BENCHEWS · SCIENTIFIC MONOGRAPH

BenchEWS – Measuring Self-Correction in Complex Systems

How complex systems lose the ability to respond to warning signals — and whether, and under which conditions, that loss can be measured

BOOK-v1.0-6bc75901

2026-11-09T19:24:41Z

Dipl.-Ing. Bernd von Mallinckrodt Publication Edition v1.0 27 September 2026

Edition and status note

This Publication Edition v1.0 is the editorially frozen English edition of the book. It has been translated and consolidated from the complete, hashable German Master v0.2 after a multi-stage adversarial audit. Open scientific questions are not hidden; they are marked as HYPOTHESIS, PROSPECTIVE, AUDIT FINDING or UNRESOLVED wherever the evidence does not support a stronger claim.

Publication of this book therefore does not claim completed scientific validation of every BenchEWS component. It publishes the present state of an open research programme in a citable and revision-safe form.

Epistemic classes: DERIVED, EMPIRICAL, FRAMEWORK, HYPOTHESIS, IMPLEMENTED, PROSPECTIVE, ANALOGY, AUDIT FINDING, UNRESOLVED.

Principal scientific limits of this edition

- The peer-reviewed mathematical core concerns spectral compression of the stationary population covariance under A1–A4.
- Φ(Σ) is not a validated proxy for the corrective option space D.
- CRTI is treated as an open research line, not as a universally validated index.
- The exact source-to-release formula binding of the current Studio 2.0 CRTI outputs remains open as SRC-004.
- Cross-domain transfers are functional or metrological hypotheses unless a genuine transfer of mechanism has been demonstrated.
- Recovery, Adaptive Reopening and Regeneration are kept distinct and treated according to their respective evidential status.

Contents

1. Chapter 1 – Systems can function while losing corrective capacity
1. Chapter 2 – Warning signals, thresholds and limits of observation
1. Chapter 3 – Self-correction as a measurable system property
1. Chapter 4 – Early Warning Signals: what can already be measured
1. Chapter 5 – Spectral compression and effective dimensionality
1. Chapter 6 – CRTI: genealogy of an open diagnostic concept
1. Chapter 7 – Feedback Channel Quality and the quality of corrective information
1. Chapter 8 – BenchEWS as a measurement architecture
1. Chapter 9 – Evidence, falsification, competing explanations and abstention
1. Chapter 10 – BenchEWS Studio: research as a reproducible instrument
1. Chapter 11 – Cross-domain transfer: what BenchEWS may claim — and what it may not
1. Chapter 12 – From warning to the restoration of self-correction

Chapter 1 – Systems can function while losing corrective capacity

Complex systems can appear externally functional for a long time even while their capacity for internal correction is already declining. Production continues, organisational routines still work, physiological regulation maintains a state, or a technical network appears stable. Visible function and remaining corrective capacity are therefore not identical.

The guiding question of this book is not whether a system will eventually collapse. It is narrower and, at the same time, more demanding: can a system continue to function while already losing the ability to detect its own deviations effectively, process relevant information, and realise suitable responses? And if so, can that loss be measured under explicit conditions?

BenchEWS treats this question as a metrological research programme. It does not claim to possess a universal theory of complex systems. Instead, it separates the functional system under study from the observer’s measurement architecture. A system can appear stable without being adaptive. Conversely, a system can be temporarily volatile while retaining a broad and effective range of corrective possibilities.

Self-correction is understood here functionally. At least three operations must be kept distinct: a relevant deviation must be detectable; correction-relevant information must be transmitted and integrated effectively within the system; and genuinely usable response options must exist. This minimal structure is not yet a theory of specific mechanisms. It is an ordering of measurement questions.

A central risk lies in confusing state with capacity. Current output can be good even though the space of responses has already narrowed. A system can successfully compensate for one disturbance and still be less capable when the next disturbance arrives. The question of self-correction is therefore stronger than the question of momentary performance.

The normative boundary is equally important. More self-correction is not automatically morally better. A harmful system can itself be highly effective at internal self-correction. Any evaluation therefore requires a functional or viability-related reference frame. BenchEWS does not measure moral goodness.

