S2 — Derived Fields, Structural Representations and Operational Semantics

Burgwald Decision Stack — Reader

Chris Reudenbach

2026-01-24

S2: Derived raster and observation products

Burgwald decision stack: meta-level description

1. Why S2 Exists at All

In the Burgwald architecture, S2 marks the first step where *data becomes representation*.

S1 organizes observations: satellite images, digital elevation models, station measurements, provider formats, spatial reference systems, temporal indexing. At that level, the system is still epistemically passive. It stores, harmonizes and validates input material, but it does not yet impose any explicit interpretation on what the data *means* in relation to landscape structure, processes or modelling goals.

S2 changes that status fundamentally. Here, raw observations are transformed into **explicitly constructed fields that encode assumptions about space, structure, scale and relevance**. These transformations are not neutral. Each derived layer embodies a decision about what aspects of reality are considered stable, meaningful, measurable and reusable for later modelling stages.

At the same time, S2 deliberately avoids three things:

1. It does not perform classification or decision-making.
2. It does not learn from data in a statistical sense.
3. It does not optimize towards downstream targets.

S2 therefore sits between *measurement* and *inference*. It produces representations that are meant to be interpretable, reproducible and epistemically conservative. The purpose is not to extract maximal information, but to construct a stable representational substrate on which later segmentation (S3), signature formation (S4/S4L) and decision logic (S5) can safely operate.

A useful mental model is:
S1 answers *"What did we measure?"*
S2 answers *"How do we represent this measurement as spatial structure?"*

2. What "Representation" Means in This Context

The term **representation** is used here in a strict technical sense.

A representation is not a model prediction, not a classification and not a semantic label. It is a *mapping* from raw observation space into a structured feature space that:

- preserves spatial alignment and traceability,
- makes scale assumptions explicit,
- constrains noise and artefacts,
- and exposes interpretable dimensions.

Examples of representations in S2 are:

- a slope raster derived from a DEM,
- a canopy height percentile derived from DOM–DEM differencing,
- a breakpoint magnitude derived from a vegetation time series,
- a segment purity value derived from land-cover overlap.

These representations are **constructed quantities**. They encode design decisions (resolution, aggregation rules, thresholds, filters) that must remain visible and auditable. This is why S2 is strongly contract-driven: every output is registered, versionable and reproducible.

Crucially, S2 representations are still *pre-semantic*. They describe structure, magnitude, variability or signal behaviour, but they do not yet say what something *is*. Semantics enter only later, either through explicit supervision (S2-5) or through learned representations (S4L).

3. Structural Overview of the Current S2 Pipeline

The current implementation of S2 consists of five complementary pipelines that each generate a different type of representation:

Pipeline	Dominant domain	Representation type	Spatial nature
S2-1	Geomorphology / Hydrology	Static structural fields	Raster
S2-2	Vegetation structure	Aggregated structural fields	Raster
S2-3	Atmospheric context	Statistical descriptors	Tables / figures
S2-4	Temporal dynamics	Signal-derived indicators	Raster
S2-5	Semantic anchoring	Supervised labels + features	Vector / table

This heterogeneity is intentional. Landscape modelling requires multiple representational modalities: geometry, structure, dynamics, context and semantics cannot be collapsed into a single homogeneous feature space without losing interpretability and control.

The unifying principle is not data type, but **epistemic function**: each pipeline produces a representation that captures a specific aspect of spatial reality under explicit assumptions.

4. S2-1 — Terrain and Hydrological Structure from DEM

4.1 From Elevation to Structural Geometry

A digital elevation model is a measurement of surface height. By itself, it contains no explicit information about slopes, curvature, connectivity or flow behaviour. All of these properties are *latent* and must be derived.

The S2-1 pipeline transforms the DEM into a set of geomorphological and hydrological representations using SAGA GIS. The central design decision is the introduction of a **canonical spatial grid**: the DEM is resampled to a uniform 10 m grid that becomes the spatial backbone for all subsequent raster-based representations.

This step is not merely technical. It establishes a fixed spatial scale at which structural reasoning takes place. Any later aggregation, segmentation or signature computation inherits this grid implicitly. The grid therefore defines the spatial resolution of structural knowledge in the system.

4.2 Morphometric Representations

On the canonical grid, classical morphometric derivatives are computed: slope, aspect and multiple curvature measures. These fields encode local geometric properties of the terrain surface. Curvature measures, in particular, express second-order geometry and relate directly to flow divergence, convexity and accumulation tendencies.

Additionally, the Topographic Position Index (TPI) is computed at multiple spatial radii. TPI expresses relative elevation with respect to a neighbourhood and thus encodes landform context (ridge, slope, valley) in a scale-dependent manner.

These fields are not meant to simulate physical processes. They are **geometric descriptors that act as proxies for process-relevant structure**. Their epistemic role is explanatory rather than predictive.

4.3 Hydrological Representations

Hydrological derivatives are computed on a sink-filled DEM to enforce hydraulic connectivity. Flow accumulation, Strahler order, watershed identifiers and distance-to-channel fields are generated.

