Supplementary Materials to

Information processing features can detect behavioral regimes of dynamical systems

* r.quax@uva.nl

Lambda parameter and information features

Relation to Lambda parameter. The Lambda parameter is a well-known 'local' characterization of a CA rule which correlates with Wolfram's complexity classes. It is local in the sense that it is computed from the state transition table, so effectively it can be seen as a prediction of the Wolfram class at $t = 0$. Here we demonstrate that our information features includes part of the predictive power of the Lambda parameter by relating the two.

Let us first consider a dynamical system with N random variables $X_i(t)$, whose values are denoted $x_i(t)$, evolves in time according to a given rule, depending on the state of the neighborhood of agent i, denoted n_i (we shall always assume the neighborhood of a node includes the node itself). We assume a probabilistic description so as to write this rule as a transition matrix

$$W_{\{x_j\}\to x_i} = \Pr(x_i, t | \{x_j\}, t-1) \tag{1}$$

$\forall j \in n_i$. This gives the probability that agent i takes value x_i at time t knowing that the neighbors of i take value x_{j_k} for $k \in \{1, ..., |n_i|\}$. With the W matrix we can express two-steps joint probabilities as

$$\Pr(x_i, t; \{x_j\}, t-1) = \Pr(\{x_j\}, t-1) W_{\{x_j\}\to x_i}. \tag{2}$$

We can now define and compute the *total information* I_{tot} as

$$\begin{aligned} I_{tot} &:= I(X_i, t : \{X_j\}, t-1) \\ &= \sum_{\{x_j\}, x_i} \Pr(x_i, t; \{x_j\}, t-1) \ln \frac{\Pr(x_i, t; \{x_j\}, t-1)}{\Pr(x_i, t) \Pr(\{x_j\}, t-1)} \\ &= \sum_{\{x_j\}, x_i} \Pr(\{x_j\}, t-1) W_{\{x_j\}\to x_i} \ln \frac{W_{\{x_j\}\to x_i}}{\Pr(x_i, t)} \end{aligned} \tag{3}$$

Note that I_{tot} is the first term needed to compute *information synergy*, as defined in the main text.

Now comes a crucial assumption: we shall assume that the agents are randomly initialized, namely that $\Pr(\{x_k\}, t-1) = 2^{-|n_k|}$ for any configuration of the neighborhood. In the case of cellular automata one is also allowed to write $n_i = n \; \forall i$, so that

$$
\begin{aligned}
I_{tot} &= 2^{-n} \sum_{\{x_j\},x_i} W_{\{x_j\}\to x_i} \ln \frac{W_{\{x_j\}\to x_i}}{2^{-n}\sum_{\{x_j\}} W_{\{x_j\}\to x_i}} \\
&= 2^{-n} \sum_{\{x_j\},x_i} W_{\{x_j\}\to x_i} \ln W_{\{x_j\}\to x_i} - 2^{-n} \sum_{\{x_j\},x_i} W_{\{x_j\}\to x_i} \ln \left(2^{-n} \sum_{\{x_j\}} W_{\{x_j\}\to x_i} \right) \\
&= -2^{-n} \sum_{\{x_j\},x_i} W_{\{x_j\}\to x_i} \ln \left(2^{-n} \sum_{\{x_j\}} W_{\{x_j\}\to x_i} \right) \\
&= -\lambda \ln \lambda - (1-\lambda)\ln(1-\lambda)
\end{aligned} \tag{4}
$$

where λ is the Langton parameter, namely the fraction of ones in the lookup table characterizing the dynamics. The last equality arises from the fact that

$$
\sum_{\{x_j\}} W_{\{x_j\}\to x_i} = \begin{cases} 2^n \lambda & \text{if } x_i = 1 \\ 2^n(1-\lambda) & \text{if } x_i = 0 \end{cases} \tag{5}
$$

while the penultimate is because for a deterministic dynamics transition elements take either value 0 or 1.