The book consequently follows a strict epistemic rule: mathematically derived results, established methods, implemented software, empirical findings, frameworks, hypotheses and analogies are kept separate. An implementation does not make a hypothesis true. A DOI does not make a claim independently confirmed. A mathematical theorem under narrow assumptions does not become a universal law of nature.

The objective therefore shifts from collapse rhetoric to a measurement question. The spectacular endpoint is not the primary object; the possible change in corrective architecture beforehand is. That requires observation models, baselines, uncertainty, falsification, and the ability to withhold diagnosis when the evidence is insufficient.

Chapter 2 – Warning signals, thresholds and limits of observation

A warning signal is never merely a property of the system. It emerges from a chain comprising system dynamics, observation operator, sampling, preprocessing, statistics and interpretation. The same latent change can therefore appear visible, invisible or even inverted depending on the measurement channel.

Formally, one may write

$y(t) = H[x(t)] + \eta(t)$,

where $x(t)$ denotes the latent system state, H the observation operator and η measurement or observation noise. Even this simple equation shows that a statistic computed from $y(t)$ is not automatically a direct statement about $x(t)$.

A useful Early Warning Signal would need to combine at least three properties: it should detect a relevant change, provide sufficient lead time, and offer enough specificity to limit false alarms. These objectives often conflict. A sensitive indicator can react early and still be nonspecific.

Critical Slowing Down is a classic example. In certain system classes, the rate of return to a stable state can decrease as a critical threshold is approached. Variance and lag-1 autocorrelation may then increase. These signatures are not universal, however. Trend, filtering, sampling, heteroscedastic noise, external forcing or changes in the observation regime can create similar patterns.

Thresholds therefore do not have a universal status. A statistical threshold, a physical limit, a regulatory boundary and a bifurcation threshold are different things. Collapsing them linguistically creates false precision.

More data do not automatically solve this problem. If the observation operator does not capture relevant directions of freedom at all, a longer measurement period cannot create the missing information. Strong aggregation can likewise smooth real change. Observability is therefore itself a property of the measurement situation.

BenchEWS separates five levels: latent construct, proxy, measurement model, causal model and diagnostic classifier. A trend can be clean at proxy level while remaining ambiguous at causal level. For that reason, abstention is a legitimate outcome.

The central consequence is simple: Early Warning is not a countdown. An indicator can show that a measured structure is changing. Whether that means an approaching transition, a changed noise structure or a measurement artefact is a separate question.

Chapter 3 – Self-correction as a measurable system property

Self-correction is not a directly observable quantity. It is a latent construct. BenchEWS therefore proposes a diagnostic profile:

$X_t = [O_t \mid F_t \mid D_t \mid R_t \mid V_t]$.

O denotes observation or access to information. F describes the processing and transmission of already available correction-relevant information. D denotes the endogenous corrective option space. R concerns the response that is actually realised. V concerns the evidence that a response was functionally successful.

These five dimensions are a diagnostic decomposition, not a claimed fundamental ontology. Nor are they automatically statistically orthogonal. O and F in particular can be empirically dependent because state estimation is built on observed information. F and R can likewise be coupled. The profile therefore separates questions; it does not assert independent random variables.

The boundary to established concepts is important. Self-correction is not identical to homeostasis, regulation, adaptation, learning, resilience or controllability. Those concepts can describe parts of the problem. BenchEWS does not attempt to rename them, but to place them within a metrological question concerning corrective capacity.

The same measurement chain applies to each dimension: first the latent construct must be defined. Then an observable, a measurement model, uncertainty information, a reference standard, benchmarks and finally a bounded inference are required. Without this chain, a numerical value remains semantically underdetermined.

D is particularly difficult. An external observer can recognise an action as possible even though the system itself has no access to it because of missing information, resources or internal representation. The relevant option space is therefore endogenous and information-conditioned. This connects BenchEWS to established reachability, controllability and viability theory without claiming that mathematics as new.

