World Scientific
www.worldscientific.com

QUANTIFYING EMERGENCE IN TERMS OF PERSISTENT MUTUAL INFORMATION

ROBIN C. BALL* and MARINA DIAKONOVA[†]

*Centre for Complexity Science and Department of Physics,
University of Warwick, Coventry, CV4 7AL, UK*
*R.C.Ball@warwick.ac.uk
†M.Diakonova@warwick.ac.uk

ROBERT S. MACKAY

*Centre for Complexity Science and Department of Mathematics,
University of Warwick, Coventry, CV4 7AL, UK*
R.S.MacKay@warwick.ac.uk

We define Persistent Mutual Information (PMI) as the Mutual (Shannon) Information between the past history of a system and its evolution significantly later in the future. This quantifies how much past observations enable long-term prediction, which we propose as the primary signature of (Strong) Emergent Behavior. The key feature of our definition of PMI is the omission of an interval of "present" time, so that the mutual information between close times is excluded: this renders PMI robust to superposed noise or chaotic behavior or graininess of data, distinguishing it from a range of established Complexity Measures. For the logistic map, we compare predicted with measured long-time PMI data. We show that measured PMI data captures not just the period doubling cascade but also the associated cascade of banded chaos, without confusion by the overlayer of chaotic decoration. We find that the standard map has apparently infinite PMI, but with well-defined fractal scaling which we can interpret in terms of the relative information codimension. Whilst our main focus is in terms of PMI over time, we can also apply the idea to PMI across space in spatially-extended systems as a generalization of the notion of ordered phases.

Keywords: Emergence; persistent mutual information; chaotic dynamical systems; complexity measure; logistic map.

1. Introduction

Our starting point is the desire to discover and quantify the extent to which the future evolution of a dynamical system can be predicted from its past, and from the standpoint of Complexity Theory we are interested in assessing this from observed data alone without any prior parametric model.

We should nevertheless admit prior classification as to the general nature of the system, and simple constraining parameters such as its size, composition and local

laws of motion. Given that information, there may prove to be entirely reproducible features of its subsequent evolution which inevitably emerge over time, such as eventual steady state behavior (including probability distributions). This we follow [1, 3] and others in terming *weak emergence*. The emergence is weak in terms of there being no choice of outcome, it can be anticipated without detailed inspection of the particular instance.

We focus on *Strong Emergence* by which we mean features of behavior significantly into the future which can only be predicted with knowledge of prior history. The implication is that the system has made conserved choices not determined by obvious conservation laws, or at least nearly conserved choices which imply the existence of associated slow variables.

More formally, we must conceive of an ensemble comprising a probability distribution of realizations of the system (and its history), from which observed histories are drawn independently. The behavior of a particular realization which can be anticipated from observing other realizations is weakly emergent, whilst that which can only be forecast from the observation of the past of each particular instance is strong emergence. A related distinction between weak and strong emergence is given in [19], but with quantification based on a metric on the underlying space, rather than purely measure-theoretic.

2. Persistent Mutual Information

Within an ensemble of histories of the system, we can quantify strong emergence in terms of mutual information between past and future history which persists across an interval of time τ. This *Persistent Mutual Information* is given by

$$I(\tau) = \int \log\left(\frac{P[x_{-0}, x_{\tau+}]}{P[x_{-0}]P[x_{\tau+}]}\right) P[x_{-0}, x_{\tau+}]dx_{-0}dx_{\tau+} \tag{1}$$

where x_{-0} designates a history of the system from far past up to present time 0; $x_{\tau+}$ is the corresponding history of the system from later time τ onwards; $P[x_{-0}, x_{\tau+}]$ is their joint probability density within the ensemble of histories; and $P[x_{-0}]P[x_{\tau+}]$ is the product of corresponding marginal probability densities for past and future taken separately. If the history variables $x(t)$ are discrete-valued, then the integration over histories is interpreted as summation; in the continuous case, $I(\tau)$ has the merit of being independent of continuous changes of variable, so long as they preserve time labelling.

