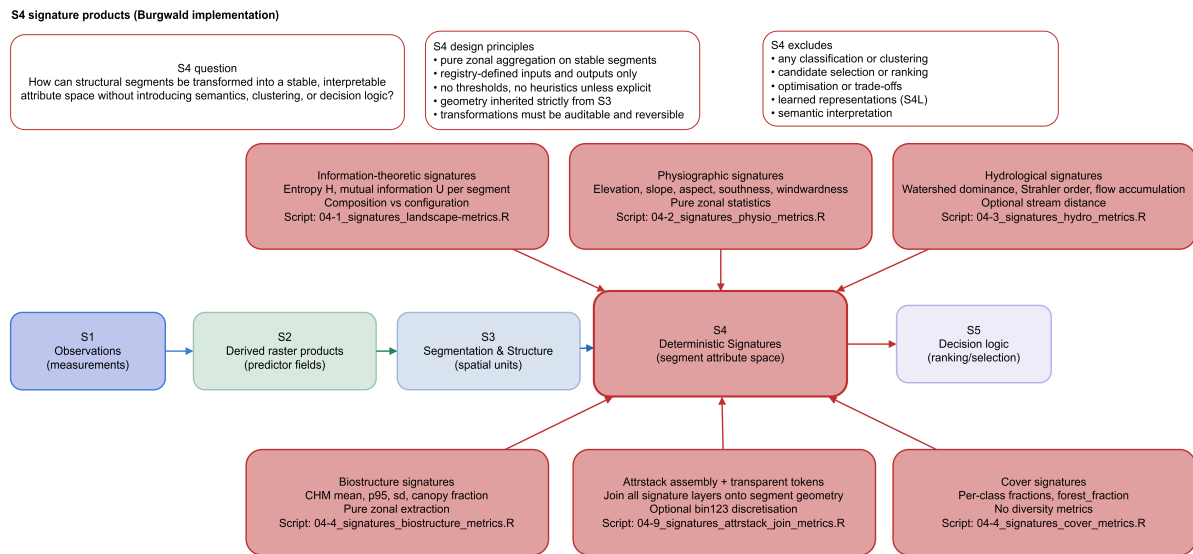# S4 — Signatures as Explicit Structural Representations

Burgwald Decision Stack — Reader

1 S4: Deterministic signatures and attribute space

Burgwald decision stack: meta-level description

S4 signature products (Burgwald implementation)

S4 question
How can structural segments be transformed into a stable, interpretable attribute space without introducing semantics, clustering, or decision logic?

S4 design principles
• pure zonal aggregation on stable segments
• registry-defined inputs and outputs only
• no thresholds, no heuristics unless explicit
• geometry inherited strictly from S3
• transformations must be auditable and reversible

S4 excludes
• any classification or clustering
• candidate selection or ranking
• optimisation or trade-offs
• learned representations (S4L)
• semantic interpretation

Information-theoretic signatures
Entropy H, mutual information U per segment
Composition vs configuration
Script: 04-1_signatures_landscape-metrics.R

Physiographic signatures
Elevation, slope, aspect, southness, windwardness
Pure zonal statistics
Script: 04-2_signatures_physio_metrics.R

Hydrological signatures
Watershed dominance, Strahler order, flow accumulation
Optional stream distance
Script: 04-3_signatures_hydro_metrics.R

S1
Observations
(measurements)

S2
Derived raster products
(predictor fields)

S3
Segmentation & Structure
(spatial units)

S4
Deterministic Signatures
(segment attribute space)

S5
Decision logic
(ranking/selection)

Biostructure signatures
CHM mean, p95, sd, canopy fraction
Pure zonal extraction
Script: 04-4_signatures_biostructure_metrics.R

Attrstack assembly + transparent tokens
Join all signature layers onto segment geometry
Optional bin123 discretisation
Script: 04-9_signatures_attrstack_join_metrics.R

Cover signatures
Per-class fractions, forest_fraction
No diversity metrics
Script: 04-4_signatures_cover_metrics.R

1.1 From Segments to Representable Objects

After S3 the system has achieved geometric stability. The MeanShift segmentation defines spatial units that are reproducible, scale-consistent and topologically clean. At this point, however, the system still does not *know* anything about these units. A segment is only a

polygon with an identifier and a shape. All semantic or structural meaning remains implicit in the raster fields underneath.

