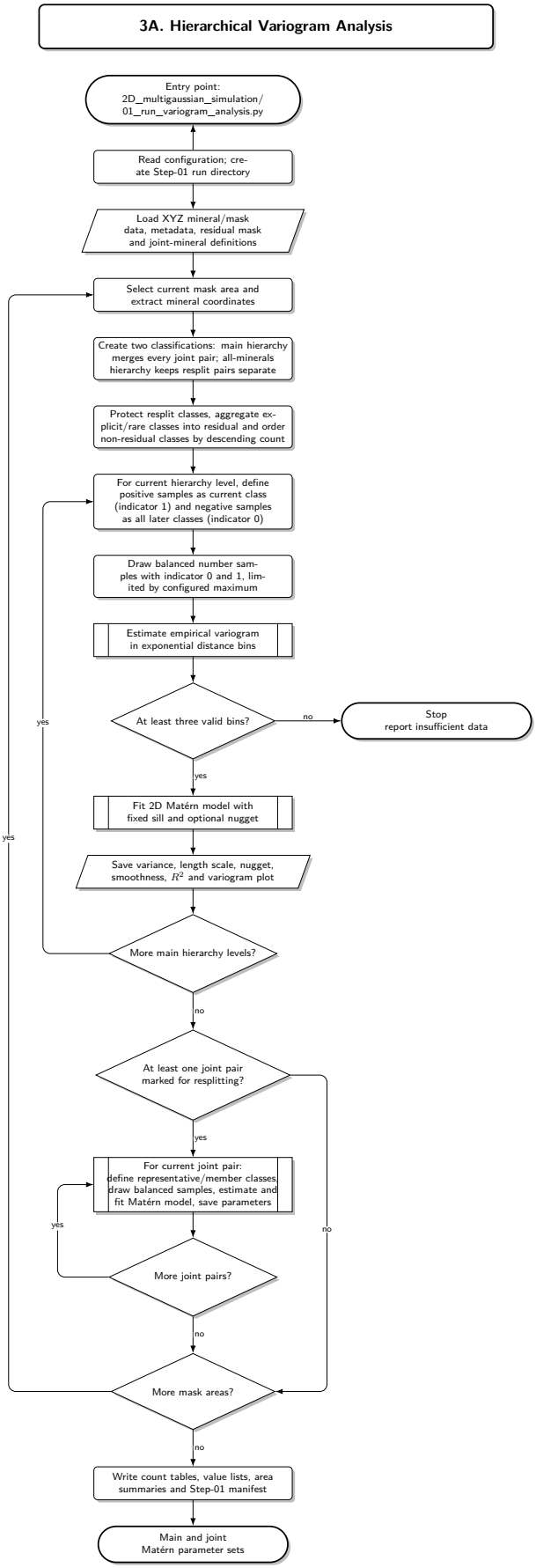
C Markov pathway preprocessing



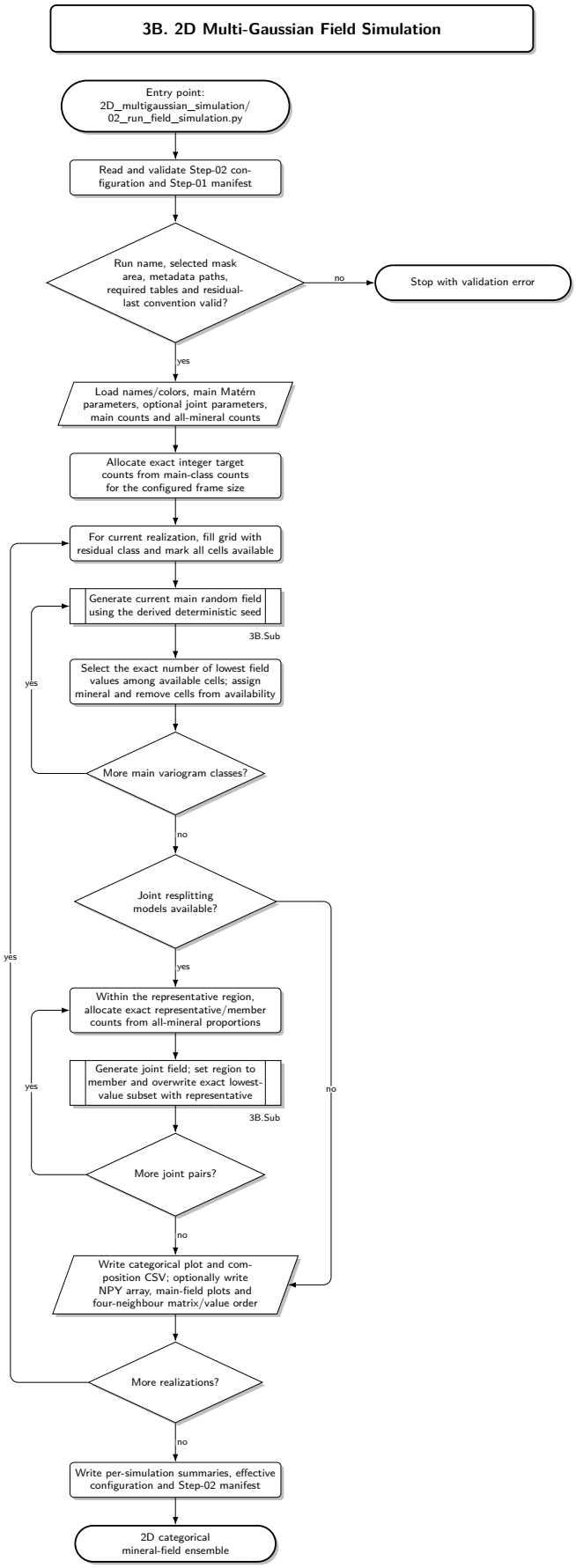