The next piece of information we consider is *memory*, which we define as

$$
I_{mem} := I(X_i, t : X_i, t-1). \tag{6}
$$

It can be computed as

$$
\begin{aligned}
I_{mem} &= \sum_{x_i, a'_i} \Pr(x_i, t; a'_i, t-1) \ln \frac{\Pr(x_i, t; a'_i, t-1)}{\Pr(x_i, t)\Pr(a'_i, t-1)} \\
&= \sum_{x_i, \{a'_j\}} \Pr(x_i, t; \{a'_j\}, t-1) \ln \frac{\sum_{\{a'_{j\neq i}\}} \Pr(x_i, t; \{a'_j\}, t-1)}{\Pr(x_i, t)\Pr(a'_i, t-1)} \\
&= 2^{-n} \sum_{x_i, \{a'_j\}} W_{\{a'_j\}\to x_i} \ln \frac{\sum_{\{a'_{j\neq i}\}} W_{\{a'_j\}\to x_i}}{2^{-1}\sum_{\{a'_j\}} W_{\{a'_j\}\to x_i}}
\end{aligned} \tag{7}
$$

We now refine the previous analysis by defining two quantities λ_0 and λ_1 as Langton parameters restricted to the case when the central site is 0 or 1 respectively. In other words we have

$$
\sum_{a'_m, m\neq i} W_{\{a'_m, a'_i\}\to x_i} = \begin{cases} 2^n(1/2-\lambda_0) & \text{if } x_i = 0 \text{ and } a'_i = 0 \\ 2^n\lambda_0 & \text{if } x_i = 1 \text{ and } a'_i = 0 \\ 2^n(1/2-\lambda_1) & \text{if } x_i = 0 \text{ and } a'_i = 1 \\ 2^n\lambda_1 & \text{if } x_i = 1 \text{ and } a'_i = 1 \end{cases} \tag{8}
$$

so that $\lambda_0 + \lambda_1 = \lambda$. This allows to rewrite I_{mem} as

$$
\begin{aligned}
I_{mem} &= 2^{-n} \sum_{x_i, x'_i, \{a'_{j\neq i}\}} W_{x'_i,\{a'_j\}\to x_i} \ln \frac{\sum_{\{a'_{j\neq i}\}} W_{x'_i,\{a'_j\}\to x_i}}{2^{-1} \sum_{x'_i,\{a'_{j\neq i}\}} W_{x'_i,\{a'_j\}\to x_i}} \\
&= 2^{-n} \sum_{\{a'_{j\neq i}\}} W_{0,\{a'_j\}\to 0} \ln \frac{1/2-\lambda_0}{2^{-1}\sum_{x'_i,\{a'_{j\neq i}\}} W_{x'_i,\{a'_j\}\to 0}} \\
&+ 2^{-n} \sum_{\{a'_{j\neq i}\}} W_{0,\{a'_j\}\to 1} \ln \frac{\lambda_0}{2^{-1}\sum_{x'_i,\{a'_{j\neq i}\}} W_{x'_i,\{a'_j\}\to 1}} \\
&+ 2^{-n} \sum_{\{a'_{j\neq i}\}} W_{1,\{a'_j\}\to 0} \ln \frac{1/2-\lambda_1}{2^{-1}\sum_{x'_i,\{a'_{j\neq i}\}} W_{x'_i,\{a'_j\}\to 0}} \\
&+ 2^{-n} \sum_{\{a'_{j\neq i}\}} W_{1,\{a'_j\}\to 1} \ln \frac{\lambda_1}{2^{-1}\sum_{x'_i,\{a'_{j\neq i}\}} W_{x'_i,\{a'_j\}\to 1}} \\
&= (1/2-\lambda_0)\ln\frac{1-2\lambda_0}{1-\lambda} + \lambda_0 \ln\frac{2\lambda_0}{\lambda} + (1/2-\lambda_1)\ln\frac{1-2\lambda_1}{1-\lambda} + \lambda_1\ln\frac{2\lambda_1}{\lambda}
\end{aligned} \tag{9}
$$

