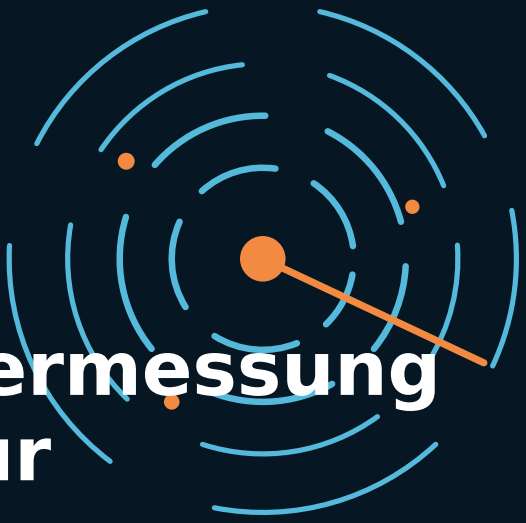
BENCHEWS · SCIENTIFIC MONOGRAPH



BenchEWS – Die Vermessung der Selbstkorrektur

Wie komplexe Systeme ihre Fähigkeit verlieren, auf Warnsignale zu reagieren – und ob und unter welchen Bedingungen sich dieser Verlust messen lässt

BOOK-v1.0-6bc75901

2026-10-07T23:24:25Z

Dipl.-Ing. Bernd von Mallinckrodt Publikationsausgabe v1.0

1. September 2026

Editions- und Statushinweis

Diese Publikationsausgabe v1.0 ist die redaktionell eingefrorene deutsche Fassung des Buches. Sie wurde aus der vollständigen, hashbaren DE-Masterfassung v0.2 nach einem mehrstufigen adversarialen Audit konsolidiert. Offene wissenschaftliche Fragen werden nicht verborgen, sondern dort als HYPOTHESIS, PROSPECTIVE, AUDIT FINDING oder UNRESOLVED gekennzeichnet, wo die Evidenz noch nicht für einen stärkeren Claim ausreicht.

Die Veröffentlichung dieses Buches ist daher kein Anspruch auf abgeschlossene wissenschaftliche Validierung aller BenchEWS-Komponenten. Sie veröffentlicht den gegenwärtigen Stand eines offenen Forschungsprogramms in einer zitier- und revisionssicheren Form.

Epistemische Klassen: DERIVED, EMPIRICAL, FRAMEWORK, HYPOTHESIS, IMPLEMENTED, PROSPECTIVE, ANALOGY, AUDIT FINDING, UNRESOLVED.

Wesentliche wissenschaftliche Grenzen dieser Ausgabe

- Der peer-reviewte mathematische Kern betrifft spektrale Kompression der stationären Populationskovarianz unter A1–A4.
- Φ(Σ) ist kein validierter Proxy für den korrektiven Optionsraum D.
- CRTI wird als offene Forschungslinie und nicht als universell validierter Index behandelt.
- Die exakte Source-to-Release-Formelbindung der aktuellen Studio-2.0-CRTI-Ausgaben bleibt als SRC-004 offen.
- Cross-Domain-Übertragungen sind funktionale oder metrologische Hypothesen, solange kein eigener Mechanismustransfer nachgewiesen ist.
- Recovery, Adaptive Reopening und Regeneration werden getrennt und ihrem jeweiligen Evidenzstatus entsprechend behandelt.

Inhalt

1. Kapitel 1 – Systeme können funktionieren und zugleich Korrekturfähigkeit verlieren
1. Kapitel 2 – Warnsignale, Schwellen und Grenzen der Beobachtung
1. Kapitel 3 – Selbstkorrektur als messbare Systemeigenschaft
1. Kapitel 4 – Early Warning Signals: Was heute bereits messbar ist
1. Kapitel 5 – Spektrale Kompression und effektive Dimensionalität
1. Kapitel 6 – CRTI: Genealogie eines offenen Diagnosekonzepts
1. Kapitel 7 – Feedback Channel Quality und die Qualität korrektiver Information
1. Kapitel 8 – BenchEWS als Messarchitektur
1. Kapitel 9 – Evidenz, Falsifikation, konkurrierende Erklärungen und Abstention
1. Kapitel 10 – BenchEWS Studio: Forschung als reproduzierbares Instrument
1. Kapitel 11 – Domänenübertragung: Was BenchEWS beanspruchen darf – und was nicht

1. Kapitel 12 – Von der Warnung zur Wiedergewinnung von Selbstkorrektur

Kapitel 1 – Systeme können funktionieren und zugleich Korrekturfähigkeit verlieren

Komplexe Systeme können nach außen lange funktionsfähig erscheinen, obwohl ihre Fähigkeit zur inneren Korrektur bereits abnimmt. Produktion läuft weiter, organisatorische Routinen funktionieren, physiologische Regulation hält einen Zustand aufrecht oder ein technisches Netz bleibt scheinbar stabil. Sichtbare Funktion und verbleibende Korrekturfähigkeit sind deshalb nicht identisch.

Die Leitfrage dieses Buches lautet nicht, ob ein System irgendwann kollabiert. Sie lautet enger und zugleich anspruchsvoller: Kann ein System noch funktionieren und zugleich bereits die Fähigkeit verlieren, eigene Abweichungen wirksam zu erkennen, relevante Information zu verarbeiten und geeignete Reaktionen zu realisieren? Und falls ja: Lässt sich dieser Verlust unter expliziten Bedingungen messen?

BenchEWS behandelt diese Frage als metrologisches Forschungsprogramm. Es behauptet nicht, eine universelle Theorie komplexer Systeme zu besitzen. Vielmehr trennt es den untersuchten Funktionszusammenhang von der Messarchitektur des Beobachters. Ein System kann stabil wirken, ohne adaptiv zu sein. Umgekehrt kann ein System zeitweise volatil erscheinen und dennoch über breite, wirksame Korrekturmöglichkeiten verfügen.