D Markov-chain pathway simulation



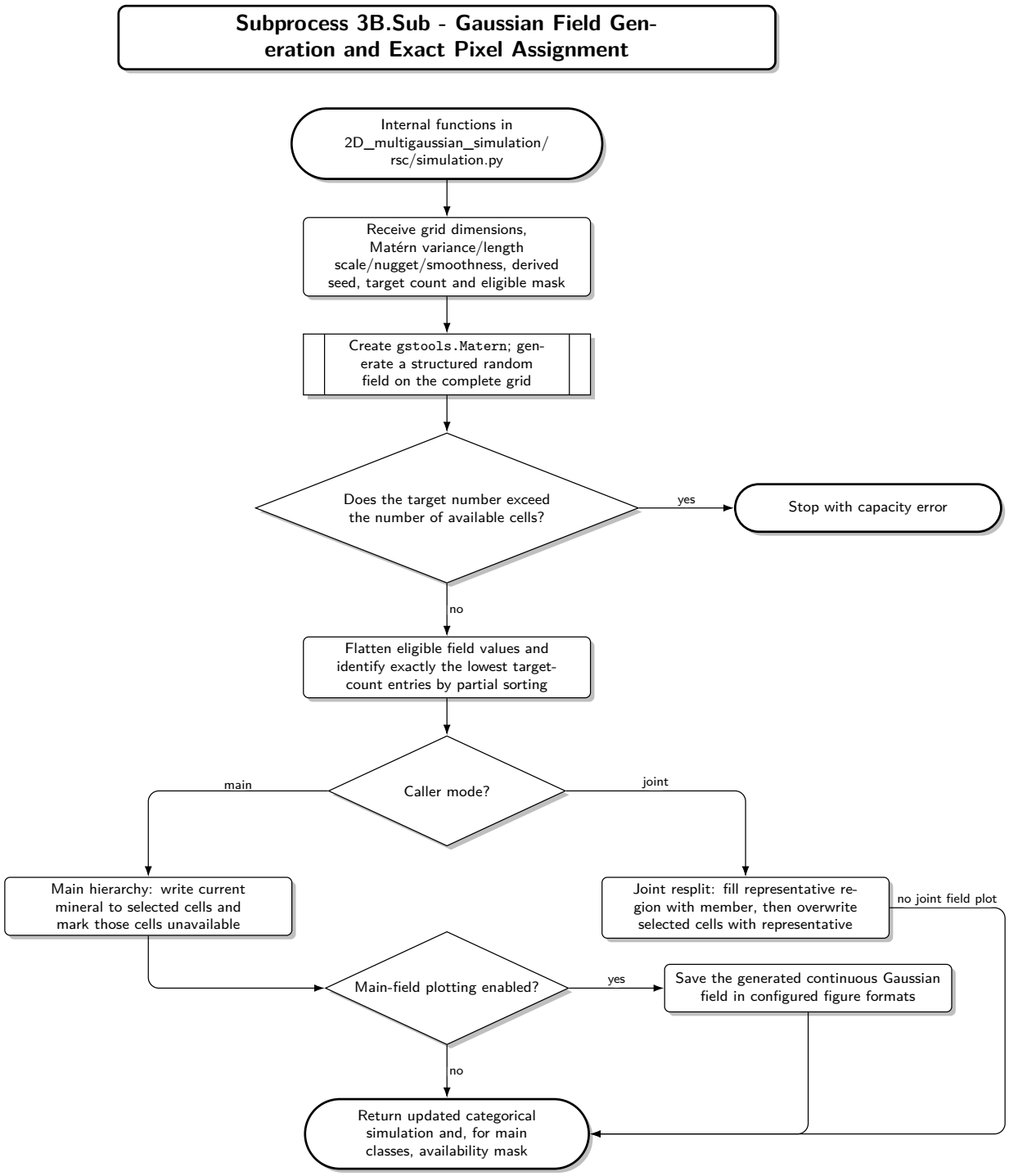
E Hierarchical variogram analysis



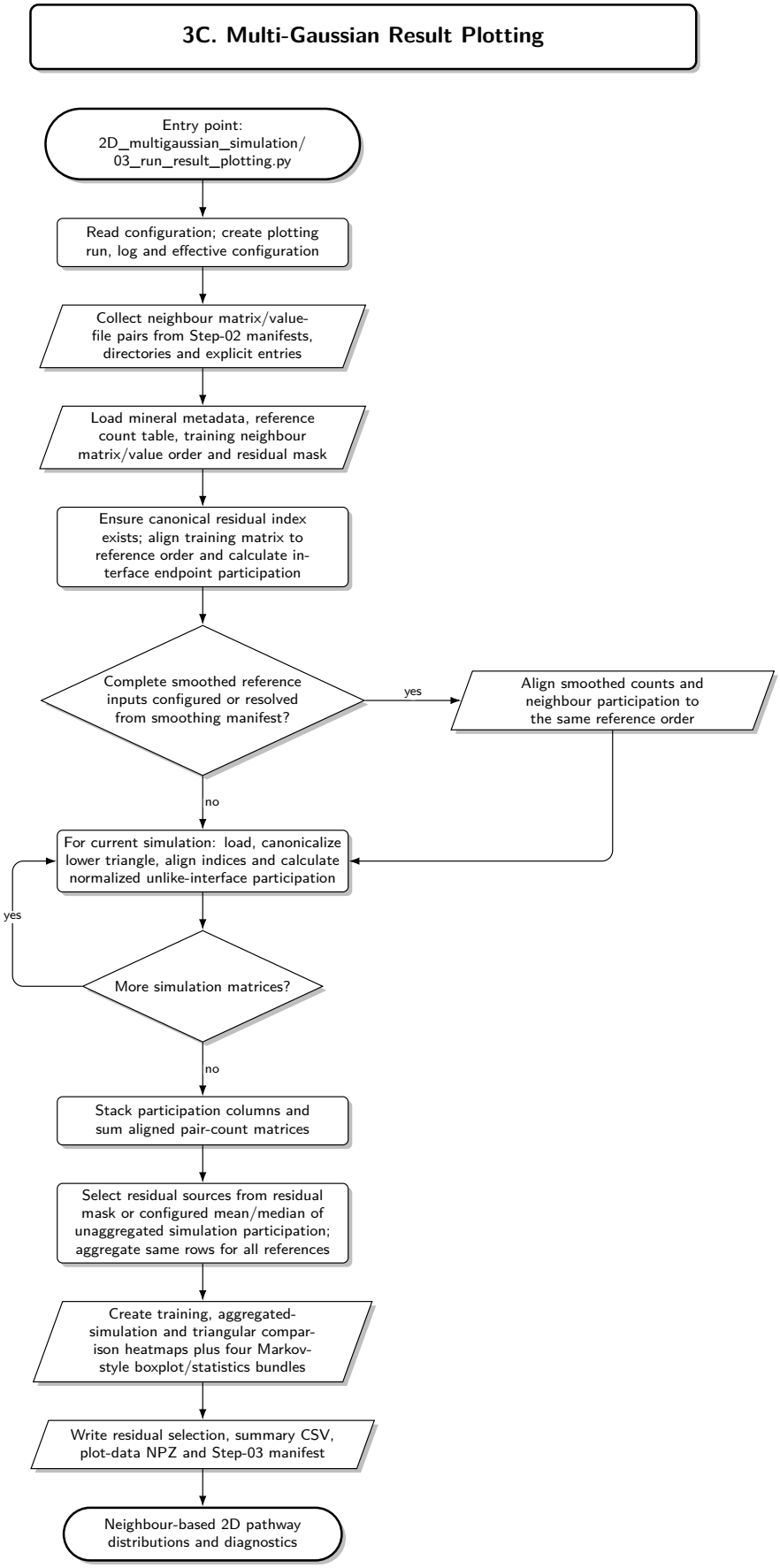