Quantitatively, $I(\tau)$ measures the deficit of (differential) Shannon Entropy in the joint history compared to that of past and future taken independently, that is

$$I(\tau) = H[P[x_{-0}]] + H[P[x_{\tau+}]] - H[P[x_{-0}, x_{\tau+}]] \tag{2}$$

where the separate (differential) Shannon Entropies for a probability density P of a set of variables y are given generically by

$$H[P] = -\int \log(P[y])P[y]dy. \tag{3}$$

For continuously distributed variables, the differential Shannon Entropies carry an infinite offset from absolute information content, but those offsets can be shown to cancel from Eq. (2) above. Thus $I(\tau)$ is precisely the amount of information (in Shannon Entropy) about the future which is determined by the past, and hence *the extent to which the future can be forecast from past observations* (of the same realization).

The key distinguishing feature of our definition above is the exclusion of information on $x_{0\tau}$, that is the intervening time interval of length τ. This ensures that $I(\tau)$ is only sensitive to system memory which persists right across time τ; any features of shorter term correlation do not contribute. The choice of τ must inevitably be informed by observation, but the extreme cases have sharp interpretation.

$I(0)$ corresponds directly to the Excess Entropy as introduced in [12] where it was called "Effective Measure Complexity", and makes no distinction of timescale in the transmission of information. This quantity has been discussed in many guises: as effective measure complexity in [12, 17], Predictive Information in [2] and as Excess Entropy in [7, 11], to name but a few. See also [22, 13, 11] for measurements of Excess Entropy and the related Entropy Rate on a variety of systems, including the logistic map.

Our sharpest measure of Strong Emergence is the *Permanently Persistent Mutual Information* (PPMI), that is the PMI $I(\infty)$ which persists to infinite time. This quantifies the degree of permanent choice spontaneously made by the system, which cannot be anticipated without observation but which persists for all time. A prominent class of example is spontaneous symmetry breaking by ordered phases of matter: here a physical system is destined to order in a state of lower symmetry than the probability distribution of initial conditions, and hence must make a choice (such as direction of magnetization) which (on the scale of microscopic times) endures forever. As a result, Strong Emergence can only be diagnosed by observing multiple independent realizations of the system, not just one long-time history. An interesting though anomalous case is presented by clock phase, where time shift leads to different phases. This is exploited in measuring PPMI for the logistic map in the following section (see Fig. 1).

PPMI corresponds to some partitioning of the attracting dynamics of the system into negligibly communicating (and negligibly overlapping) subdistributions. If the dynamics evolves into partition i with probability p_i, then the PPMI is simply given by

$$I(\infty) = -\sum_i p_i \log(p_i) \tag{4}$$

which is the entropy of the discrete distribution p_i. For deterministic dynamics, each p_i is simply determined by the sampling of its associated basin of attraction in the distribution of initial conditions, so in this case, the PPMI is sensitive to the
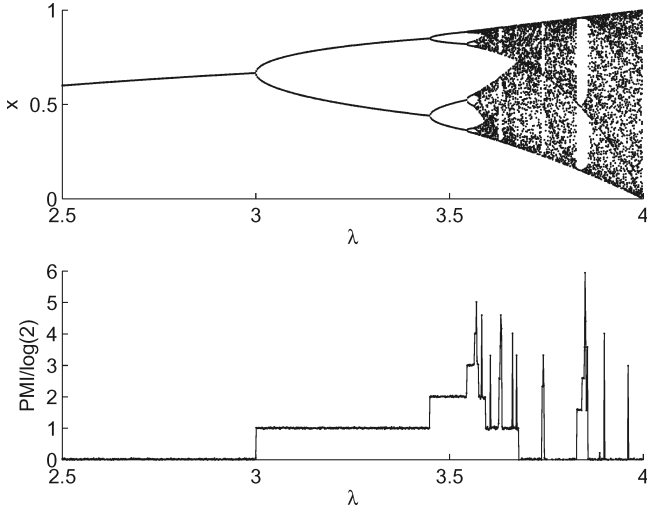
Fig. 1.   Bifurcation diagram (above) and measured Persistent Mutual Information (below) for the logistic map, as a function of the map control parameter λ. For each value of λ, the map was allowed a minimum of 10^5 iterations to settle and then the Mutual Information measured across a "time" interval of 10^5 iterations. Each MI measurement used the distance to 4th nearest neighbor to estimate probability density ($k = 4$), based on a sample of $N = 5000$ iterate pairs. Before chaos sets in, PMI increases stepwise in jumps of $\log 2$, reflecting the doubling of the resolved period. It is also seen to pick up the band periodicity after the onset of chaos (resolving some more fine periods within the band hopping), as well as a nonchaotic period three regime.

latter. However, for stochastic dynamics, it is possible for the p_i to be predominantly determined by the distribution of early time fluctuations.