Selbstkorrektur wird hier funktional verstanden. Mindestens drei Operationen müssen auseinandergehalten werden: Eine relevante Abweichung muss erkennbar sein; korrektivrelevante Information muss im System wirksam weitergeleitet und integriert werden; und es müssen tatsächlich nutzbare Reaktionsmöglichkeiten existieren. Diese Minimalstruktur ist noch keine Theorie über konkrete Mechanismen. Sie ist eine Ordnung von Messfragen.

Ein zentrales Risiko besteht in der Verwechslung von Zustand und Fähigkeit. Ein aktueller Output kann gut sein, obwohl die Reaktionsmöglichkeiten bereits geschrumpft sind. Ein System kann eine einzelne Störung erfolgreich kompensieren und dennoch bei der nächsten Störung schlechter aufgestellt sein. Daher ist die Frage nach Selbstkorrektur stärker als die Frage nach momentaner Performance.

Ebenso wichtig ist die normative Grenze. Mehr Selbstkorrektur ist nicht automatisch moralisch besser. Auch ein schädliches System kann intern hochgradig selbstkorrigierend sein. Jede Bewertung benötigt daher einen funktionalen oder viabilitätsbezogenen Referenzrahmen. BenchEWS misst keine moralische Güte.

Das Buch folgt daraus einer strengen epistemischen Regel: mathematisch hergeleitete Ergebnisse, etablierte Methoden, implementierte Software, empirische Befunde, Frameworks, Hypothesen und Analogien werden getrennt. Eine Implementierung macht eine Hypothese nicht wahr. Ein DOI macht einen Claim nicht unabhängig bestätigt. Ein mathematischer Satz unter engen Annahmen wird nicht zu einem universellen Naturgesetz.

Damit verschiebt sich das Ziel von einer Kollapsmetaphorik zu einer Messfrage. Nicht der spektakuläre Endpunkt ist der primäre Gegenstand, sondern die mögliche Veränderung der

Korrekturarchitektur davor. Genau dafür braucht es Beobachtungsmodelle, Baselines, Unsicherheit, Falsifikation und die Möglichkeit, bei unzureichender Evidenz keine Diagnose auszugeben.

Kapitel 2 – Warnsignale, Schwellen und Grenzen der Beobachtung

Ein Warnsignal ist niemals nur eine Eigenschaft des Systems. Es entsteht aus einer Kette von Systemdynamik, Beobachtungsoperator, Sampling, Vorverarbeitung, Statistik und Interpretation. Deshalb kann dieselbe latente Veränderung je nach Messkanal sichtbar, unsichtbar oder sogar invertiert erscheinen.

Formal kann man schreiben

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

wobei $x(t)$ den latenten Systemzustand, H den Beobachtungsoperator und η Mess- oder Beobachtungsrauschen bezeichnet. Schon diese einfache Gleichung zeigt: Eine Statistik aus $y(t)$ ist nicht automatisch eine direkte Aussage über $x(t)$.

Ein gutes Early Warning Signal müsste mindestens drei Eigenschaften vereinen: Es müsste eine relevante Veränderung erkennen, hinreichenden zeitlichen Vorlauf bieten und genügend Spezifität besitzen, um Fehlalarme zu begrenzen. Diese Ziele stehen häufig in Spannung. Ein empfindlicher Indikator kann früh reagieren und gleichzeitig unspezifisch sein.

Critical Slowing Down ist ein klassisches Beispiel. In bestimmten Systemklassen kann sich die Rückkehrgeschwindigkeit zu einem stabilen Zustand verringern, wenn eine kritische Schwelle näher rückt. Dann können Varianz und Lag-1-Autokorrelation steigen. Diese Signaturen sind jedoch nicht universell. Trend, Filterung, Sampling, heteroscedastisches Rauschen, externe Forcierung oder Regimewechsel im Beobachtungskanal können ähnliche Muster erzeugen.

Schwellen besitzen daher keinen universellen Status. Ein statistischer Threshold, ein physikalischer Grenzwert, eine regulatorische Grenze und eine Bifurkationsschwelle sind verschiedene Dinge. Wer sie sprachlich zusammenzieht, erzeugt Scheingenauigkeit.

Mehr Daten lösen dieses Problem nicht automatisch. Wenn der Beobachtungsoperator relevante Freiheitsrichtungen gar nicht erfasst, kann zusätzliche Messdauer die fehlende Information nicht erzeugen. Ebenso kann stark aggregierte Beobachtung reale Veränderungen glätten. Observability ist daher selbst eine Eigenschaft der Messsituation.

BenchEWS trennt fünf Ebenen: latentes Konstrukt, Proxy, Messmodell, kausales Modell und diagnostischer Klassifikator. Ein Trend kann auf Proxy-Ebene sauber sein und auf Kausalebene trotzdem mehrdeutig bleiben. Aus diesem Grund ist Abstention ein legitimes Ergebnis.

Die zentrale Konsequenz lautet: Early Warning ist nicht Countdown. Ein Indikator kann zeigen, dass sich eine gemessene Struktur verändert. Ob dies einen Übergang, eine veränderte Rauschstruktur oder einen Messartefakt bedeutet, ist eine zusätzliche Frage.

Kapitel 3 – Selbstkorrektur als messbare Systemeigenschaft

Selbstkorrektur ist kein direkt beobachtbarer Messwert. Sie ist ein latentes Konstrukt. BenchEWS schlägt deshalb ein diagnostisches Profil vor:

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

O steht für Beobachtung beziehungsweise Informationszugang. F beschreibt die Verarbeitung und Weiterleitung bereits verfügbarer korrektivrelevanter Information. D bezeichnet den endogenen korrektiven Optionsraum. R betrifft die tatsächlich realisierte Reaktion. V betrifft die Evidenz darüber, ob eine Reaktion funktional erfolgreich war.