F Two-dimensional multi-Gaussian field simulation



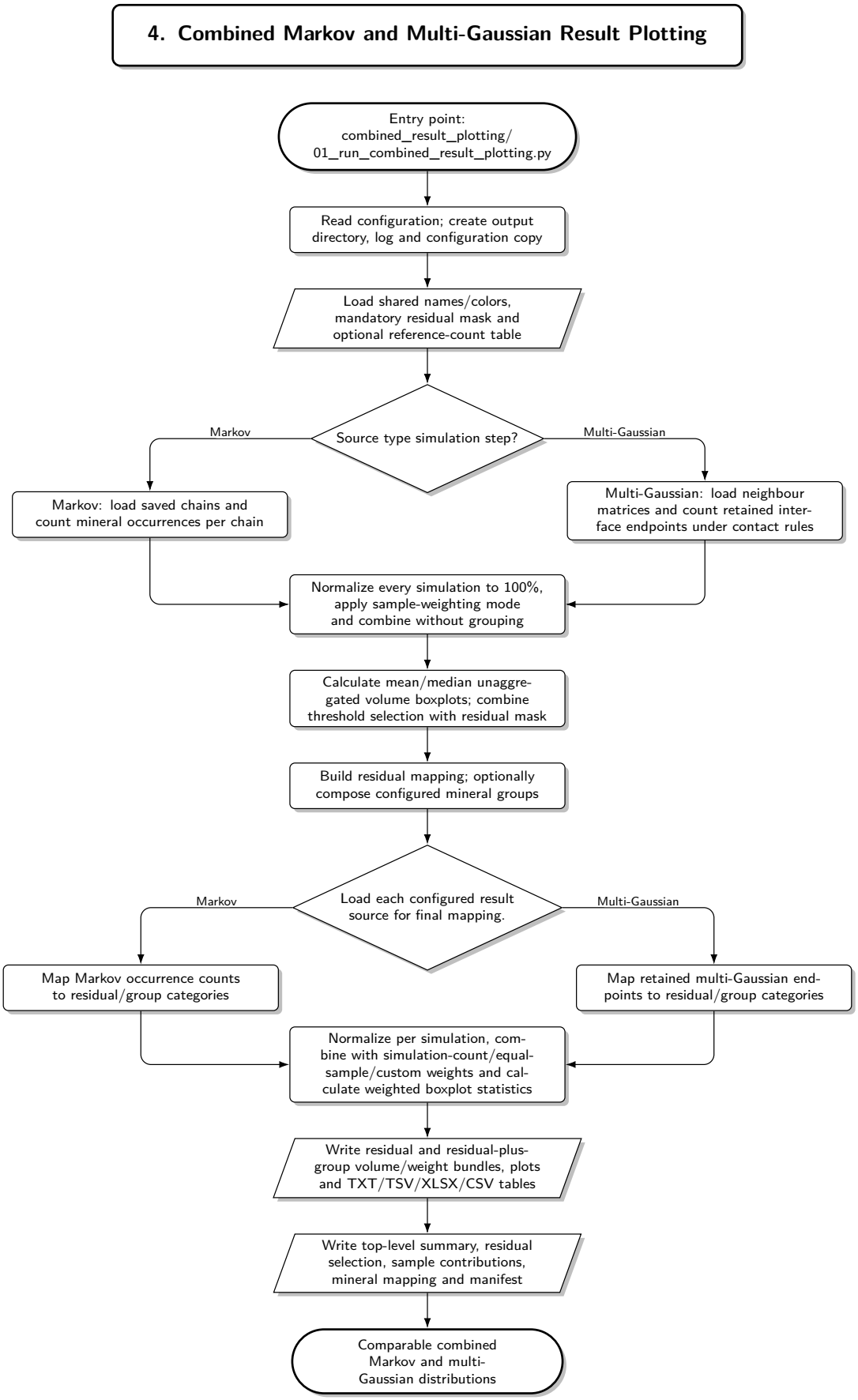
G Gaussian-field generation and exact