3. PPMI in the Logistic Map

We consider time series from the logistic map $x_{n+1} = \lambda x_n(1 - x_n)$ as a simplest non-trivial example which brings out some non-trivial points. Depending on the control parameter λ, its attracting dynamics can be a limit cycle, fully chaotic, or the combined case of banded chaos where there is a strictly periodic sequence of bands visited but chaotic evolution within the bands [5].

In the case of a periodic attracting orbit with period T, the choice of phase of the cycle is a permanent choice leading directly to positive PPMI. For this case, the phase separation is in the time domain, so the attractor can be fully sampled by shifting the start time of observation. If we assume the latter is uniformly distributed over a range of time large enough compared to the period, then the observed phase is a result of a uniformly selected and symmetry breaking choice in which each phase has $p_i = 1/T$, and this leads to PPMI

$$I(\infty) = \log(T). \tag{5}$$

This is a generic result and not special to the logistic map. Because there is just an attracting orbit, the Excess Entropy gives the same value.

For fully chaotic attracting dynamics such as at $\lambda = 4$, we have to be careful in principle about limits. Provided the probability densities are measured with only limited resolution δx in x, then we expect past and future to appear effectively independent for $\tau > \tau(\delta x)$ and hence $I(\tau) \to 0$ and there is zero PPMI. Thus, for chaotic motion, the associated values are quite different: the Excess Entropy is positive and reflects the complexity of its dynamics, whereas the PPMI is zero reflecting the absence of long-time correlations.

Both of the above results can be seen in measured numerical data for the logistic map in Fig. 1. What is more pleasing still is the behavior of PMI for banded chaos, where a T-periodic sequence of bands shows through to give $I(\infty) \geq \log(T)$ (assuming random initiation phase as before) with equality when the Tth iterate restricted to one band is mixing. The fact that the numerical results overshoot $\log(T)$ for many parameter values can be attributed to the presence of a finer partition than the T bands, for example, around an attracting periodic orbit of period a multiple of T or into sub-bands with a period a multiple of T. Even in cases where the dynamics really is T-periodic mixing (meaning it cyclically permutes T bands and the Tth iterate restricted to one is mixing), and hence PPMI is precisely $\log(T)$, the numerics might pick up some long-time correlation that does not decay until after the computational time interval. One can see this in the more detailed data of Fig. 2 where there are background plateaux corresponding to the periodicity of the observed major bands, decorated by narrower peaks corresponding to higher
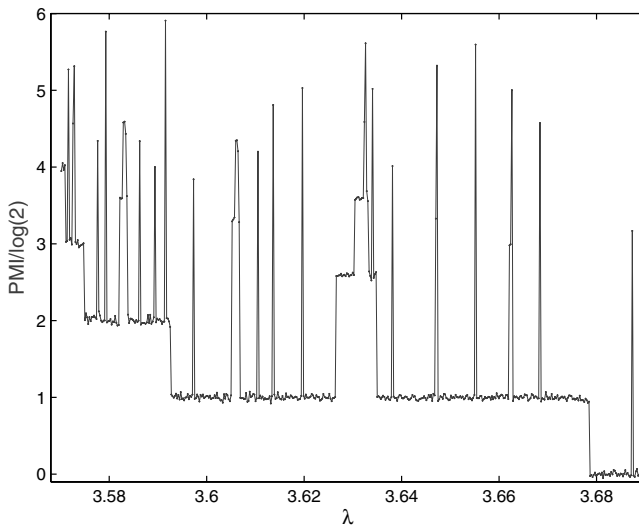