Diese fünf Dimensionen sind eine diagnostische Zerlegung, keine behauptete fundamentale Ontologie. Sie sind außerdem nicht automatisch statistisch orthogonal. Gerade O und F können empirisch abhängig sein, weil Zustandsschätzung auf Beobachtungsinformation aufbaut. Ebenso können F und R gekoppelt sein. Das Profil dient deshalb der Trennung von Fragen, nicht der Behauptung unabhängiger Zufallsvariablen.

Wichtig ist die Abgrenzung zu etablierten Begriffen. Selbstkorrektur ist nicht identisch mit Homöostase, Regulation, Adaptation, Lernen, Resilienz oder Controllability. Diese Konzepte können Teile des Problems beschreiben. BenchEWS versucht nicht, sie umzubenennen, sondern sie in eine metrologische Frage nach Korrekturfähigkeit einzuordnen.

Für jede Dimension gilt dieselbe Messkette: Zuerst muss das latente Konstrukt definiert werden. Dann braucht es eine Observable, ein Messmodell, Unsicherheitsangaben, einen Referenzstandard, Benchmarks und schließlich eine begrenzte Inferenz. Ohne diese Kette bleibt ein numerischer Wert semantisch unterbestimmt.

D ist besonders heikel. Ein externer Beobachter kann eine Handlung als möglich erkennen, die dem System selbst aufgrund fehlender Information, Ressourcen oder interner Repräsentation gar nicht zugänglich ist. Der relevante Optionsraum ist deshalb endogen und informationskonditioniert. Dies verbindet BenchEWS mit etablierter Reachability-, Controllability- und Viability-Theorie, ohne deren Mathematik als neu zu beanspruchen.

Auch R und V müssen getrennt werden. Eine Handlung kann ausgeführt werden und dennoch wirkungslos sein. Umgekehrt kann ein positiver Output eintreten, ohne dass die intendierte Intervention ursächlich verantwortlich war. Validation benötigt deshalb eigene Evidenz.

Das Profil darf derzeit nicht zu einem globalen Score aggregiert werden. Es gibt weder eine allgemein begründete gemeinsame Skala noch validierte Gewichte. Ein Mittelwert könnte funktional völlig unterschiedliche Defizite verdecken.

Entscheidend ist außerdem eine klare Grenze zum mathematischen Kern des späteren Kapitels 5: Spektrale Kompression einer Kovarianzmatrix ist nicht identisch mit dem Verlust korrektiver Optionen D. Derzeit existiert keine validierte Brücke, die $\Phi(\Sigma)$ als Proxy für D rechtfertigt. Jede solche Verbindung ist eine separate Hypothese und müsste mit unabhängig gemessenem D getestet werden.

Kapitel 4 – Early Warning Signals: Was heute bereits messbar ist

BenchEWS beginnt nicht bei null. Die Literatur zu Early Warning Signals stellt etablierte Statistiken und theoretische Mechanismen bereit, die als Baselines dienen müssen.

Varianz misst Streuung um den Mittelwert. In manchen Critical-Slowing-Down-Szenarien kann sie vor einer Instabilität steigen. Sie ist jedoch unspezifisch: veränderte Forcierung, heteroscedastisches Rauschen, Mischverteilungen oder Vorverarbeitung können denselben Effekt erzeugen.

Lag-1-Autokorrelation misst zeitliche Persistenz zwischen benachbarten Beobachtungen. Für einen AR(1)-Prozess

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

nähert sich a in bestimmten CSD-Szenarien dem Wert 1. Doch Samplingfrequenz, Trend, Filterung und Nichtstationarität beeinflussen die Schätzung stark.

Skewness kann Verteilungsasymmetrien erfassen, ist aber empfindlich gegenüber Ausreißern, kleinen Stichproben und Mischverteilungen. Kendall-τ kann Trends quantifizieren, sagt jedoch allein nichts über den zugrunde liegenden Mechanismus.

Surrogatverfahren können Nullmodelle bereitstellen. Ihre Aussage hängt vollständig davon ab, welche Eigenschaften der Surrogate erhalten bleiben. Ein p-Wert gegen ein bestimmtes Nullmodell ist kein Mechanismusbeweis.

Für Klassifikationsaufgaben können ROC-Kurven, TPR, FPR und AUC nützlich sein. Auch hier gilt: AUC misst Diskrimination, nicht Kalibrierung und nicht automatisch praktischen Nutzen. Lead Time und Fehlalarmrate müssen gemeinsam betrachtet werden.

In multivariaten Systemen wird die Kovarianzmatrix zu einem natürlichen Objekt. Ihre Eigenwerte und Eigenrichtungen beschreiben, wie gemeinsame Variation verteilt ist. Diese Struktur ist nicht einfach 'bessere Varianz'; sie liefert andere geometrische Information.

BenchEWS Studio 2.0 dokumentiert vier wissenschaftliche Module: Varianz, Lag-1-Autokorrelation, Skewness und ein experimentelles multivariates CRTI-Modul. Die ersten drei sind etablierte Statistiken; CRTI besitzt einen deutlich offeneren Evidenzstatus.

Der Studio-Reference-Benchmark ist ein technischer Pipeline-Test. Ein deterministischer Referenzwert zeigt, dass ein Codepfad reproduzierbar ausgeführt werden kann. Er validiert keinen Tipping-Mechanismus.

Jede neue BenchEWS-Metrik muss daher gegen diese Baselines antreten. Ein sinnvoller Zusatznutzen könnte frühere Warnung, geringere Fehlalarmrate, bessere Robustheit, bessere Mechanismendifferenzierung oder bessere Skaleninvarianz sein. Ob ein solcher Zusatznutzen existiert, ist empirisch zu prüfen und darf nicht aus einer mathematischen Herleitung allein abgeleitet werden.