pixel assignment



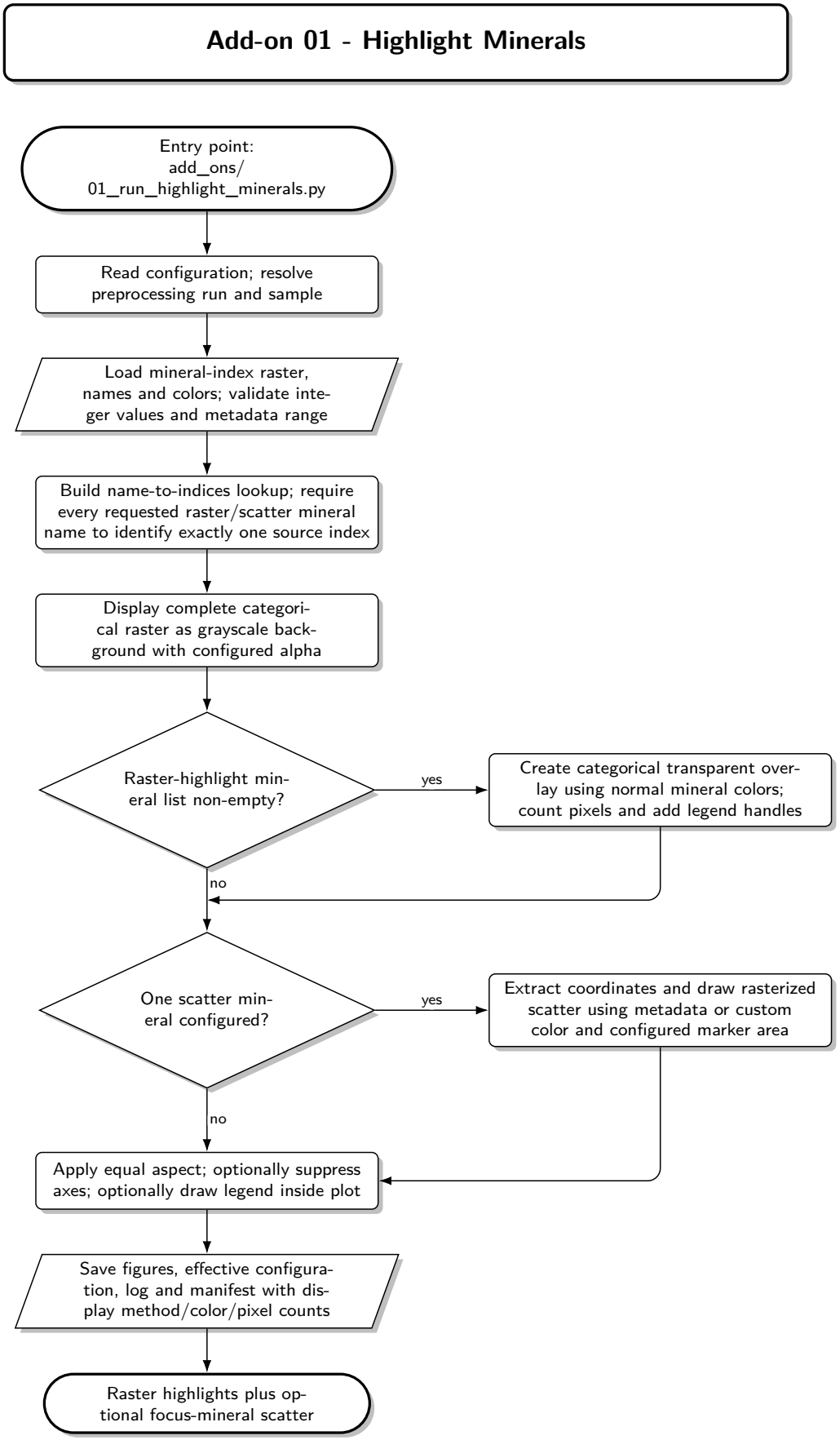
H Multi-Gaussian result plotting



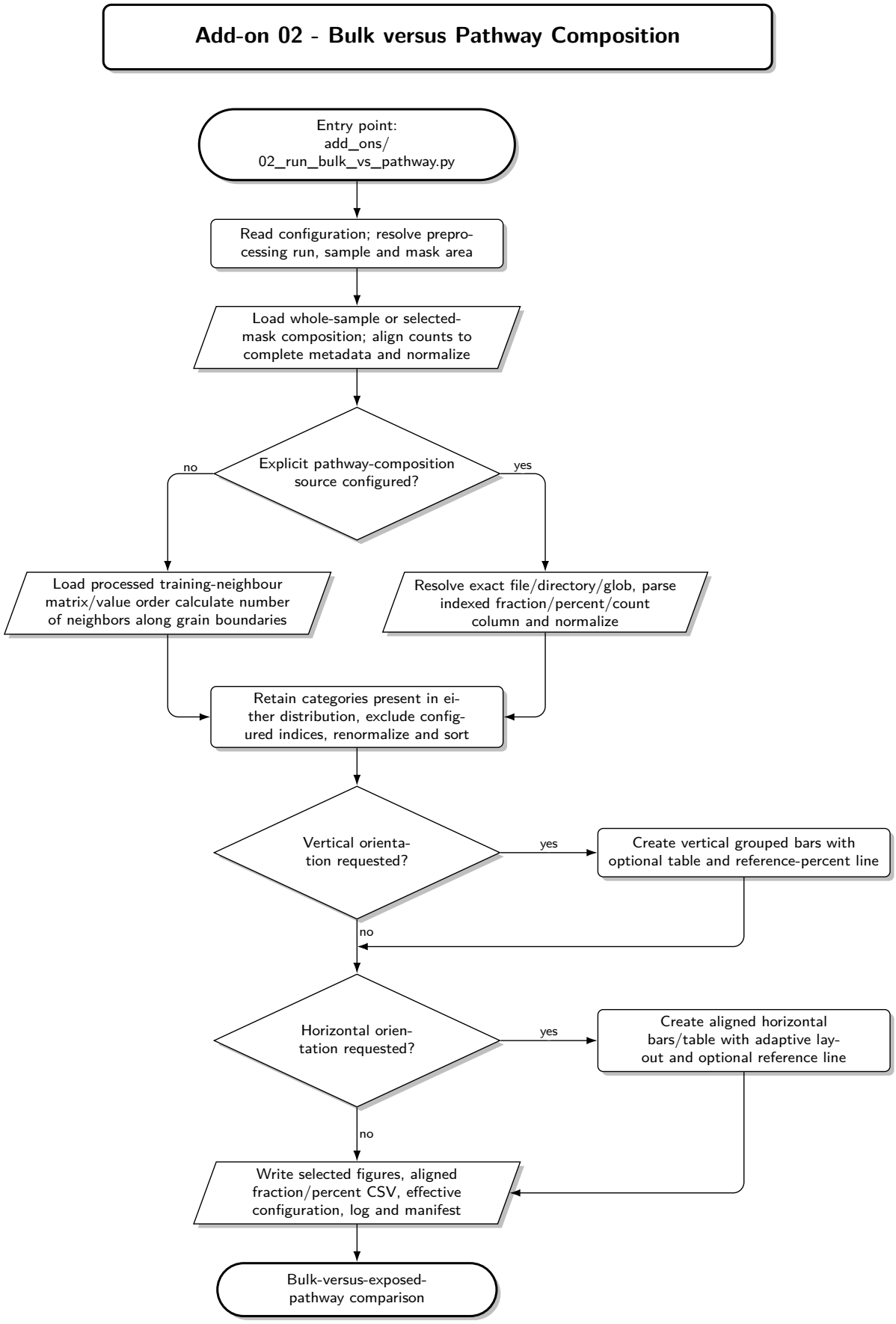