R and V must also be separated. An action can be executed and still be ineffective. Conversely, a positive output can occur without the intended intervention being causally responsible. Validation therefore requires its own evidence.

At present the profile must not be aggregated into a global score. There is neither a generally justified common scale nor validated weights. An average could conceal functionally very different deficits.

A clear boundary to the mathematical core in Chapter 5 is also essential: spectral compression of a covariance matrix is not identical to the loss of corrective options D. There is currently no validated bridge that would justify $\Phi(\Sigma)$ as a proxy for D. Any such link is a separate hypothesis and would have to be tested against an independently measured D.

Chapter 4 – Early Warning Signals: what can already be measured

BenchEWS does not begin from zero. The Early Warning Signals literature provides established statistics and theoretical mechanisms that must serve as baselines.

Variance measures dispersion around the mean. In some Critical-Slowing-Down scenarios it can rise before an instability. It is, however, nonspecific: changing forcing, heteroscedastic noise, mixture distributions or preprocessing can produce the same effect.

Lag-1 autocorrelation measures temporal persistence between adjacent observations. For an AR(1) process

$$x_{t+1} = a\, x_t + \varepsilon_t$$

a approaches 1 in certain CSD scenarios. Sampling frequency, trend, filtering and nonstationarity can strongly affect the estimate.

Skewness can capture distributional asymmetry but is sensitive to outliers, small samples and mixture distributions. Kendall’s τ can quantify trends but by itself says nothing about the underlying mechanism.

Surrogate methods can provide null models. Their meaning depends entirely on which properties the surrogates preserve. A p-value against one specific null model is not proof of a mechanism.

For classification tasks, ROC curves, TPR, FPR and AUC can be useful. The same caveat applies: AUC measures discrimination, not calibration and not automatically practical usefulness. Lead time and false-alarm rate must be assessed together.

In multivariate systems, the covariance matrix becomes a natural object. Its eigenvalues and eigenvectors describe how shared variation is distributed. This structure is not simply “better variance”; it contains different geometric information.

BenchEWS Studio 2.0 documents four scientific modules: variance, lag-1 autocorrelation, skewness, and an experimental multivariate CRTI module. The first three are established statistics; CRTI has a considerably more open evidential status.

The Studio Reference Benchmark is a technical pipeline test. A deterministic reference value shows that a code path can be reproduced. It does not validate a tipping mechanism.

Any new BenchEWS metric must therefore compete against these baselines. Meaningful incremental value could consist of earlier warning, a lower false-alarm rate, greater robustness, better mechanism differentiation or improved scale invariance. Whether such incremental value exists is an empirical question and cannot be inferred from a mathematical derivation alone.

Chapter 5 – Spectral compression and effective dimensionality

The mathematically strongest original component of the programme is a peer-reviewed result concerning the stationary covariance of a linear stochastic system. Consider

$dx_t = A\,x_t\,dt + B\,dW_t,$

with $Q = BB^*$ and a stable drift matrix A. The stationary covariance Σ satisfies the Lyapunov equation

$A\Sigma + \Sigma A^* + Q = 0$

and equivalently

$\Sigma = \int_0^\infty e^{At}\,Q\,e^{A^*t}\,dt.$

Let the dominant eigenvalue of the drift matrix be $\lambda_1 = -\varepsilon$ with $\varepsilon \to 0+$, while the remaining modes stay stable behind a fixed spectral gap. Under assumptions A1–A4 of the published theorem,

$\lambda_1(\Sigma) = (v_1^*Qv_1)/(2\varepsilon) + O(\varepsilon^{-1/2}),$

while the remaining covariance eigenvalues remain $O(1)$. A1 requires an algebraically simple, isolated dominant mode; A2 sufficient excitation of that mode by the noise; A3 bounded non-normality or eigenvector conditioning; and A4 a bounded stable residual contribution.

From the normalised covariance eigenvalues

$p_i = \lambda_i(\Sigma) / \mathrm{tr}(\Sigma)$

one defines the spectral entropy

$H = -\Sigma_i p_i \log p_i$

and the effective rank

$\Phi(\Sigma) = \exp(H)$.