Kapitel 5 – Spektrale Kompression und effektive Dimensionalität

Der mathematisch stärkste originäre Baustein des Programms ist ein peer-reviewtes Resultat zur stationären Kovarianz eines linearen stochastischen Systems. Betrachtet wird

dx_t = A x_t dt + B dW_t,

mit Q = BB* und einer stabilen Driftmatrix A. Die stationäre Kovarianz Σ erfüllt die Lyapunov-Gleichung

AΣ + ΣA* + Q = 0

und äquivalent

Σ = ∫_0^∞ e^{At} Q e^{A*t} dt.

Sei der dominante Eigenwert der Driftmatrix λ_1 = -ε mit ε → 0+, während die übrigen Modi durch einen festen Spektralabstand stabil bleiben. Unter den Annahmen A1–A4 des veröffentlichten Satzes gilt asymptotisch

λ_1(Σ) = (v_1*Qv_1)/(2ε) + O(ε^{-1/2}),

während die übrigen Kovarianzeigenwerte O(1) bleiben. A1 fordert einen algebraisch einfachen, isolierten dominanten Modus; A2 hinreichende Anregung dieses Modus durch das Rauschen; A3 beschränkte Nichtnormalität beziehungsweise Konditionierung; A4 einen beschränkten stabilen Residualbeitrag.

Aus den normalisierten Kovarianzeigenwerten

p_i = λ_i(Σ) / tr(Σ)

wird die spektrale Entropie

H = -Σ_i p_i log p_i

und der effektive Rang

Φ(Σ) = exp(H)

definiert. Für ein gleichmäßig verteiltes Spektrum liegt Φ nahe d; bei Konzentration auf eine Richtung nähert sich Φ dem Wert 1. Unter A1–A4 folgt

p_1 → 1, H → 0, Φ(Σ) → 1.

Dieses Resultat beschreibt die Geometrie stationärer Kovarianz. Es bedeutet nicht, dass der physikalische Zustandsraum eindimensional wird. Es bedeutet ebenso wenig, dass korrektive Optionen D verschwinden. Der Satz liefert keinen allgemeinen Kollapsdetektor und keine universelle CRTI-Definition.

Die Grenzen sind wesentlich: Hopf-artige Mehrmodenfälle, defekte Eigenstrukturen, starke Nichtnormalität, fehlende Anregung des kritischen Modus, hochdimensionale Bulk-Spektren und externe Schocks liegen außerhalb oder an den Grenzen des einfachen Bildes.

Ein neuer Auditbefund betrifft defekte Jordan-Strukturen. Eigene Reproduktion bestätigt für einfache Jordan-Beispiele, dass die Divergenzrate schneller als O(ε^{-1}) werden kann. Dieser Befund liegt außerhalb von A1 und berührt daher den Kernbeweis nicht. Er wird hier als AUDIT FINDING geführt; eine allgemeine mathematische Aussage für defekte Fälle ist noch nicht

eingefroren.

Eine zweite Grenze betrifft endliche Fenster. Φ(Σ) ist eine Populationsaussage. Ein Rolling-Window-Schätzer Φ̂_W kann gerade dann sättigen oder stark nach oben verzerrt sein, wenn die Relaxationszeit 1/ε die Fensterlänge übersteigt. Die relevante dimensionslose Relation ist qualitativ εW. Diese Schätzerfrage ist nicht durch den Populationstheorem gelöst und benötigt eigene Theorie und Simulation.

Auch Standardisierung verändert das Messobjekt. Φ auf einer Kovarianzmatrix in festen Einheiten ist nicht generell gleichbedeutend mit Φ auf einer Korrelationsmatrix nach z-Standardisierung. Eigene Auditrechnungen zeigen, dass die Korrelationstransformation die asymptotische Kompression abschwächen oder bei Nullkomponenten des dominanten Modus verhindern kann. Daher muss jede Anwendung explizit dokumentieren, ob Kovarianz oder Korrelation analysiert wird.

Innerhalb des idealisierten A1–A4-Regimes tragen Varianz des dominanten Modus, dessen Autokorrelation und Φ letztlich Information über dieselbe Stabilitätsmarge ε. Ein inkrementeller diagnostischer Nutzen von Φ gegenüber klassischen CSD-Baselines ist daher eine empirische Hypothese, kein Theorem.

Die wissenschaftlich zulässige Aussage lautet somit eng: Unter A1–A4 konzentriert sich die stationäre Populationskovarianz asymptotisch auf einen dominanten Modus. Alles Weitere – endliche Fenster, Cross-Domain-Transfer, Verbindung zu D, Frühwarnnutzen – benötigt zusätzliche Evidenz.

Kapitel 6 – CRTI: Genealogie eines offenen Diagnosekonzepts

CRTI besitzt keine geradlinige, vollständig harmonisierte Geschichte. Genau diese Tatsache wird hier nicht geglättet, sondern zum Gegenstand des Kapitels gemacht.

Historisch wurde CRTI als Compression–Response Transition Index geführt. Eine frühe Form kombinierte eine Kompressionsgröße Φ und eine Relaxationsgröße R in dem Skalar

$T = R / \Phi$.

Diese Form wurde später aufgegeben. Historische Manuskripte gingen zu einer zweikanaligen Fragility Signature über: sinkende strukturelle Dimensionalität und steigende Relaxationszeit sollten gemeinsam ausgewertet und gegen Surrogate getestet werden.

Doch auch die historischen Definitionen sind nicht einheitlich. Ein Manuskript definierte ein beobachtbares Φ_obs als effective rank eines whitened lag-covariance operators

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

Für lineare Prozesse gilt jedoch $C_\tau = F^\tau C_0$; K_τ ist damit ähnlich zu F^τ. Seine Eigenwerte hängen folglich am Propagator, nicht an der stationären Kovarianzkonzentration. Eigene Auditrechnung bestätigt diese Unabhängigkeit im linearen Fall. Das historische Φ_obs misst daher nicht dasselbe Objekt wie Φ(Σ) aus Kapitel 5.