I Combined Markov and multi-Gaussian result plotting



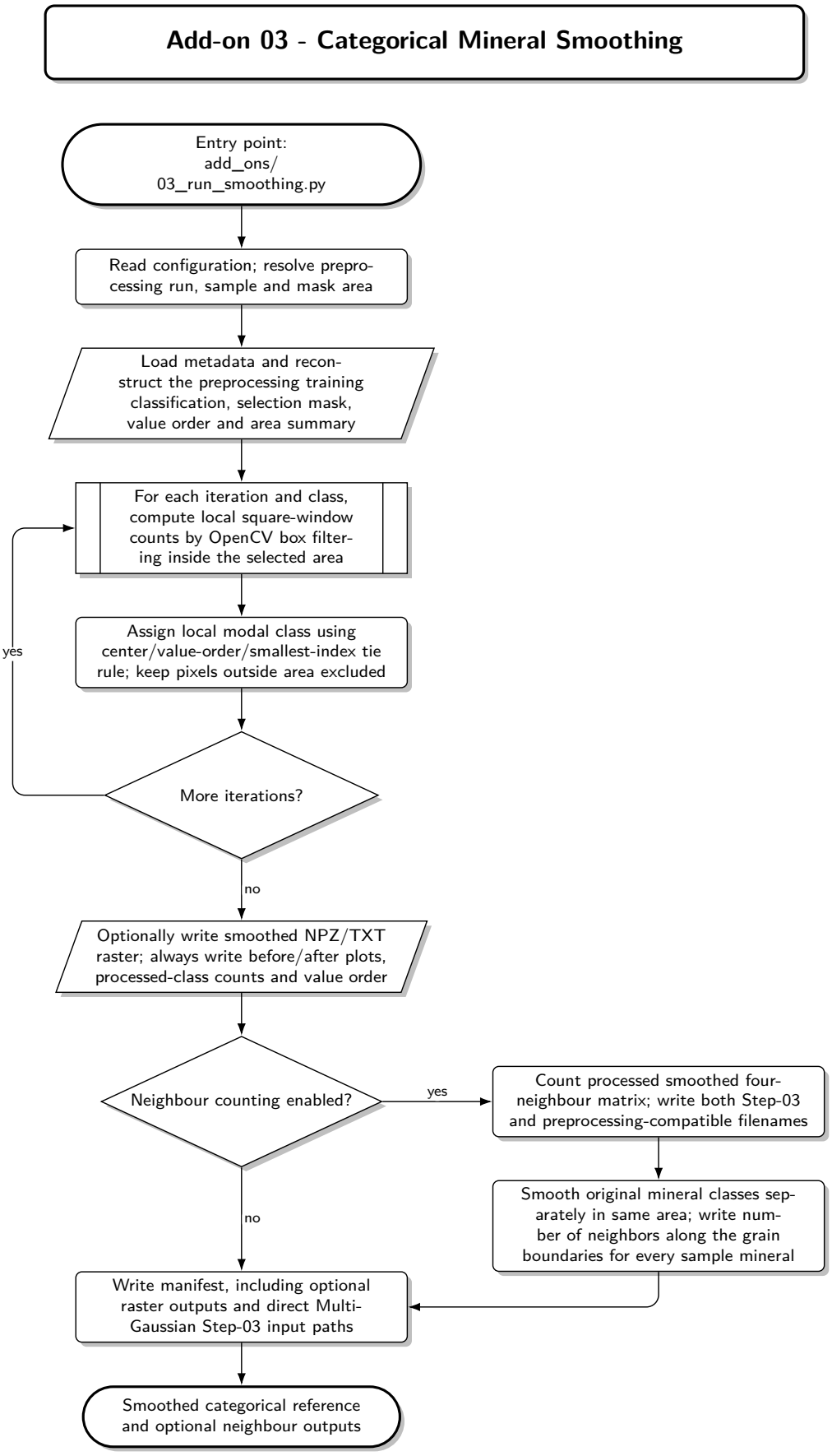
J Add-on 01: Highlight minerals