Fig. 2.   PMI for a chaotic region of the control parameter (sample size $N = 5000$, $k = 4$, settle time 10^{10}, time separation 10^{6}). In this range, chaotic bands are known to merge (see bifurcation diagram in Fig. 1). PMI picks out the relevant decrease in overall periodicity as 16 bands merge pairwise into eight, four, and finally two at λ slightly less than 3.68. PMI also detects a period tripling regime, which can be seen around $\lambda = 3.63$.

periodicities. The steps in PMI are particularly clear where bands merge because these special points have strong mixing dynamics within each band cycle [20]. Figures 1 and 2 for the PPMI can be usefully contrasted with the Excess Entropy graphed in Fig. 1 of [11]. On the period-doubling side, they display the same values (log of the period), whereas in the regions of banded chaos the PPMI picks out the number of bands whilst the Excess Entropy is complicated by its sensitivity to short-time correlations in the chaotic decoration.

4. Issues Measuring PMI

Measuring Mutual Information and in particular the implied measurement of the entropy of the joint distribution suffers from standard challenges in measuring the entropy of high dimensional data. The naive "histogram" method, in which probability densities are estimated directly from frequency counts in pre-selected (multidimensional) intervals, is easy to apply but can require very large sample sizes in order to ensure that the significant frequencies are estimated from multiple (rather than single) counts. In practice, we found the kth neighbor approach of Kraskov *et al.* [16] a more effective tool (from now on referred to as the k-NN method). It is more limited in sample size due to unfavorable order of algorithm, but this was outweighed by its automatic adjustment of spatial resolution to the actual density of points.

The basis of the k-NN method is to estimate the entropy of a distribution from the following estimate of the logarithm of local probability density about each sampled point:

$$\log(p) \simeq \log\left(\frac{k/N}{\epsilon_k^d}\right) + [\Psi(k) - \Psi(N) - \log(k/N)] = -\log\epsilon_k^d + \Psi(k) - \Psi(N) \quad (6)$$

where N is the total number of sample points and ϵ_k^d is the volume of space out to the location of the kth nearest neighbor of the sample point in question. The combination $\frac{k/N}{\epsilon_k^d}$ in the first logarithm is simply interpretable as an amount of sampled probability in the neighborhood divided by corresponding volume. In the remaining term $\Psi(z) = \Gamma'(z)/\Gamma(z) \simeq \log(z) - \frac{1}{2z}$ as $|z| \to \infty$ is the digamma function and the whole of this term is only significant for small k, where it corrects a slight bias associated with finite sampling of the neighborhood [16].

We interpret $\log(N/k) \simeq \Psi(N) - \Psi(k)$ as the (logarithmic) probability resolution in these measurements. When N and either k is large or held fixed, the variation of these two forms is equivalent and we generally show the first for simplicity of exposition. The main exception is for entropy data on the standard map below, where taking data down to $k = 1$ significantly enhances our appreciation of scaling and the use of the more accurately scaled second form is important.

Because PMI is invariant under changes of variable, there is considerable scope for choice of how to parametrize past and future before feeding into the PMI measurement. For the logistic map, we exploited its deterministic nature, by which one value of x_n provides all information about the past up to iterate n which is relevant

to the future, and similarly all influence of the past on the future beyond iterate n' is captured by the value of $x_{n'}$. Note however that we did not require to identify minimalist causal states in the sense discussed in Sec. 5 below.

For systems without known causal coordinates, the practical measurement of PMI has a rich time parametrization. In principle, what we can directly measure is the mutual information $I(t_1, t_2; t_3, t_4)$ between time intervals $[t_1, t_2]$ and $[t_3, t_4]$. If we assume stationarity, then this is more naturally parametrized as $I(\tau; T_-, T_+)$ in terms of the intervals $T_- = t_2 - t_1$ and $T_+ = t_4 - t_3$ of past and future respectively as well as the intervening interval $\tau = t_3 - t_2$. Then the full PMI is defined as

$$I(\tau) = \lim_{T_-, T_+ \to \infty} I(\tau; T_-, T_+). \tag{7}$$

If the PPMI is desired, it is computationally efficient to set $\tau \to \infty$ before taking the limits above, because the dimension of space in which entropy must be measured is set by $T_- + T_+$ alone. By contrast with PMI, the Predictive Information developed at length in [2] is in the present notation $I(0; T, T)$.