The third and last piece of information we consider is *transfer*, defined as

$$
I_{trans} := I(X_i, t : X_k, t-1). \tag{10}
$$

where j is any neighbor of i different from i itself. Its computation is quite similar to that of memory and we get

$$
\begin{aligned}
I_{trans} &= \sum_{x_i, a'_k} \Pr(x_i, t; a'_k, t-1) \ln \frac{\Pr(x_i, t; a'_k, t-1)}{\Pr(x_i, t)\Pr(a'_k, t-1)} \\
&= \sum_{x_i, \{a'_j\}} \Pr(x_i, t; \{a'_j\}, t-1) \ln \frac{\sum_{\{a'_{j\neq k}\}} \Pr(x_i, t; \{a'_j\}, t-1)}{\Pr(x_i, t)\Pr(a'_k, t-1)} \\
&= 2^{-n} \sum_{x_i, \{a'_j\}} W_{\{x'_j\}\to x_i} \ln \frac{\sum_{\{a'_{j\neq k}\}} W_{\{x'_j\}\to x_i}}{2^{-1}\sum_{\{x'_j\}} W_{\{x'_j\}\to x_i}}
\end{aligned} \tag{11}
$$

We can follow the same way as previously by defining a set of parameters $(\lambda_0^{(L)}, \lambda_1^{(L)}, \lambda_0^{(R)}, \lambda_1^{(R)})$ where the superscript index L or R refers to the left of right neighbor respectively. In other words, $\lambda_0^{(L)}$ is the fraction of ones in the lookup table when the left neighbor is zero, *etc.* As for memory we obviously have $\lambda_0^{(L)} + \lambda_1^{(L)} = \lambda$ and $\lambda_0^{(R)} + \lambda_1^{(R)} = \lambda$. The reasoning exposed for the memory information may be carried over without modification so as to get finally

$$
\begin{aligned}
I_{trans}^{(L)} &= (1/2-\lambda_0^{(L)})\ln\frac{1-2\lambda_0^{(L)}}{1-\lambda} + \lambda_0^{(L)}\ln\frac{2\lambda_0^{(L)}}{\lambda} \\
&+ (1/2-\lambda_1^{(L)})\ln\frac{1-2\lambda_1^{(L)}}{1-\lambda} + \lambda_1^{(L)}\ln\frac{2\lambda_1^{(L)}}{\lambda}
\end{aligned} \tag{12}
$$

$$
\begin{aligned}
I_{trans}^{(R)} &= (1/2-\lambda_0^{(R)})\ln\frac{1-2\lambda_0^{(R)}}{1-\lambda} + \lambda_0^{(R)}\ln\frac{2\lambda_0^{(R)}}{\lambda} \\
&+ (1/2-\lambda_1^{(R)})\ln\frac{1-2\lambda_1^{(R)}}{1-\lambda} + \lambda_1^{(R)}\ln\frac{2\lambda_1^{(R)}}{\lambda}
\end{aligned} \tag{13}
$$

The case of Rule 110 As an example consider Rule 110, which is an elementary cellular automaton which is well known for its complex behavior. The lookup table of this dynamics is given by

$$W_{x'_{i-1}x'_i x'_{i+1}\to x_i} = \begin{array}{c|cc|} & 0 & 1 \\ \hline 000 & 1 & 0 \\ 001 & 0 & 1 \\ 010 & 0 & 1 \\ 011 & 0 & 1 \\ 100 & 1 & 0 \\ 101 & 0 & 1 \\ 110 & 0 & 1 \\ 111 & 1 & 0 \end{array} \tag{14}$$

One immediately gets $\lambda = 5/8$, $\lambda_0 = \lambda_1^{(L)} = \lambda_0^{(R)} = 1/4$ and $\lambda_1 = \lambda_0^{(L)} = \lambda_1^{(R)} = 3/8$. It results that

$$I_{tot}^{110} = -\frac{5}{8}\ln\frac{5}{8} - \frac{3}{8}\ln\frac{3}{8} \approx 0.66156 \tag{15}$$

$$\begin{aligned} I_{mem} &= (1/2-\lambda_0)\ln\frac{1-2\lambda_0}{1-\lambda} + \lambda_0 \ln\frac{2\lambda_0}{\lambda} + (1/2-\lambda_1)\ln\frac{1-2\lambda_1}{1-\lambda} + \lambda_1\ln\frac{2\lambda_1}{\lambda} \\ &= \frac{1}{4}\ln\frac{4}{3} + \frac{1}{4}\ln\frac{4}{5} + \frac{1}{8}\ln\frac{2}{3} + \frac{3}{8}\ln\frac{6}{5} \approx 0.03382 \end{aligned} \tag{16}$$