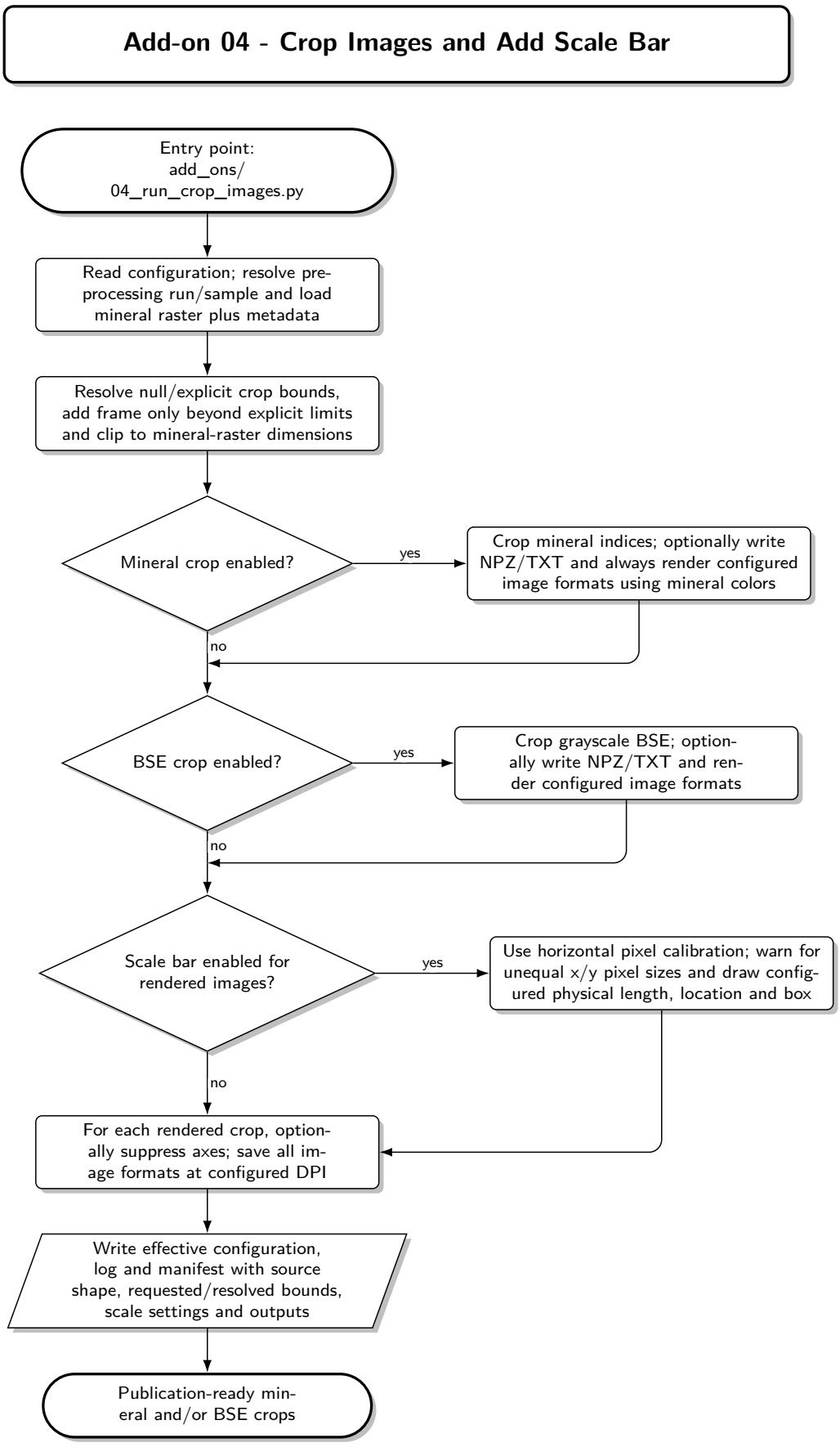
K Add-on 02: Bulk versus pathway composition



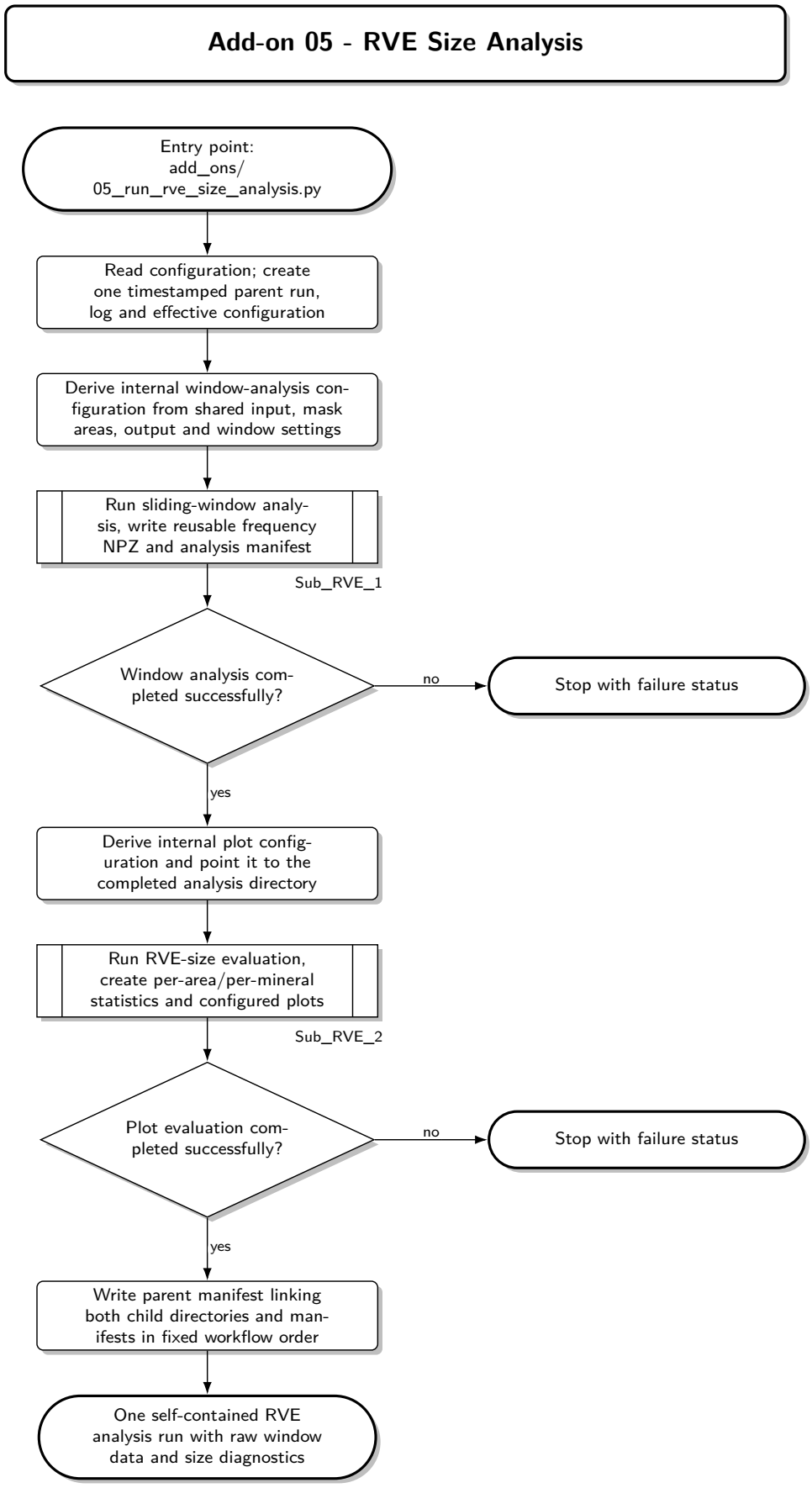