For a uniformly distributed spectrum, Φ is close to d; when the spectrum concentrates in one direction, Φ approaches 1. Under A1–A4,

$p_1 \to 1$, $H \to 0$, $\Phi(\Sigma) \to 1$.

This result describes the geometry of stationary covariance. It does not mean that the physical state space literally becomes one-dimensional. Nor does it mean that corrective options D disappear. The theorem is neither a general collapse detector nor a universal definition of CRTI.

The limitations matter: Hopf-like multimode cases, defective eigenstructures, strong non-normality, failure to excite the critical mode, high-dimensional bulk spectra and external shocks lie outside or at the boundary of the simple picture.

A newer audit finding concerns defective Jordan structures. Independent reproduction for simple Jordan examples confirms that the divergence rate can become faster than $O(\varepsilon^{-1})$. This lies outside A1 and therefore does not affect the core proof. It is retained here as an AUDIT FINDING; no general theorem for defective cases is frozen in this edition.

A second limitation concerns finite windows. Φ(Σ) is a population statement. A rolling-window estimator Φ_W can saturate or be strongly biased upward precisely when the relaxation time $1/\varepsilon$ exceeds the window length. Qualitatively, the relevant dimensionless relation is εW. This estimator problem is not solved by the population theorem and requires its own theory and simulation.

Standardisation also changes the measurement object. Φ computed on a covariance matrix in fixed units is not generally equivalent to Φ computed on a correlation matrix after z-standardisation. Audit calculations show that the correlation transform can weaken the asymptotic compression or prevent it when the dominant mode contains zero components. Every application must therefore document explicitly whether covariance or correlation is analysed.

Within the idealised A1–A4 regime, variance of the dominant mode, its autocorrelation and Φ ultimately carry information about the same stability margin ε. Any incremental diagnostic value of Φ over classical CSD baselines is therefore an empirical hypothesis, not a theorem.

The scientifically admissible statement is consequently narrow: under A1–A4, the stationary population covariance asymptotically concentrates onto one dominant mode. Everything further — finite windows, cross-domain transfer, a link to D, or early-warning utility — requires additional evidence.

Chapter 6 – CRTI: genealogy of an open diagnostic concept

CRTI does not have a single, linear, fully harmonised history. Rather than smoothing over that fact, this chapter makes it part of the scientific record.

Historically, CRTI was introduced as the Compression–Response Transition Index. An early form combined a compression quantity Φ and a relaxation quantity R in the scalar

$T = R / \Phi$.

This form was later abandoned. Historical manuscripts moved toward a two-channel fragility signature: decreasing structural dimensionality and increasing relaxation time were to be evaluated jointly and tested against surrogates.

Yet even the historical definitions are not uniform. One manuscript defined an observable Φ_{obs} as the effective rank of a whitened lag-covariance operator

$K_\tau = C_0^{-1/2} C_\tau C_0^{-1/2}$.

For linear processes, however, $C_\tau = F^\tau C_0$; K_τ is therefore similar to F^τ. Its eigenvalues depend on the propagator rather than on the stationary covariance concentration. Independent audit calculation confirms this independence in the linear case. Historical Φ_{obs} therefore does not measure the same object as $\Phi(\Sigma)$ in Chapter 5.

R_{obs} also diverged between the manuscript and the historical code. The manuscript used an AR(1)-based relaxation time, schematically

$R_{obs} = -1 / \log(\hat{\rho})$,

whereas a code appendix computed a norm-based quantity

$r = 1 - ||C_\tau||_2 / ||\Sigma_{reg}||_2$.

These quantities differ not only in scale; they can move in opposite directions under slowing dynamics.

In addition, the historical surrogate logic was not fully reproduced in code; earlier audits also documented very high false-alarm rates and historical AUC claims that could not be reproduced. These negative findings are deliberately retained. Historical CRTI artefacts are therefore development history, not the current canonical specification.