This situation is typical for object-based spatial analysis: segmentation produces spatial objects, but meaning only emerges once attributes are explicitly constructed and interpreted, not from geometry alone.[1] [2]

S4 exists to close exactly this gap. It transforms segments into **explicit representable objects** by attaching deterministic, interpretable descriptors. These descriptors are called *signatures*.

In practical terms, S4 consumes the stable segment geometry, the derived raster fields from S2 (relief, hydrology, biostructure), and categorical overlays used for diagnostic structure, and produces a consolidated attribute stack in which each segment carries a structured vector of metrics and derived tokens.

No learning, optimisation or classification occurs in S4. Every transformation is deterministic and reversible.

1.2 What a Signature Means in This Architecture

A signature is not a feature vector in the machine-learning sense. It is a **representational contract**: a fixed mapping from spatial structure into a finite, interpretable descriptor space. This notion aligns with geographic object models where meaning is explicitly attached through attributes rather than inferred implicitly from pixels or geometry.[3] [4]

Formally, for a segment s and a set of aligned representation layers L_k, a signature is defined as:

$$\mathbf{z}(s) = [f_1(L_1 \mid s), f_2(L_2 \mid s), ..., f_n(L_n \mid s)]$$

Each operator f_i is an explicit aggregation or transformation (mean, quantile, entropy, adjacency statistic, fraction). There is no parameter fitting and no data-driven adaptation at this level.

Three properties matter operationally.

First, **determinism**: the same inputs always produce the same signature.

Second, **interpretability**: each component corresponds to a physically or structurally meaningful quantity.

---

[1]Blaschke, T. (2010). *ISPRS Journal of Photogrammetry and Remote Sensing.*
[2]Hay, G. J., & Castilla, G. (2008). *Photogrammetric Engineering & Remote Sensing.*
[3]Hay, G. J., & Castilla, G. (2008). *Photogrammetric Engineering & Remote Sensing.*
[4]Couclelis, H. (1992). *Spatio-Temporal Reasoning in Geographic Space.*

Third, **scale binding**: the signature is valid only at the segmentation scale and raster resolution used upstream. This explicitly acknowledges the modifiable areal unit problem and aggregation effects that inevitably arise in spatial analysis.[5] [6] [7]

A signature therefore encodes structured description, not inference. This distinction is essential because later stages (S4L, S5) rely on the stability and auditability of these representations.

1.3 Why S4 Is Not Just "Feature Engineering"

Without an explicit signature layer, feature construction tends to drift toward whatever improves downstream model performance. The result is usually a growing collection of opaque predictors whose meaning, scale behaviour and coupling to physical processes becomes unclear.[8] [9]

S4 enforces a separation of concerns.

S2 answers how raw measurements are transformed into stable spatial fields.
S3 answers how space is partitioned into stable units.
S4 answers how each unit is described structurally in a controlled way.

Only after this representational discipline is established does it make sense to introduce learning or optimisation. This mirrors the broader distinction between handcrafted representations and learned embeddings in representation learning literature.[10] [11]

1.4 Structural Domains of the Current Signatures

The concrete S4 implementation reflects the actual scripts in the repository. Each domain contributes a coherent block of metrics that is later joined into the attribute stack.

1.4.1 Landscape configuration (information-theoretic metrics)

The landscape metrics pipeline computes composition and configuration descriptors per segment using a categorical raster.

Shannon entropy measures compositional diversity:

$$H = -\sum_k p_k \log(p_k)$$

[5]O'Sullivan, D., & Unwin, D. (2010). *Geographic Information Analysis.*

[6]Atkinson, P. M., & Tate, N. J. (2000). *Professional Geographer.*

[7]Goodchild, M. F. (2011). *GIS&T Body of Knowledge.*

[8]Guyon, I., & Elisseeff, A. (2003). *JMLR.*

[9]Lipton, Z. C. (2018). *Communications of the ACM.*

[10]Bengio, Y., Courville, A., & Vincent, P. (2013). *IEEE TPAMI.*

[11]Raghu, M., et al. (2017). *ICML.*

where p_k is the area-weighted proportion of class k inside the segment. The entropy concept originates from information theory and has long been used to characterise landscape heterogeneity.[12] [13]

Configuration is captured using relative mutual information:

$$U = \frac{I}{H}$$

with

$$I = \sum_{i,j} p(i,j) \, \log\left(\frac{p(i,j)}{p(i)\,p(j)}\right)$$

Here $p(i,j)$ is the joint probability of adjacent class pairs inside the segment. Similar information-based pattern measures are well established in landscape ecology.[14] [15]

The important point is not the numerical value itself but the diagnostic role: H and U jointly express whether a segment is internally homogeneous, mixed, or spatially organised. Their values depend on segmentation scale and raster resolution; this is an explicit feedback between S3 and S4 rather than an artefact.