$$\begin{aligned} I_{trans}^{(L)} &= (1/2-\lambda_0^{(L)})\ln\frac{1-2\lambda_0^{(L)}}{1-\lambda} + \lambda_0^{(L)}\ln\frac{2\lambda_0^{(L)}}{\lambda} \\ &\quad + (1/2-\lambda_1^{(L)})\ln\frac{1-2\lambda_1^{(L)}}{1-\lambda} + \lambda_1^{(L)}\ln\frac{2\lambda_1^{(L)}}{\lambda} \\ &= \frac{1}{8}\ln\frac{2}{3} + \frac{3}{8}\ln\frac{6}{5} + \frac{1}{4}\ln\frac{4}{3} + \frac{1}{4}\ln\frac{4}{5} \approx 0.03382 \end{aligned} \tag{17}$$

$$\begin{aligned} I_{trans}^{(R)} &= (1/2-\lambda_0^{(R)})\ln\frac{1-2\lambda_0^{(R)}}{1-\lambda} + \lambda_0^{(R)}\ln\frac{2\lambda_0^{(R)}}{\lambda} \\ &\quad + (1/2-\lambda_1^{(R)})\ln\frac{1-2\lambda_1^{(R)}}{1-\lambda} + \lambda_1^{(R)}\ln\frac{2\lambda_1^{(R)}}{\lambda} \\ &= \frac{1}{4}\ln\frac{4}{3} + \frac{1}{4}\ln\frac{4}{5} + \frac{1}{8}\ln\frac{2}{3} + \frac{3}{8}\ln\frac{6}{5} \approx 0.03382 \end{aligned} \tag{18}$$

Discussion about the Langton parameter relation So far we have proved that the basic quantities of information processing could be expressed in terms of generalized Langton parameters. This is nevertheless true only when in the initial state the nodes are completely uncorrelated. When the nodes are initially correlated, the values are significantly shifted. Figs. 1, 2, 3 and 4 show the behavior of the total, memory and left- and right-transfer information respectively after the nodes have become correlated. The dotted line pictures the value of each information quantity in the uncorrelated state, highlighting the shift between this value and the corresponding value in the stationary regime.

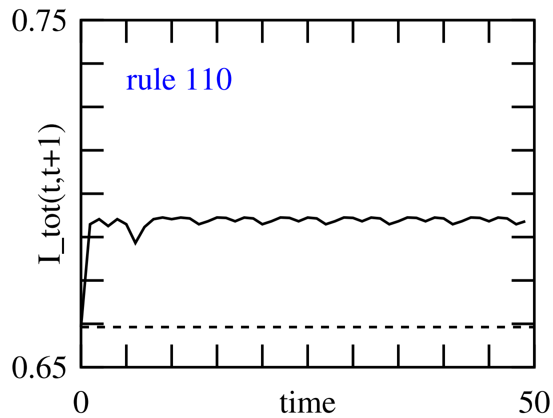

Figure 1: Total information evaluated for rule 110 over 50 time steps, starting from an uncorrelated initial state.

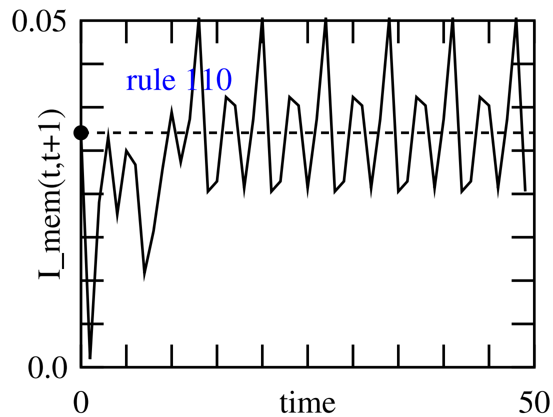

Figure 2: Memory information evaluated for rule 110 over 50 time steps, starting from an uncorrelated initial state.

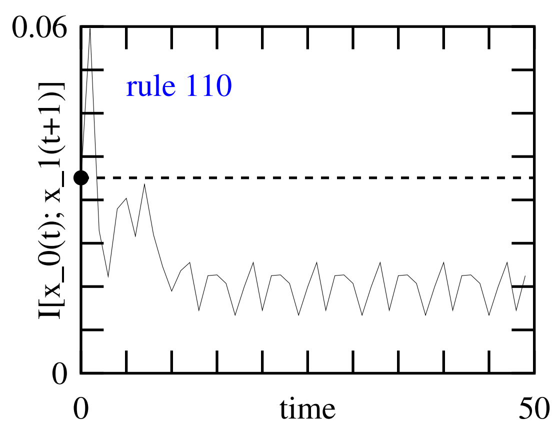

Figure 3: Left-transfer information evaluated for rule 110 over 50 time steps, starting from an uncorrelated initial state.