Practical measurement of PMI entails some limited resolution, whether explicitly by histogram methods or implicitly through the depth of sampling in k-NN and other adaptive methods. This inevitably leads to long periodic orbits being capped in their apparent period and hence their measured $I(\tau)$. We can be fairly concrete in the case of measurement by the k-NN method, which looks out across a neighborhood whose aggregate measure is k/N. The longest period one can thereby detect is of the order N/k so we are led to expect $I(\tau) \simeq \log(\min(T, N/k))$. One point
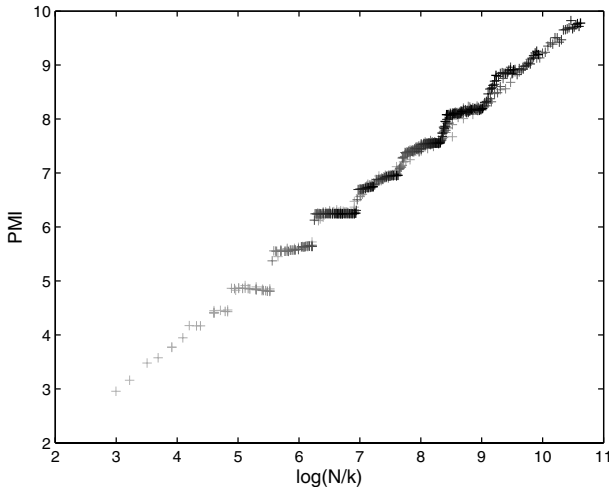

Fig. 3. PMI of the logistic map at the period-doubling accumulation point plotted against the effective resolution with which probability density has been measured, given by $\log(N/k)$. ($N = 1000, 2000, \ldots, 41000$ where runs with higher N correspond to darker points; $k = 1, 2, 3, 4, 5, 10, \ldots, 50$, time separation 10^8, settle time 10^{10}.) The periodicity measured is resolution limited, with apparent overall slope of this plot ca. 0.9 in fair agreement with slope unity predicted on the basis of resolvable period $\propto N/k$ (see text).

where we can check this quantitatively is the accumulation point of the period-doubling sequence [6, 10, 15]: Fig. 3 shows the measured results agreeing with the expectation that $I(\tau) \simeq \log(N/k)$.

5. Relationship with Statistical Complexity

Statistical Complexity (in certain contexts, it is equivalent to "True Measure Complexity" first introduced in [12]; see also [8, 21]) is built on the projection of past and future down to optimal causal states $S_-(t)[x_{t-}]$ such that $P[x_{t+}|x_{t-}] = P[x_{t+}|S_-(t)]$. In terms of these, one readily obtains that the PMI is given by the Mutual Information between time-separated forward and reverse causal states, that is

$$I(\tau) = H[P[S_-(t)]] + H[P[S_+(t+\tau)]] - H[P[S_-(t), S_+(t+\tau)]] \qquad (8)$$

as a straightforward generalization of the corresponding result for the Excess Entropy $I(0)$ [12].

   In general, one cannot simplify the above formula (I_{-+} in a natural extension of the notation) to use other combinations of choice between S_- or S_+. However, we conjecture that under fairly general conditions they are all equivalent for the PPMI, i.e., $\tau \to \infty$. It is an interesting open question whether for general time gap τ it can be proved that $I_{-+} \geq (I_{--}, I_{++}) \geq I_{+-}$, and perhaps also $I_{--} = I_{++}$. For $\tau = 0$, similar forward and reverse time dependencies for ϵ-machines have been considered in [9], where it was noted that in general $I_{--} \neq I_{++}$, and bidirectional machines were defined that incorporate this time asymmetry.