Studio 2.0 documents an experimental multivariate module whose outputs are labelled Phi_obs and R_obs. Runtime evidence shows that numerical values are produced. The authoritative formula binding of those current outputs remains SRC-004: UNRESOLVED. A companion document also describes the spectral quantity at one point in terms resembling the leading share p_1, whereas the peer-reviewed quantity Φ is an effective rank. That passage must not be treated as a definition.

CRTI is therefore not presented in this book as a finished index. Working definition: CRTI denotes a research line testing whether spectral concentration and an independent relaxation- or recovery-related quantity jointly provide additional diagnostic value.

Binding symbol separation:

$\Phi(\Sigma)$: effective rank of stationary population covariance, DERIVED under A1–A4. p_1: leading covariance share $\lambda_1/tr\Sigma$, not identical to Φ. $\hat{\Phi}_W$: finite-window estimator of Φ, still requiring statistical validation. Φ_{obs}^{H1}: historical whitened-lag quantity, DEFINED/HISTORICAL. φ^{H2}: historical code quantity, IMPLEMENTED/HISTORICAL. Phi_{obs}^{S2}: Studio 2.0 output, IMPLEMENTED/RUNTIME EVIDENCE, FORMULA UNRESOLVED. R_{obs}^{H1}: historical AR(1)-based relaxation time. r^{H2}: historical norm-based code quantity. R_{obs}^{S2}: Studio 2.0 output, FORMULA UNRESOLVED. $T = R/\Phi$: historical scalar CRTI proposal, ABANDONED.

The status is therefore clear: the spectral-compression theorem is peer-reviewed. CRTI as a combined compression-recovery architecture has not yet been established as a single validated measurand.

Chapter 7 – Feedback Channel Quality and the quality of corrective information

A system can detect a deviation and still fail to respond effectively. Between observation and action lie estimation, selection, transmission and execution. Feedback Channel Quality examines parts of this information chain.

In the current BenchEWS framework, F includes in particular State Estimation Q_E, Gating Q_G, a directed decision bias b_M and Command Transmission Q_CT. These components are grounded predominantly in established estimation, decision and communication theory. Their mathematical foundations are not claimed as new.

The boundary between O and F is critical. O asks what relevant information is available at all. F asks what happens to already available information inside the system. State estimation, however, necessarily depends on O. F is therefore not a statistically independent quantity. The framework must explicitly condition on observation quality and must not count the same evidential deficit twice.

Q_E describes how effectively available observational information is transformed into an internal state representation. A perfect estimator cannot create information that is absent from the observation channel. This implies an information ceiling.

Q_G concerns the logic of selection and forwarding. Thresholds encode error costs. A directional bias is therefore not automatically a defect; it can be rational when misses and false alarms have different consequences.

Q_CT asks whether a selected command or corrective information reaches the executing part of the system with sufficient fidelity and in time. Decision and execution remain distinct.

Earlier FCQ versions contained broader components; later audits moved Sensing into O, Execution Outcome into R, and Response/Refinement more toward V. This refactoring matters: a framework gains strength not by accumulating terms but by improving conceptual separation.

FCQ is not a global quality score. A system can have good state estimation, poor gating and good transmission. An average would obscure the bottleneck.

Nor is causal diagnosis automatically identifiable. From a missing output alone one cannot infer whether observation, gating, transmission or execution failed. Bottleneck diagnosis should therefore be posterior-based and abstain when ambiguity remains.

A future FCQ test would need synthetic systems with deliberately manipulated stages, posterior calibration tests, false-discovery measurement in intact systems, and robustness tests under model misspecification. The present status remains: FRAMEWORK / CANDIDATE DIAGNOSTIC ARCHITECTURE, not an empirically validated universal metric.

Chapter 8 – BenchEWS as a measurement architecture

BenchEWS is not a single metric. It is a measurement and evidence architecture. Two orders must be separated: the functional architecture of the system under study and the observer's measurement architecture.

The metrological chain is:

Latent Capacity → Observable → Measurement Model → Measurement Uncertainty → Reference Standard → Benchmark → Inference.

This chain prevents a calculated number from being reported directly as a latent system property.

The O|F|D|R|V diagnostic profile runs across that chain. The same metrological sequence must be followed for each dimension. BenchEWS can therefore be understood as a matrix: functional dimensions horizontally, evidential stages vertically.