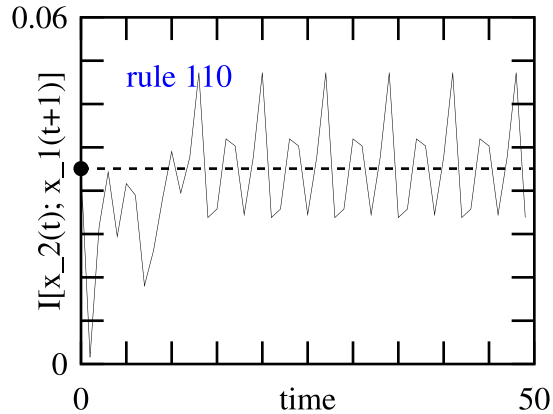

Figure 4: Right-transfer information evaluated for rule 110 over 50 time steps, starting from an uncorrelated initial state.

Robustness of financial cube plots

Different sliding window size for computing mutual information. Here we show information feature plots for the financial time series (Fig. 4 and Fig. 5 in the main text) for varying sliding window size ($w = 1400$ data points in main text). The lower the sliding window size, the fewer datapoints and thus the more difficult it is to accurately estimate mutual information, but the higher temporal resolution is obtained. A larger sliding window leads to a coarser temporal resolution but to a better estimate. We show here $w = 1000$ and $w = 2000$ in Figs. 5 through 8.

Alternative calculation of the instability indicator

An alternative calculation of the instability indicator ω would be to project all points to the 'trend' vector $\vec{x}_t - \vec{x}_{t-\Delta t}$ and computing the average perpendicular distance of the points to their projections. This results in almost indistinguishable indicator curves, as shown in Fig. 9.

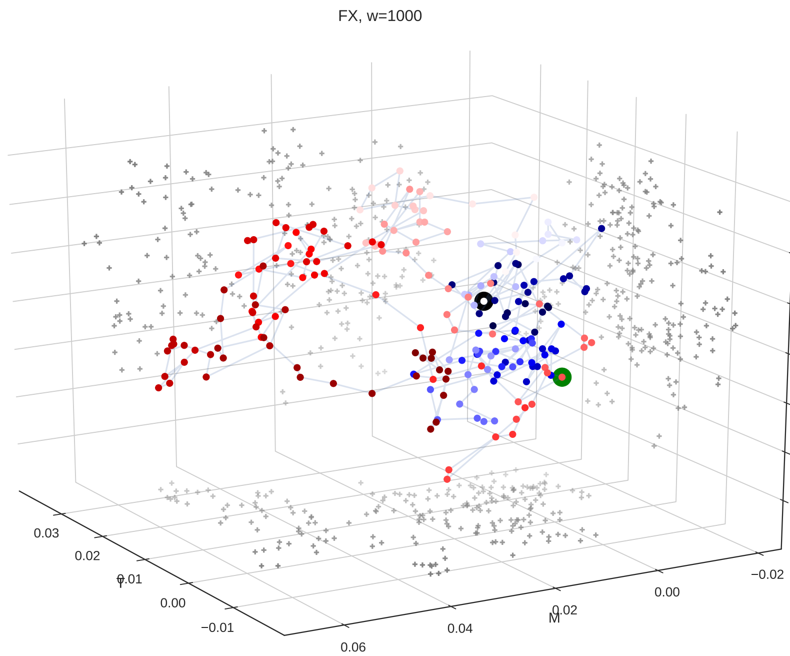

Figure 5: 200 time points showing the progression of the three information features memory (M), transfer (T), and synergy (S) computed with a time delay of 1 day (similar to $t = 1$ for ECA). The color indicates the time difference with September 15, 2008 (big black dot), which we consider the starting point of the 2008 crisis, from dark blue (long before) to dark red (long after) and white at the crisis date. The data spans from 1999-01-01 through 2017-04-21; the large green dot is the last time point also present in the IRS data in 2011. Mutual information is calculated using a sliding window size of $w = 1000$ days; the 200 windows partially overlap and are placed uniformly over the dataset, where the first and last window include the first and last day of the dataset, resp.