Auch R_obs divergierte zwischen Manuskript und historischem Code. Die Manuskriptfassung verwendete eine AR(1)-basierte Relaxationszeit, sinngemäß

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

während ein Code-Appendix eine normbasierte Größe

r = 1 - ||C_τ||_2 / ||Σ_reg||_2

berechnete. Diese Größen haben nicht nur unterschiedliche Skalen, sondern können bei Verlangsamung in entgegengesetzte Richtungen laufen.

Darüber hinaus war die historische Surrogatlogik nicht vollständig im Code reproduziert; frühere Audits dokumentieren außerdem sehr hohe Fehlalarme und nicht reproduzierbare historische AUC-Angaben. Diese Negativbefunde werden ausdrücklich erhalten. Historische CRTI-Artefakte sind deshalb Entwicklungsgeschichte, nicht aktuelle kanonische Spezifikation.

Studio 2.0 dokumentiert ein experimentelles multivariates Modul mit Ausgaben, die als Phi_obs und R_obs bezeichnet werden. Runtime-Evidenz zeigt, dass numerische Werte erzeugt werden. Die autoritative Formelbindung dieser aktuellen Ausgaben ist jedoch weiterhin SRC-004: UNRESOLVED. Ein Companion-Dokument beschreibt die spektrale Größe zudem an einer Stelle wie den führenden Anteil p_1, während die peer-reviewte Größe Φ ein effective rank ist. Diese Beschreibung darf nicht als Definition übernommen werden.

CRTI wird in diesem Buch deshalb nicht als fertiger Index geführt. Arbeitsdefinition: CRTI bezeichnet eine Forschungslinie, die prüft, ob spektrale Konzentration und eine unabhängige Relaxations- oder Recovery-bezogene Information gemeinsam zusätzlichen diagnostischen Wert liefern.

Verbindliche Symboltrennung:

Φ(Σ): effective rank der stationären Populationskovarianz, DERIVED unter A1–A4. p_1: führender Kovarianzanteil λ_1/trΣ, nicht identisch mit Φ. Φ̂_W: finite-window estimator von Φ, statistisch noch zu validieren. Φ_obs^H1: historische whitened-lag-Größe, DEFINED/HISTORICAL. φ^H2: historische Codegröße, IMPLEMENTED/HISTORICAL. Phi_obs^S2: Studio-2.0-Ausgabe, IMPLEMENTED/RUNTIME EVIDENCE, FORMULA UNRESOLVED. R_obs^H1: historische AR(1)-Relaxationszeit. r^H2: historische normbasierte Codegröße. R_obs^S2: Studio-2.0-Ausgabe, FORMULA UNRESOLVED. T = R/Φ: historische, aufgegebene Skalarform.

Damit ist der Status klar: Der spektrale Kompressionssatz ist peer-reviewed. CRTI als kombinierte Compression-Recovery-Architektur ist noch nicht als einheitliches, validiertes Measurand etabliert.

Kapitel 7 – Feedback Channel Quality und die Qualität korrektiver Information

Ein System kann eine Abweichung erkennen und trotzdem nicht wirksam reagieren. Zwischen Beobachtung und Handlung liegen Schätzung, Auswahl, Weiterleitung und Ausführung. Feedback Channel Quality untersucht Teile dieser Informationskette.

Im aktuellen BenchEWS-Rahmen gehören zu F insbesondere State Estimation Q_E, Gating Q_G, ein gerichteter Entscheidungsbias b_M und Command Transmission Q_CT. Diese Komponenten beruhen überwiegend auf etablierter Schätz-, Entscheidungs- und Kommunikationstheorie. Ihre mathematischen Grundlagen werden nicht als neu beansprucht.

Die Grenze zwischen O und F ist kritisch. O fragt, welche relevante Information überhaupt verfügbar ist. F fragt, was mit bereits verfügbarer Information innerhalb des Systems geschieht. State Estimation hängt jedoch zwangsläufig von O ab. F ist daher keine statistisch unabhängige Größe. Das Framework muss die Konditionierung auf Beobachtungsqualität explizit

berücksichtigen und darf dieselbe Evidenz nicht doppelt zählen.

Q_E beschreibt, wie gut verfügbare Beobachtungsinformation in eine interne Zustandsrepräsentation überführt wird. Ein perfekter Schätzer kann Information, die der Beobachtungskanal nicht enthält, nicht erzeugen. Daraus folgt ein Information Ceiling.

Q_G betrifft die Selektions- und Weiterleitungslogik. Thresholds kodieren Fehlerkosten. Ein gerichteter Bias ist deshalb nicht automatisch ein Defekt; er kann rational sein, wenn Misses und False Alarms unterschiedliche Konsequenzen besitzen.

Q_CT fragt, ob ein ausgewählter Befehl oder eine korrektive Information den ausführenden Teil des Systems mit ausreichender Fidelity und rechtzeitig erreicht. Entscheidung und Ausführung bleiben getrennt.

Frühere FCQ-Versionen enthielten breitere Komponenten; spätere Audits verschoben Sensing nach O, Execution Outcome nach R und Response/Refinement eher nach V. Diese Refaktorierung ist wichtig: Ein Framework gewinnt nicht durch möglichst viele Begriffe, sondern durch Trennschärfe.

FCQ ist kein globaler Qualitätswert. Ein System kann gute Zustandsabschätzung, schlechtes Gating und gute Übertragung besitzen. Ein Mittelwert würde den Bottleneck verdecken.

Auch die kausale Diagnose ist nicht automatisch identifizierbar. Aus einem fehlenden Output allein kann man nicht schließen, ob Beobachtung, Gating, Übertragung oder Ausführung versagt hat. Deshalb sollte eine Bottleneck-Diagnose posteriorbasiert sein und bei Mehrdeutigkeit abstain.