L Add-on 03: Categorical mineral smoothing



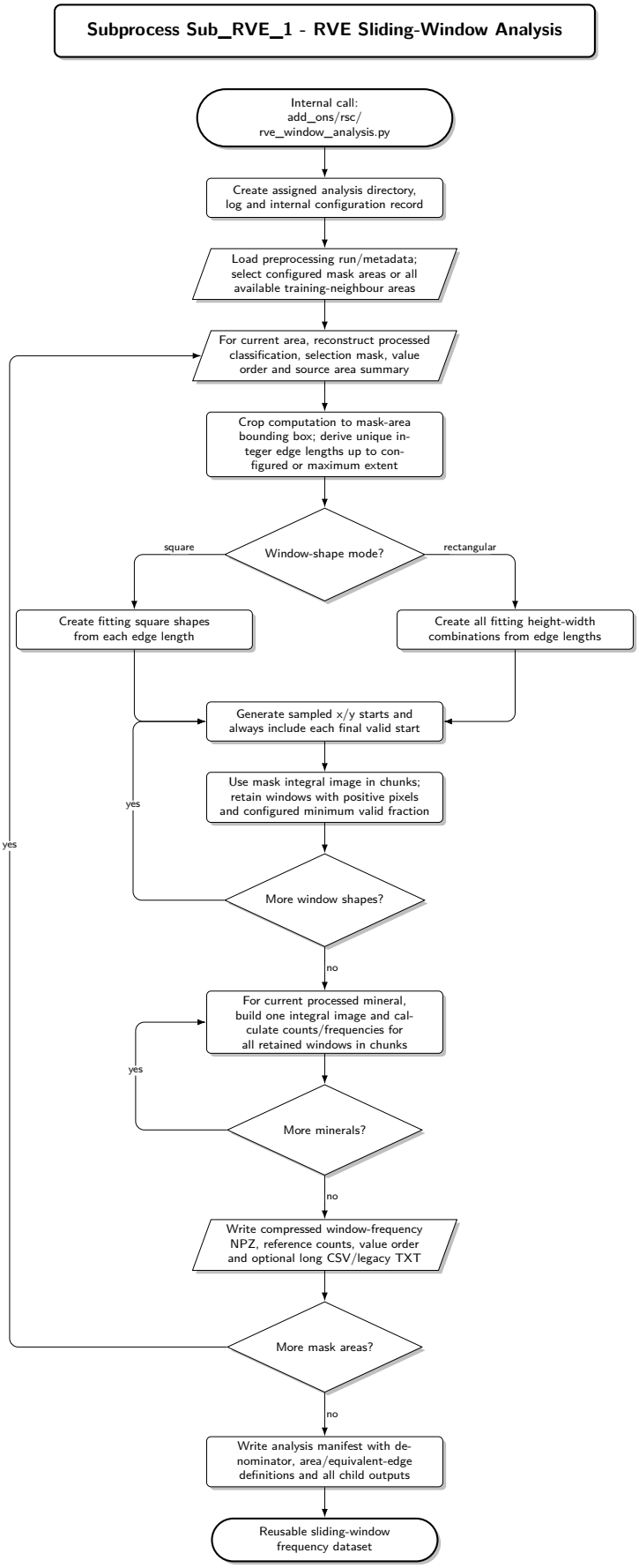
M Add-on 04: Crop images and add a scale bar



N Add-on 05: Combined RVE size analysis



O RVE sliding-window analysis subprocess



P RVE size evaluation and plotting subprocess