6. Fractal and Multifractal PMI: Example of the Standard Map

As we already observed for the accumulation points of the standard map, where the attractor of the dynamics is a Cantor fractal adding machine, the measured PMI may go to infinity as the resolution increases. The archetypal case of this is where the probability measures themselves have fractal support or more generally exhibit multifractal scaling. The general phenomenology is that dividing space of dimension d into cells c of linear width ϵ of integrated measure $\mu_c = \int_c P(x)d^dx$, the density in a cell is estimated as μ_c/ϵ^d and hence to leading order as $\epsilon \to 0^+$ one expects

$$H[P] = -\sum_c \mu_c \log\left[\frac{\mu_c}{\epsilon^d}\right] = d\log\epsilon - D\log\epsilon + \text{constant} \qquad (9)$$

where D is the information dimension of the integrated (natural) measure μ, defined to be

$$D = \lim_{\epsilon \to 0^+} \frac{\sum_c \mu_c \log \mu_c}{\log \epsilon}. \qquad (10)$$

Applying this to the PMI through Eq. (2) then leads to

$$I(\tau) = I_0(\tau) - (D_- + D_+ - D_{-+})\log\epsilon \qquad (11)$$

where D_{-+} is the information dimension of the joint distribution of past and future, D_- and D_+ the information dimensions of the respective marginal distributions, and $I_0(\tau)$ the extrapolated resolution-free PMI (note the dimensions of the underlying spaces cancel from $I(\tau)$ because $d_{-+} = d_- + d_+$).

Applying the equivalent analysis to the k-NN method, we have to be careful to insist that Eq. (1) is used, meaning in particular that it is a neighborhood of k neighbors in the *joint* distribution which is taken to determine the ratio of joint and marginal probability densities within the logarithm. With this understanding, $\log \epsilon$ in the above expression for PMI can be written in terms of probability distribution $\frac{1}{D_{-+}} \log(k/N)$ leading to

$$I(\tau) = I_0(\tau) + \Gamma \log(N/k) \tag{12}$$

where

$$\Gamma = \frac{D_- + D_+ - D_{-+}}{D_{-+}} \tag{13}$$

is the relative information codimension.

Our first check is the accumulation point of period doubling of the logistic map, at which the dynamics causally orbits its Cantor set attractor. In this case all the information dimensions above are each equal to the fractal dimension of the Cantor set, leading to $\Gamma = 1$ in agreement with our earlier observations and interpretation based on resolution limited period. Figure 3 shows the directly measured PMI with apparent $\Gamma \simeq 0.9$ in fair agreement, where sampling over the attractor set is achieved by using different times of measurement. A uniform distribution of measurement times approximates uniform sampling of the unique invariant probability measure of the attractor.

The standard map provides a much more subtle test of this phenomenology. This two-dimensional map

$$p' = (p + K \sin x)_{\text{modulo } 2\pi}, \quad x' = (x + p')_{\text{modulo } 2\pi} \tag{14}$$

is strictly area-preserving in the x, p plane (reduced modulo 2π) and has uniform invariant measure. Thus if we launch this dynamics with random initial conditions, the marginal distributions (joint of x, p at fixed iteration) remain strictly uniform forever. The joint distribution between distant iterations of the standard map is far from simple and uniform, at least for moderate values of the map parameter K. Figure 4 shows the measured PMI as a function of the probability resolution k/N, displaying clear fractal behavior for values of $0 \leq K < 6$.