1.4.2 Physiographic embedding

Physiographic metrics attach terrain-based context to each segment: mean elevation, mean slope, mean aspect and derived southness, optionally complemented by windwardness.

Southness is derived from aspect θ as:

$$\text{southness} = \frac{\cos(\theta - \pi) + 1}{2}$$

This mapping is commonly used to linearise aspect into an interpretable radiation-exposure proxy.[16]

Physiographic metrics provide slow-varying structural constraints and anchor each segment in long-term geomorphological context.[17]

---

[12]Shannon, C. E. (1948). *Bell System Technical Journal.*

[13]McGarigal, K., Cushman, S. A., & Ene, E. (2012). *FRAGSTATS v4.*

[14]Li, H., & Reynolds, J. F. (1993). *Landscape Ecology.*

[15]Cushman, S. A., & McGarigal, K. (2002). *Landscape Ecology.*

[16]Wilson, J. P., & Gallant, J. C. (2000). *Terrain Analysis.*

[17]Wilson, J. P., & Gallant, J. C. (2000). *Terrain Analysis.*

1.4.3 Hydrological embedding

Hydrological metrics aggregate watershed identifiers, Strahler order, flow accumulation and proximity to streams.

These quantities encode topological embedding and connectivity rather than hydrodynamic behaviour, following classical geomorphological interpretations of drainage structure.[18] [19]

1.4.4 Biostructural embedding

Biostructure metrics aggregate canopy structure fields derived from surface and terrain models: mean canopy height, upper height percentiles, height variability and canopy fraction.

These metrics follow established practice in lidar-based vegetation structure analysis and deliberately avoid species-level interpretation.[20] [21]

1.5 From Metrics to Tokens: Controlled Symbolic Compression

Several continuous metrics are discretised into three robust bins using global tertiles and encoded as symbolic tokens.

Symbolic compression supports readability, filtering and rule construction, but must remain layered on top of quantitative representations rather than replacing them.[22]

Tokens are therefore pragmatic abstractions, not learned regimes.

1.6 The Attribute Stack][as the S4 Product

All metric tables are joined onto the stable segment geometry to form the attribute stack. Geometry is inherited exclusively from the segmentation backbone. No geometries are merged or altered.

The attribute stack is the canonical S4 artefact: a spatial table in which each segment carries its full signature vector and optional tokens.

---

[18]Strahler, A. N. (1957). *Transactions of the American Geophysical Union.*
[19]Moore, I. D., Grayson, R. B., & Ladson, A. R. (1991). *Hydrological Processes.*
[20]Lefsky, M. A. et al. (2002). *BioScience.*
[21]Dubayah, R., & Drake, J. (2000). *Journal of Forestry.*
[22]Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Elements of Statistical Learning.*

1.7 The Boundary Between S4 and S4L

S4 ends with explicit, human-interpretable signatures. At this point, the system has not yet constructed types or regimes.

S4L introduces learned representations: embeddings, latent regimes or similarity structures derived from the S4 attribute space using machine-learning models.[23] [24]

This separation avoids the failure mode where representation construction and decision logic become entangled in a single opaque pipeline.[25]

Random Forest models operate on the S4 attribute space but are documented separately and do not belong to S4 conceptually.

1.8 Why This Layer Matters Architecturally

S4 establishes a stable representational interface between geometry and intelligence. It enables explicit reasoning about structure, transparent auditing of assumptions and controlled extension toward learned representations.

The transformation chain becomes:

Observations $\rightarrow$ Derived Fields $\rightarrow$ Segments $\rightarrow$ Signatures (S4) $\rightarrow$ Learned Representations (S4L) $\rightarrow$ Decisions

S4 is therefore not a technical convenience. It is the epistemic stabiliser of the entire architecture.

[23]Bengio, Y., Courville, A., & Vincent, P. (2013). *IEEE TPAMI.*
[24]Raghu, M., et al. (2017). *ICML.*
[25]Lipton, Z. C. (2018). *Communications of the ACM.*