A global score is deliberately avoided. There is no validated common scale, no justified compensation logic and no empirically supported weighting scheme. A profile is diagnostically more informative than a rating.

Observation quality has a gating function: if O is insufficient, downstream causal interpretations can be blocked. Likewise, an unstable estimator can block interpretation even though a numerical value remains computable. Computable is not the same as interpretable.

Evidence states should be explicit: MEASURED, ESTIMATED, INTERPRETABLE, PROSPECTIVE. Implemented, Validated and Release Authorized must likewise remain distinct.

Provenance is not an administrative detail. Data, preprocessing, algorithm, parameters, software version, reference and interpretation rule must be traceable. CRTI illustrates what happens when one label is reused across multiple definitions.

Modularity is epistemic design. If an experimental module fails, established statistics or other parts of the framework must not automatically fail with it. Conversely, a passed software test must not upgrade the epistemic status of other layers.

BenchEWS is therefore best understood as an Evidence Interface: it should make visible what was measured, how it was measured, which interpretation is admissible, what uncertainty remains, and when no statement is justified.

Chapter 9 – Evidence, falsification, competing explanations and abstention

A scientific early-warning system is not good because it warns often. It is good when it is clear when a warning is justified and when it should not be issued.

BenchEWS separates falsifiability of the measurement architecture from falsifiability of specific diagnostic claims. A construct can be measurable and still have no Early Warning utility. A classifier can perform well and still measure the wrong construct.

Central kill conditions include: lack of reproducible estimability; collapse onto established quantities without incremental value; no incremental diagnostic gain; no genuine lead time; an

unacceptable false-alarm rate; lack of construct validity; lack of reproducibility; failed cross-domain transfer; recovery that cannot be distinguished from artefact; or uncertainty that overwhelms the signal completely.

Before a confirmatory study, these conditions require quantitative criteria. Otherwise thresholds can be moved after the fact. A current weakness of the programme is that several kill conditions still lack domain-specific preregistered limits. That does not refute the framework, but it remains an open operationalisation task.

Competing explanations must be retained. For spectral concentration, possibilities include approaching instability, changed noise structure, observation drift, external shocks, non-normality or high-dimensional bulk effects. A plausible explanation is not proof.

Positive and negative controls are necessary, as are adversarial benchmarks. Particularly important are cases designed to trigger known weaknesses: nearly degenerate eigenvalues, strong non-normality, changing sampling rate, sensor drift, multiple critical modes, or tipping without a classical CSD signature.

Prospective validation is stronger than retrospective pattern recognition. Base rates must be considered: P(Signal|Transition) is not P(Transition|Signal). Calibration and discrimination are different properties.

Abstention is itself a scientific output. DATA_INSUFFICIENT, MODEL_NOT_APPLICABLE, NON_IDENTIFIABLE, UNCERTAINTY_TOO_HIGH, CONTRADICTORY_EVIDENCE and NOT_VALIDATED_FOR_DOMAIN are distinct reasons not to issue a diagnosis.

A Claim Ladder helps prevent overextension: Observation → Statistical Interpretation → Model-Based Interpretation → Diagnostic Interpretation → Operational Claim. Each step requires additional evidence.

The minimal rule is: never claim more than the weakest necessary evidential stage can support.

Chapter 10 – BenchEWS Studio: research as a reproducible instrument

BenchEWS Studio is not proof of the research programme. It is the software and provenance infrastructure intended to make parts of the programme executable and auditable.

The documented workflow separates Reference Benchmark, Idea, Project, Configuration, Validation, Execution, Verification, Interpretation, Publication, Reuse, Object Browser and Scientific Modules. This separation matters scientifically: calculation and interpretation are not the same event.

Configuration Validation checks the formal consistency of an analysis configuration. Scientific Validation is different: it asks whether a method actually measures the scientific object it claims to measure. Release Authorization concerns the concrete software artefact. These three levels must not be collapsed into one global PASS.