Ein künftiger FCQ-Test müsste synthetische Systeme mit gezielt manipulierten Stufen verwenden, Kalibrierung der Posterioren prüfen, False Discovery in intakten Systemen messen und Robustheit unter Modellmissspezifikation testen. Der gegenwärtige Status bleibt: FRAMEWORK / CANDIDATE DIAGNOSTIC ARCHITECTURE, nicht empirisch validierte Universalmetrik.

Kapitel 8 – BenchEWS als Messarchitektur

BenchEWS ist keine einzelne Kennzahl. Es ist eine Mess- und Evidenzarchitektur. Zwei Ordnungen müssen unterschieden werden: die funktionale Architektur des untersuchten Systems und die Messarchitektur des Beobachters.

Die metrologische Kette lautet:

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

Diese Kette verhindert, dass eine berechnete Zahl unmittelbar als latente Systemeigenschaft ausgegeben wird.

Das Diagnoseprofil O|F|D|R|V steht quer dazu. Für jede Dimension muss dieselbe metrologische Kette durchlaufen werden. BenchEWS kann deshalb als Matrix verstanden werden: horizontal die funktionalen Dimensionen, vertikal die Evidenzstufen.

Ein globaler Score wird bewusst vermieden. Es fehlt eine validierte gemeinsame Skala, eine begründete Kompensationslogik und ein empirisch gestütztes Gewichtungsschema. Ein Profil ist diagnostisch informativer als ein Rating.

Beobachtungsqualität besitzt eine Gate-Funktion: Wenn O unzureichend ist, können nachgelagerte Kausalinterpretationen blockiert werden. Ebenso kann ein instabiler Schätzer eine Interpretation blockieren, obwohl ein Zahlenwert berechenbar bleibt. Computable ist nicht gleich interpretable.

Evidenzstatus sollten explizit sein: MEASURED, ESTIMATED, INTERPRETABLE, PROSPECTIVE. Ebenso müssen Implemented, Validated und Release Authorized auseinandergehalten werden.

Provenienz ist kein Verwaltungsdetail. Daten, Preprocessing, Algorithmus, Parameter, Softwareversion, Referenz und Interpretationsregel müssen nachvollziehbar sein. CRTI zeigt, was geschieht, wenn derselbe Name über mehrere Definitionen hinweg verwendet wird.

Modularität ist epistemisches Design. Wenn ein experimentelles Modul scheitert, dürfen etablierte Statistiken oder andere Frameworkteile nicht automatisch mitfallen. Umgekehrt darf ein bestandener Softwaretest nicht den Status anderer Ebenen aufwerten.

BenchEWS ist daher am besten als Evidence Interface zu verstehen: Es soll sichtbar machen, was gemessen wurde, wie es gemessen wurde, welche Interpretation zulässig ist, welche Unsicherheit bleibt und wann keine Aussage gerechtfertigt ist.

Kapitel 9 – Evidenz, Falsifikation, konkurrierende Erklärungen und Abstention

Ein wissenschaftliches Frühwarnsystem ist nicht gut, weil es häufig warnt. Es ist gut, wenn klar ist, wann eine Warnung gerechtfertigt ist und wann sie nicht ausgegeben werden darf.

BenchEWS unterscheidet die Falsifizierbarkeit der Messarchitektur von der Falsifizierbarkeit konkreter diagnostischer Claims. Ein Konstrukt kann messbar sein und dennoch keinen Early-Warning-Nutzen besitzen. Ein Klassifikator kann gute Performance zeigen und dennoch ein falsches Konstrukt messen.

Zentrale Kill Conditions sind: fehlende reproduzierbare Schätzbarkeit; Kollaps auf etablierte Größen ohne Zusatznutzen; kein inkrementeller diagnostischer Gewinn; kein echter Vorlauf; unvertretbare Fehlalarmrate; fehlende Konstruktvalidität; fehlende Reproduzierbarkeit; gescheiterter Cross-Domain-Transfer; Recovery, die sich nicht von Artefakten unterscheiden lässt; oder Unsicherheit, die das Signal vollständig überdeckt.

Diese Bedingungen benötigen vor einer konfirmatorischen Studie quantitative Kriterien. Andernfalls können Schwellen nachträglich verschoben werden. Ein aktueller Schwachpunkt des Programms ist, dass mehrere Kill Conditions noch keine domänenspezifisch präregistrierten Grenzwerte besitzen. Das ist keine Widerlegung des Frameworks, aber eine offene Operationalisierungsaufgabe.

Konkurrierende Erklärungen müssen mitgeführt werden. Für spektrale Konzentration kommen etwa approaching instability, veränderte Noise-Struktur, Beobachtungsdrift, externe Schocks, Nichtnormalität oder High-Dimensional-Bulk-Effekte infrage. Eine plausible Erklärung ist kein Beweis.

Positive und negative Kontrollen sind ebenso notwendig wie adversariale Benchmarks. Besonders wichtig sind Fälle, die bekannte Schwächen provozieren: nahezu degenerierte Eigenwerte, starke Nichtnormalität, wechselnde Samplingrate, Sensor-Drift, multiple kritische Moden oder Tipping

ohne klassische CSD-Signatur.

Prospektive Validierung ist stärker als retrospektive Mustererkennung. Base Rates müssen berücksichtigt werden: P(Signal|Transition) ist nicht P(Transition|Signal). Kalibrierung und Diskrimination sind unterschiedliche Eigenschaften.

Abstention ist ein eigener wissenschaftlicher Output. DATA_INSUFFICIENT, MODEL_NOT_APPLICABLE, NON_IDENTIFIABLE, UNCERTAINTY_TOO_HIGH, CONTRADICTORY_EVIDENCE und NOT_VALIDATED_FOR_DOMAIN sind unterschiedliche Gründe, keine Diagnose zu stellen.

Eine Claim Ladder hilft, Überdehnung zu vermeiden: Observation → Statistical Interpretation → Model-Based Interpretation → Diagnostic Interpretation → Operational Claim. Jede Stufe benötigt zusätzliche Evidenz.