The corresponding estimates of the relative information codimension are shown in Fig. 5. For the left end point we can anticipate $\lim_{K \to 0} \Gamma(K) = 1/3$ on theoretical grounds, because this limit corresponds to a continuous time Hamiltonian dynamics which is energy (Hamiltonian) conserving. The joint distribution can therefore only explore $D_{-+} = 4 - 1$ degrees of freedom leading to $\Gamma = 1/3$. Note this result depends on the assumption that the shear in the dynamics between close energy shells is sufficient to destroy correlation of their in-shell coordinates, and we did
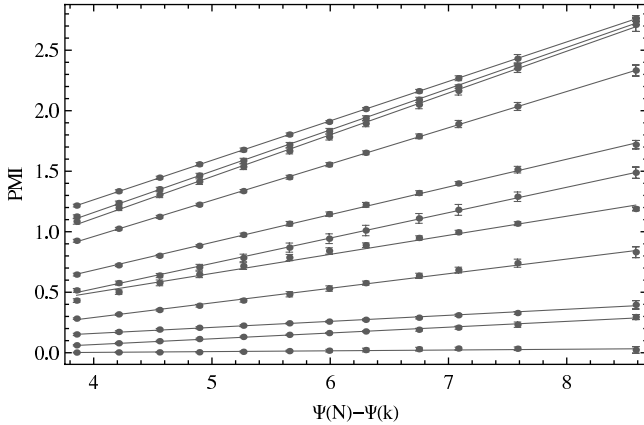
Fig. 4.   PMI for the standard map at 3000 iterations, as a function of the probability resolution used to measure it, for map parameters $K = 0.1, 0.5, 0.9, 1.0, 1.1, 1.2, 1.5, 2, 3, 4, 6$ (top to bottom). The resolution of probability is plotted as $\Psi(N) - \Psi(k) \simeq Ln(N/k)$, where $N = 3000$ is the number of sample points used, and k is the rank of neighbor used in the measurement of Mutual Information (see text) which ranges from 1 to 64 across each plot. For each map parameter, there is a clear linear dependence on this logarithmic resolution, consistent with fractal phenomenology and the interpretation of the slope as a relative information codimension. The data have been averaged over five independent sets of such measurements with the error bars showing the 2σ error in each mean, and the drawn lines correspond to the slopes plotted in Fig. 5.
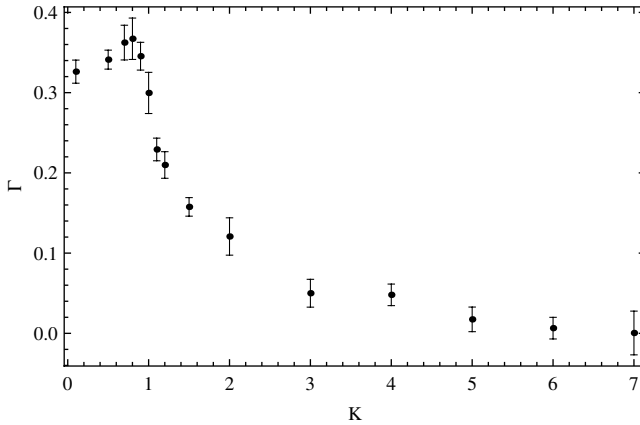

Fig. 5.   Relative information codimension Γ for the standard map (at 3000 iterations) as a function of map parameter K. These are the slopes of the data shown in Fig. 4, and the error bars are estimated as 2σ based on separate best fit to the five independent runs of data. The intercept at $K = 0$ matches theoretical expectation of $1/3$ (see text) and the fall to zero at large K is consistent with dominance by chaotic dynamics. It is interesting that Γ peaks in the vicinity of $K_g = 0.97$ where the golden KAM curve breaks up, but anomalously slow dynamical relaxation in this region (see [4] and references therein) means that the peak may not reflect the limit of infinite iterations and hence true PPMI.

correspondingly find that we had to use a large number of map iterations for the expected behavior to emerge.

The apparent peak in Γ around $K = 1$ is particularly interesting because this is the vicinity of K_c where momentum becomes diffusive, closely associated (and often identified) with the breakup of the golden KAM (Kolmogorov–Arnold–Moser) curve at $K_g = 0.971635\ldots$ [14, 18]. Dynamical anomalies have been observed around this critical value of K which might underlie the peak we observe in PPMI. On the other hand, corresponding long-time dynamical correlations pose a threat to whether our results are adequately converged.

For larger K, our measurements are consistent with $\Gamma(K) = 0$ for $K > K_1$ where $6 < K_1 < 7$, and indeed the full PMI is within the uncertainty of zero in this regime, meaning the map appears fully chaotic to the level we have resolved.

7. Conclusions

We have shown that Persistent Mutual Information is a discriminating diagnostic of hidden information in model dynamical systems, and the Permanent PMI is a successful indicator of Strong Emergence.

The detailed behavior of the logistic map is sufficiently re-entrant, with periodicity and cascades of period multiplication intermingled amidst chaos, that we are unlikely to have the last word on the full quantitative behavior of the PPMI as a function of map parameter λ beyond the first cascade.