A Reference Benchmark can show that installation and pipeline work reproducibly. It cannot show that a statistic is a valid early-warning indicator. Unit, integration, regression and smoke tests reduce classes of software error; they do not replace construct validity.

Reproducible research requires a source-of-truth chain: repository, commit, tag, release artefact, checksum, build configuration, dependencies and documentation version.

Numerical reproducibility adds a second rule: the measurement rule must exist before the measurement is judged. BenchEWS therefore uses versioned Tolerance Contracts for non-exact cross-platform parity comparisons. A contract defines the admissible numerical discrepancy, is frozen before evaluation, and is cryptographically bound to the decision. If it is missing, still PENDING, or modified, the comparison cannot produce PASS. Widening a tolerance after observing the result would be post-hoc tuning; a scientifically justified change must instead create a new contract version. For CRTI between macOS/Desktop and Linux/Web, the relevant contract deliberately remains PENDING without a numerical tolerance until frozen Mac reference outputs and the associated Python/NumPy/BLAS environment are available. This governance does not change Scientific-Core mathematics; it changes the evidential rule under which reproducibility may be claimed. Runtime screenshots can show that an interface produced a value. They do not show which internal formula was executed.

This is exactly where the open SRC-004 gap remains for CRTI. Studio 2.0 documents runtime outputs, but the authoritative source-to-release binding of the current CRTI formulas has not yet been closed in the book audit.

The audit of the Scientific Release Companion also identified a documentation problem: one passage describes a leading-eigenvalue fraction rising toward 1, whereas the peer-reviewed quantity Φ is an effective rank that falls toward 1. The AIP paper is also labelled too broadly in several places as a “CRTI derivation”. This book does not adopt those formulations; the Companion requires a separate correction or erratum.

Studio should ultimately function as an Evidence Interface: Result, Method, Module Version, Input Adequacy, Method Adequacy, Execution Status, Evidence/Uncertainty, Contradictions, Interpretation and Non-Claim should be visible. An experimental module should not visually suggest the same epistemic status as an established statistic.

The present status of Studio 2.0 is therefore conservative: workflow implemented/documented; classical statistical modules implemented; CRTI runtime documented/experimental; CRTI scientific validation not established; authoritative CRTI formula provenance still open in the book audit.

Chapter 11 – Cross-domain transfer: what BenchEWS may claim — and what it may not

BenchEWS uses terms that sound meaningful across many domains: observation, feedback, response options, compression and recovery. Shared language, however, does not establish shared mechanics.

This book distinguishes five transfer levels: analogy, functional comparability, model mapping, empirical transfer and mechanism transfer. Each successive level requires more evidence.

A Domain Adapter translates between an abstract construct and a domain-specific operationalisation. It must preserve meaning, direction, ordering and uncertainty semantics sufficiently well. If it fails to do so, the transfer claim fails for that dimension.

A technical sensor and an organisational reporting pathway can both functionally concern access to information. Their error structures are nonetheless different. A common O construct may be useful; an identical measurement model is unlikely to be.

The same holds for F, D, R and V. Controllability, reachability and viability are established mathematical fields. BenchEWS does not claim new mathematics merely because those tools are placed inside a common measurement architecture.

The mathematical theorem in Chapter 5 may be transferred only through an explicit model chain:

Theorem → Model Class → Domain Model → Observable → Empirical Test.

Not: Theorem → Real World.

Social, organisational and economic systems require particular caution because intentionality, strategy, Goodhart effects and reflexivity can alter the measurement itself. In such settings, metaphors and functional parallels are legitimate so long as they are clearly labelled as such.

Cross-domain transfer therefore does not mean ranking forests, organisations and power grids on one universal self-correction scale. A more plausible target is process comparability: different domains may use different measurements while following the same evidential discipline.

The strongest current cross-domain claim is therefore this: BenchEWS proposes a common metrological logic of questioning. It does not claim one common mechanism for all complex adaptive systems.

Chapter 12 – From warning to the restoration of self-correction

A research programme on degradation remains incomplete if it does not ask whether lost corrective capacity can be regained. Recovery, however, is not simply loss with the sign reversed.