Again, these are not hydrodynamic simulations. They represent topological structure and potential flow organization under simplified assumptions. Their purpose is to encode landscape connectivity and drainage structure in a form that can later be used as predictors or stratification variables.

Taken together, S2-1 constructs a **structural geometry layer of the landscape**: static, deterministic, scale-defined and fully interpretable.[1]

5. S2-2 — Vegetation Biostructure from Surface and Ground Models

5.1 From Height Measurements to Structural Representation

DOM and DEM provide absolute height measurements of surface and ground. Their difference yields a proxy for vegetation height, but only if both grids are geometrically aligned. The pipeline enforces strict grid harmonization before differencing.

The differential canopy height model is defined as:

$$\mathrm{dCHM}(x, y) = \max(0,\ \mathrm{DOM}(x, y) - \mathrm{DEM}(x, y))$$

Negative values are explicitly clipped to zero. This is a semantic decision: negative canopy height has no physical interpretation and represents artefacts rather than signal.

5.2 Scale Stabilisation via Block Aggregation

The raw dCHM exists at 1 m resolution. Using it directly would inject excessive spatial noise and violate scale consistency with the DEM-derived fields. Therefore, strict block aggregation to the canonical 10 m grid is applied.

No smoothing kernels, no adaptive windows and no segmentation are used. The aggregation is purely geometric and preserves spatial alignment.

For each 10 m cell, four descriptors are computed: mean height, 95th percentile height, within-cell height variability and canopy fraction above a fixed threshold.

Formally, canopy fraction is defined as:

$$f_{\text{canopy}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\text{dCHM}_i > h_{\text{thr}}), \quad h_{\text{thr}} = 2\,\text{m}.$$

5.3 Epistemic Role

These fields do not represent species, biomass or ecological state. They represent **vertical structure and heterogeneity** in a deliberately coarse but robust way. They bridge the gap between raw geometric measurement and landscape-scale structural reasoning.

In later stages, they act as explanatory variables for segmentation and signature formation rather than as direct ecological indicators.[2]

6. S2-3 — Atmospheric Context from DWD Wind Observations

Unlike the raster-based pipelines, S2-3 operates on station time series and produces contextual descriptors rather than spatial fields.

Hourly wind measurements from DWD are ingested directly from provider archives. Station metadata are harmonized and spatial filtering is applied using a 20 km buffer around the AOI.

From these data, wind roses, directional distributions, seasonal summaries and aggregated statistics are generated. The outputs remain tables and figures; they are not rasterized or injected into the spatial predictor stack.

The epistemic function of this pipeline is **contextualization**. It characterizes the atmospheric regime in which local spatial patterns are embedded. This information supports interpretation, hypothesis generation and later stratification, but it does not directly drive segmentation or classification.

S2-3 therefore intentionally remains orthogonal to the spatial modelling chain.[3]

7. S2-4 — Temporal Disturbance Signals from Satellite Time Series

S2-4 introduces temporal structure into S2.

For each pixel of a Sentinel-2 data cube, a vegetation index time series is derived. NDVI is transformed into kNDVI to stabilize variance and reduce saturation effects:

$$\mathrm{NDVI} = \frac{NIR - RED}{NIR + RED}, \qquad \mathrm{kNDVI} = \tanh(\mathrm{NDVI}^2).$$

The kNDVI time series is analysed using `bfastmonitor` to detect structural breaks. Two quantities are retained:

- the temporal location of a detected breakpoint,
- the estimated magnitude of change.

These are not labels of disturbance types. They are **signal descriptors** indicating the presence and strength of temporal discontinuities. Interpretation remains external and contextual.

This pipeline introduces a dynamic dimension into S2 while preserving the principle of non-inference.

8. S2-5 — Semantic Anchoring via Supervision Transfer

S2-5 introduces the first explicit semantic linkage in the pipeline.

A categorical land-cover raster is intersected with segmentation polygons. For each segment, area-weighted class shares are computed and mapped into project-specific semantic classes.

The dominant class share defines the segment's purity:

$$\mathrm{purity} = \max_k(\mathrm{share}_k).$$

Segments below minimum purity or minimum area thresholds are excluded from training data generation. For the remaining segments, predictor means are attached to form a training matrix.

Two artefacts are produced: a fully annotated segment layer for diagnostics and a filtered training table for supervised learning.

This step is epistemically sensitive. It embeds assumptions about class geometry, scale compatibility and label noise. It must therefore remain transparent and adjustable. S2-5 does not claim semantic truth; it provides **operationally usable supervision under explicit constraints**.

9. Conceptual Synthesis

S2 constructs five complementary representation layers:

Dimension	Representation
Geometry	Terrain and hydrology
Vertical structure	Canopy biostructure
Environmental context	Wind climatology
Temporal dynamics	Disturbance signals
Semantic anchoring	Supervised labels

The unifying principle is not algorithmic similarity, but epistemic clarity: each representation encodes a specific aspect of spatial reality under controlled assumptions, without collapsing into premature inference or optimisation.

S2 therefore establishes the representational foundation upon which the higher abstraction levels of the Burgwald architecture operate.