For the standard map, PPMI reveals some of the subtlety only otherwise accessible through dynamical properties such as explicit orbits. Precise relationships remain an open issue, particularly around critical map parameter K_c. The observed fractal behavior with a deficit between the joint information dimension and those of the marginals is we suggest a general phenomenology. Whether it reflects truly fractal and multifractal behavior in any particular case should rest on a wider multifractal analysis of the joint probability measure, which we intend to address in later work on a wider range of non-trivial dynamical systems.

Application to intrinsically stochastic systems and real-world data are outstanding challenges. We can however readily invoke a wide variety of examples associated with ordering phenomena in statistical physics where a dynamically persistent and spatially coherent order parameter emerges. In these cases, there is clearly PPMI in time, but we can also consider just a time slice and let one spatial coordinate take over the role of time in our PMI analysis. Spin Glasses are a key instance where the two viewpoints are not equivalent: these have order and hence Mutual Information persists in time but not in space.

References

[1] Bar-Yam, Y., A mathematical theory of strong emergence using multiscale variety, *Complexity* **9** (2004) 15–24.

[2] Bialek, W., Nemenman, I. and Tishby, N., Predictability, complexity, and learning, *Neural Comput.* **13** (2001) 2409–2463.

[3] Chalmers, D. J., *Strong and Weak Emergence*, The Re-Emergence of Emergence. (Oxford University Press., 2002).

[4] Chirikov, B. V. and Shepelyansky, D. L., Asymptotic statistics of Poincaré recurrences in Hamiltonian systems with divided phase space, *Phys. Rev. Lett.* **82** (1999) 528–531.

[5] Collet, J.-P. and Eckmann, P., *Iterated Maps on the Interval as Dynamical System*, Modern Birkhäuser Classics (Birkhäuser, 1980).

[6] Coullet, P. and Tresser, C., Iterations d'endomorphismes et groupe de renormalisation, *J. Phys. Colloque C* **539** (1978) C5 – 25.

[7] Crutchfield, J. and Packard, N., Symbolic dynamics of noisy chaos, *Physica D* **7** (1983) 201–223.

[8] Crutchfield, J. and Young, K., Inferring statistical complexity, *Phys. Rev. Lett.* **63** (1989) 105–108.

[9] Ellison, C. J., Mahoney, J. R. and Crutchfield, J. P., Prediction, retrodiction, and the amount of information stored in the present, *J. Stat. Phys.* **136** (2009) 1005–1034.

[10] Feigenbaum, M. J., Quantitative universality for a class of nonlinear transformations., *J. Stat. Phys.* **19** (1978) 25–52.

[11] Feldman, D. P., McTague, C. S. and Crutchfield, J. P., The organization of intrinsic computation: Complexity-entropy diagrams and the diversity of natural information processing, *CHAOS* **18** (2008).

[12] Grassberger, P., Toward a quantitative theory of self-generated complexity, *International Journal of Theoretical Physics* **25** (1986) 907–938.

[13] Grassberger, P., On symbolic dynamics of one-humped maps of the interval, *Z. Naturforschung* **43a** (1988) 671–680.

[14] Greene, J. M., A method for determining a stochastic transition, *Journal of Mathematical Physics* **20** (1979) 1183–1201.

[15] Grossman, S. and Thomae, S., *Z. Naturforschung* **32a** (1977) 1353.

[16] Kraskov, A., Stogbauer, H., and Grassberger, P., Estimating mutual information, *Phys. Rev. E* **69** (2004).

[17] Lindgren, K. and Nordahl, M. G., Complexity measures and cellular automata, *Complex Systems* **2** (1988) 409–440.

[18] MacKay, R. and Percival, I., Converse KAM: Theory and Practice, *Commun. Math. Phys.* **98** (1985) 469–512.

[19] MacKay, R. S., Nonlinearity in complexity science, *Nonlinearity* **21** (2008) T273–T281.

[20] Misiurewicz, M., Absolutely continuous measures for certain maps of an interval, *Publications Mathematiques* (1981) 17–51.

[21] Shalizi, C., Shalizi, K. and Haslinger, R., Quantifying self-organization with optimal predictors, *Phys. Rev. Lett.* **93** (2004).

[22] Zambella, D. and Grassberger, P., Complexity of forcasting in a class of simple models, *Complex Systems* **2** (1988) 269–303.