Die Minimalregel lautet: Behaupte nie mehr, als die schwächste notwendige Evidenzstufe trägt.

Kapitel 10 – BenchEWS Studio: Forschung als reproduzierbares Instrument

BenchEWS Studio ist nicht der Beweis des Forschungsprogramms. Es ist die Software- und Provenienzinfrastruktur, mit der Teile des Programms ausführbar und überprüfbar werden sollen.

Der dokumentierte Workflow trennt Reference Benchmark, Idea, Project, Configuration, Validation, Execution, Verification, Interpretation, Publication, Reuse, Object Browser und Scientific Modules. Diese Trennung ist wissenschaftlich bedeutsam: Eine Berechnung und ihre Interpretation sind nicht dasselbe Ereignis.

Configuration Validation prüft formale Konsistenz einer Analysekonfiguration. Scientific Validation ist etwas anderes: Sie fragt, ob ein Verfahren tatsächlich das behauptete wissenschaftliche Objekt misst. Release Authorization wiederum betrifft das konkrete Softwareartefakt. Diese drei Ebenen dürfen nicht in ein globales PASS verschmolzen werden.

Ein Reference Benchmark kann zeigen, dass Installation und Pipeline reproduzierbar funktionieren. Er kann nicht zeigen, dass eine Statistik ein valider Frühindikator ist. Unit-, Integration-, Regression- und Smoke-Tests reduzieren Softwarefehlerklassen; sie ersetzen keine Konstruktvalidität.

Für reproduzierbare Forschung braucht es eine Source-of-Truth-Kette: Repository, Commit, Tag, Releaseartefakt, Checksum, Buildkonfiguration, Dependencies und Dokumentationsversion.

Für numerische Reproduzierbarkeit kommt eine zweite Regel hinzu: Die Messlatte muss vor der Messung feststehen. BenchEWS verwendet deshalb für nicht-exakte Cross-Platform-Paritätsvergleiche versionierte Tolerance Contracts. Ein solcher Contract definiert die zulässige numerische Abweichung, wird vor der Auswertung eingefroren und kryptografisch an die Entscheidung gebunden. Fehlt er, ist er noch PENDING oder wurde er verändert, darf der Vergleich kein PASS erzeugen. Eine nach Kenntnis des Ergebnisses verbreiterte Toleranz wäre Post-hoc-Tuning; eine fachlich begründete Änderung muss stattdessen eine neue Contract-Version erzeugen. Für CRTI zwischen macOS/Desktop und Linux/Web bleibt der entsprechende Contract bewusst ohne numerische Toleranz PENDING, bis eingefrorene Mac-Referenzausgaben und die

zugehörige Python-/NumPy-/BLAS-Umgebung vorliegen. Diese Governance verändert nicht die Mathematik des Scientific Core, sondern die Evidenzregel, unter der Reproduzierbarkeit behauptet werden darf. Runtime-Screenshots können zeigen, dass eine Oberfläche einen Wert ausgegeben hat. Sie zeigen nicht, welche Formel intern ausgeführt wurde.

Genau hier bleibt für CRTI die offene SRC-004-Lücke. Studio 2.0 dokumentiert Runtime-Ausgaben, aber die autoritative Source-to-Release-Bindung der aktuellen CRTI-Formeln ist im Buchaudit noch nicht geschlossen.

Zusätzlich wurde im Audit des Scientific Release Companion ein Dokumentationsproblem identifiziert: Eine Passage beschreibt eine gegen 1 steigende leading-eigenvalue fraction, während die peer-reviewte Größe Φ ein effective rank ist, der gegen 1 fällt. Außerdem wird das AIP-Paper an mehreren Stellen zu stark als 'CRTI derivation' etikettiert. Das Buch übernimmt diese Formulierungen nicht; das Companion benötigt separat ein Erratum beziehungsweise eine korrigierte Fassung.

Studio sollte langfristig als Evidence Interface funktionieren: Result, Method, Module Version, Input Adequacy, Method Adequacy, Execution Status, Evidence/Uncertainty, Contradictions, Interpretation und Non-Claim sollten sichtbar sein. Ein experimentelles Modul darf visuell nicht denselben epistemischen Status wie eine etablierte Statistik suggerieren.

Der gegenwärtige Status von Studio 2.0 ist daher konservativ: Workflow implemented/documented; klassische Statistikmodule implemented; CRTI runtime documented/experimental; CRTI scientific validation not established; authoritative CRTI formula provenance still open in the book audit.

Kapitel 11 – Domänenübertragung: Was BenchEWS beanspruchen darf – und was nicht

BenchEWS verwendet Begriffe, die in vielen Domänen sinnvoll klingen: Beobachtung, Feedback, Reaktionsmöglichkeiten, Kompression, Recovery. Eine gemeinsame Sprache beweist jedoch keine gemeinsame Mechanik.

Dieses Buch unterscheidet fünf Transferstufen: Analogie, funktionale Vergleichbarkeit, Modell-Mapping, empirischer Transfer und Mechanismustransfer. Jede Stufe benötigt mehr Evidenz.

Ein Domain Adapter übersetzt zwischen einem abstrakten Konstrukt und einer domänenspezifischen Operationalisierung. Er muss Bedeutung, Richtung, Ordnung und Unsicherheitssemantik hinreichend bewahren. Wenn dies nicht gelingt, scheitert der Transferclaim für diese Dimension.

Ein technischer Sensor und ein organisatorischer Meldeweg können funktional beide Informationszugang betreffen. Ihre Fehlerstrukturen sind dennoch verschieden. Ein gemeinsames O-Konstrukt kann sinnvoll sein, ein identisches Messmodell wahrscheinlich nicht.