Adaptive Reopening and Self-Correction Recovery are kept distinct. Adaptive Reopening concerns specifically the possible re-expansion of the corrective option space D. Self-Correction Recovery concerns multidimensional change in the full O|F|D|R|V profile.

An increase in D alone does not constitute complete recovery. A system can have new options but still fail to realise them because feedback is poor or execution is blocked. “Reopened but not realised” is therefore a distinct diagnostic state.

Recovery can be partial, asymmetric and hysteretic. Different dimensions may have different recovery times. A system can arrive at a new functional state without returning to the original one.

Regeneration goes conceptually further. It denotes possible reorganisation through which new feedback pathways, new options or new functional structures emerge. Within the BenchEWS programme, this remains prospective rather than an established validated metric.

Viability Theory provides established mathematics for asking whether admissible paths exist under constraints. A horizon-relative lock-in can therefore be more meaningful than absolute irreversibility. A scientifically cautious statement is: under model M, control set U, resource budget B, uncertainty and horizon T, no viable recovery path was found. That is not the same as saying that recovery is impossible in reality.

Candidates such as a regenerative variety margin or conditional irreversibility belong to the later research frontier. They are not results of Studio 2.0 and are not validated universal indicators.

The decisive next phase of the programme is empirical: Definition Freeze, Ground Truth by Construction, adversarial benchmarks, independent implementations, controllable real systems, cross-domain replication and prospective validation.

The programme must allow its own failure. Self-Correction Capacity may prove impossible to operationalise as a common cross-domain measurand. Individual dimensions may work while the overall profile fails. CRTI may provide no incremental value. Such outcomes would still be scientifically valuable.

The guiding question therefore remains open: how much corrective capacity remains in a system — and under what conditions can its loss or restoration be measured reliably?

Appendix A – Binding CRTI symbol table

Symbol	Meaning	Status
Φ(Σ)	effective rank of stationary population covariance	DERIVED under A1–A4
p_1	$\lambda_1/\mathrm{tr}\Sigma$	defined spectral quantity, not Φ
Φ̂_W	finite-window estimator of Φ	AUDIT FINDING / unvalidated
Φ_obs^H1	historical effective rank of whitened lag operator	HISTORICAL
φ^H2	historical code effective-rank-like quantity	IMPLEMENTED/HISTORICAL
Phi_obs^S2	Studio 2.0 runtime output	IMPLEMENTED; formula UNRESOLVED
R_obs^H1	AR(1)-based relaxation time	HISTORICAL
r^H2	norm-based historical code quantity	IMPLEMENTED/HISTORICAL
R_obs^S2	Studio 2.0 runtime output	formula UNRESOLVED
T=R/Φ	historical scalar CRTI proposal	ABANDONED

Appendix B – Open research and provenance questions

1. SRC-004: authoritative Studio 2.0 CRTI source-to-release binding and current mathematical formula binding.
2. SRC-005: correction of the Scientific Release Companion concerning Φ versus p_1 and the designation of the AIP result.
3. SRC-006: preregistrable quantitative kill conditions for at least one confirmatory synthetic benchmark.
4. AUD-001: mathematical clarification of the defective Jordan edge case outside A1.
5. Expansion of the bibliographic apparatus, particularly for Viability Theory, non-normal dynamics/pseudospectra, and information-theoretic limits of control.

These items form part of the open research and documentation agenda. They are not presented in this edition as already solved and therefore do not block publication of the book as a documented state of the research programme.

Appendix C – Core references

- B. von Mallinckrodt, “Spectral compression of the stationary covariance matrix as a structural precursor to systemic transitions: A Lyapunov–resolvent analysis,” AIP Advances 16, 095208 (2026). DOI 10.1063/5.0337045.
- BenchEWS Research Programme Orientation Guide. DOI 10.5281/zenodo.21611529.
- Feedback Channel Quality paper. DOI 10.5281/zenodo.21705727.
- Structural Compression & Narrative Persistence. DOI 10.5281/zenodo.21616059.
- BenchEWS Studio software all-versions DOI 10.5281/zenodo.21807143.