Dasselbe gilt für F, D, R und V. Controllability, Reachability und Viability sind etablierte mathematische Gebiete. BenchEWS beansprucht keine neue Mathematik, nur weil diese Werkzeuge in eine gemeinsame Messarchitektur eingeordnet werden.

Der mathematische Satz aus Kapitel 5 darf nur über eine explizite Modellkette übertragen werden:

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

Nicht: Theorem → Real World.

Soziale, organisatorische und ökonomische Systeme erfordern besondere Vorsicht, weil Intentionalität, Strategie, Goodhart-Effekte und Reflexivität eine Messung selbst verändern können. In solchen Bereichen sind Metaphern und funktionale Parallelen legitim, solange sie klar als solche gekennzeichnet werden.

Cross-Domain bedeutet daher nicht, Wälder, Organisationen und Stromnetze auf einer einzigen universellen Selbstkorrekturskala zu ranken. Plausibler ist Process Comparability: verschiedene Domänen können unterschiedliche Messgrößen besitzen, aber dieselbe Evidenzdisziplin verwenden.

Der gegenwärtig stärkste Cross-Domain-Claim lautet daher: BenchEWS schlägt eine gemeinsame metrologische Fragelogik vor. Ein gemeinsamer Mechanismus aller komplexen adaptiven Systeme wird nicht behauptet.

Kapitel 12 – Von der Warnung zur Wiedergewinnung von Selbstkorrektur

Ein Forschungsprogramm über Degradation bleibt unvollständig, wenn es nicht fragt, ob verlorene Korrekturfähigkeit wiedergewonnen werden kann. Recovery ist jedoch nicht einfach Loss mit umgekehrtem Vorzeichen.

Adaptive Reopening und Self-Correction Recovery werden getrennt. Adaptive Reopening betrifft spezifisch die mögliche Wiedererweiterung des korrektiven Optionsraums D. Self-Correction Recovery betrifft die multidimensionale Veränderung des gesamten Profils O|F|D|R|V.

Ein Anstieg von D allein bedeutet keine vollständige Recovery. Ein System kann neue Optionen besitzen, sie aber wegen schlechtem Feedback oder fehlender Ausführung nicht realisieren. Deshalb ist 'reopened but not realized' ein eigener diagnostischer Zustand.

Recovery kann partiell, asymmetrisch und hysteretisch sein. Unterschiedliche Dimensionen können verschiedene Erholungszeiten besitzen. Ein System kann in einen neuen funktionsfähigen Zustand gelangen, ohne zum ursprünglichen Zustand zurückzukehren.

Regeneration geht konzeptionell darüber hinaus. Sie bezeichnet die mögliche Neuorganisation, durch die neue Feedbackwege, neue Optionen oder neue funktionsfähige Strukturen entstehen. Dieser Begriff ist im BenchEWS-Programm derzeit prospektiv und nicht als validierte Metrik etabliert.

Viability Theory bietet etablierte Mathematik für die Frage, ob innerhalb von Constraints zulässige Pfade existieren. Ein horizon-relative lock-in kann deshalb sinnvoller sein als absolute Irreversibilität. Eine wissenschaftlich vorsichtige Aussage lautet: Unter Modell M, Kontrollsatz U, Ressourcenbudget B, Unsicherheit und Horizont T wurde kein viabler Recovery-Pfad gefunden. Das ist nicht identisch mit 'Recovery ist in der Realität unmöglich'.

Kandidaten wie regenerative variety margin oder conditional irreversibility gehören zur späteren Forschungsfront. Sie sind keine Ergebnisse von Studio 2.0 und keine validierten universellen Kennzahlen.

Die entscheidende nächste Phase des Programms ist empirisch: Definition Freeze, Ground Truth by Construction, adversariale Benchmarks, unabhängige Implementierungen, reale kontrollierbare Systeme, Cross-Domain-Replikation und prospektive Validierung.

Das Programm muss dabei sein eigenes Scheitern zulassen. Vielleicht ist Self-Correction Capacity nicht als gemeinsames Cross-Domain-Measurand operationalisierbar. Vielleicht funktionieren einzelne Dimensionen, das Gesamtprofil aber nicht. Vielleicht besitzt CRTI keinen Zusatznutzen. Solche Ergebnisse wären wissenschaftlich wertvoll.

Die Leitfrage bleibt deshalb offen: Wie viel Korrekturfähigkeit bleibt einem System – und unter welchen Bedingungen lässt sich ihr Verlust oder ihre Wiedergewinnung zuverlässig messen?

Anhang A – Verbindliche CRTI-Symboltabelle

Symbol	Bedeutung	Status
Φ(Σ)	effective rank der stationären Populationskovarianz	DERIVED unter A1–A4
p_1	$\lambda_1/tr\Sigma$	definierte Spektralgröße, nicht Φ
Φ̂_W	finite-window estimator von Φ	AUDIT FINDING / unvalidiert
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

Anhang B – Offene Forschungs- und Provenienzfragen

1. SRC-004: autoritative Studio-2.0-CRTI-Source-to-release-Bindung und aktuelle mathematische Formelbindung.
2. SRC-005: Korrektur des Scientific Release Companion hinsichtlich Φ versus p_1 sowie der Bezeichnung des AIP-Resultats.
3. SRC-006: präregistrierbare quantitative Kill Conditions für mindestens einen konfirmatorischen synthetischen Benchmark.
4. AUD-001: mathematische Clarification zum defekten Jordan-Randfall außerhalb von A1.
5. Erweiterung des bibliographischen Apparats insbesondere zu Viability Theory, nichtnormaler Dynamik/Pseudospektren und informationstheoretischen Grenzen von Kontrolle.

Diese Punkte sind Teil der offenen Forschungs- und Dokumentationsagenda. Sie werden in dieser Ausgabe nicht als bereits gelöst dargestellt und blockieren deshalb nicht die Veröffentlichung des Buches als dokumentierten Forschungsstand.

Anhang C – Kernreferenzen

- 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.