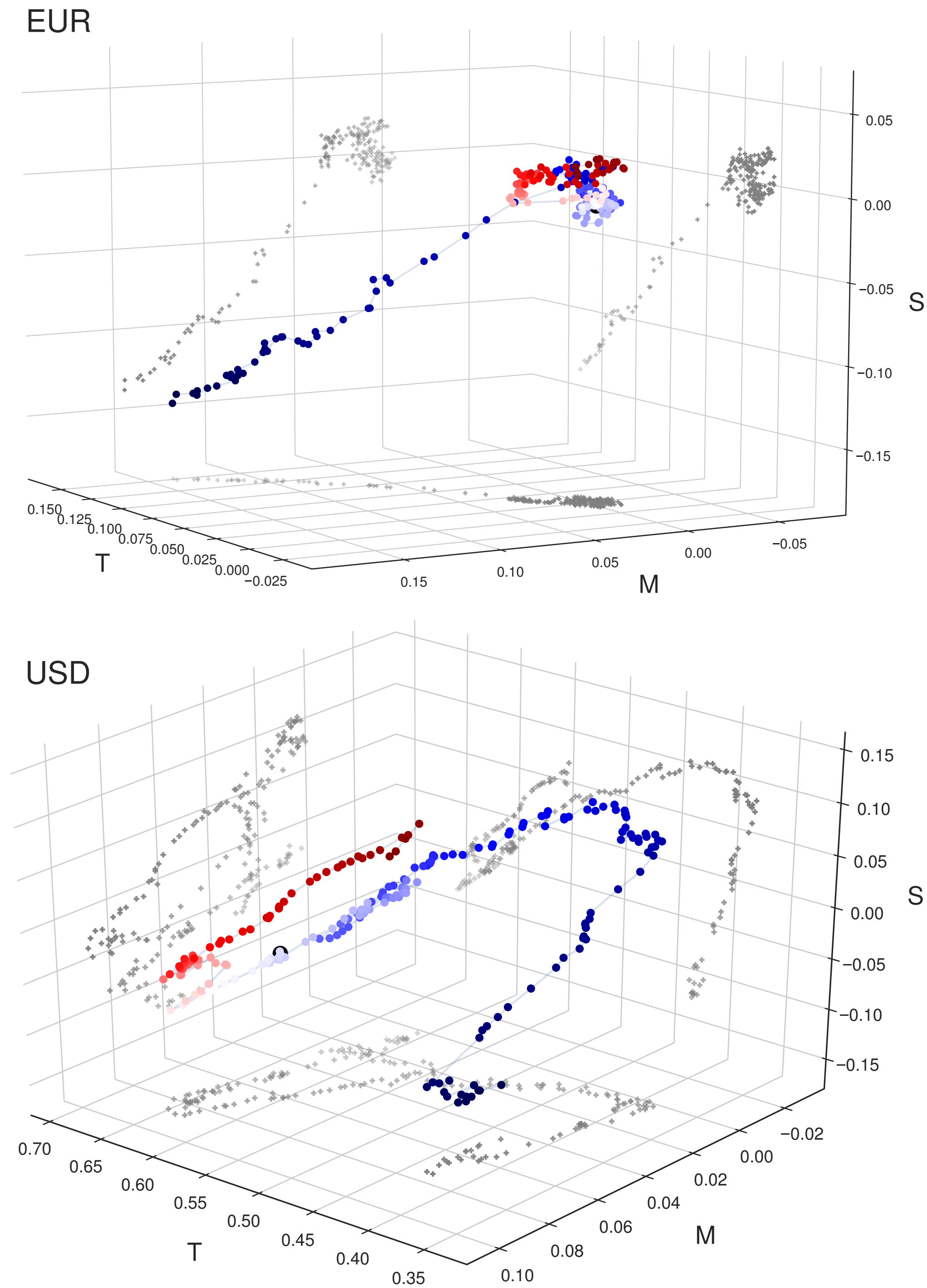

Figure 6: 200 time points showing the progression of the three information features memory (M), transfer (T), and synergy (S) computed with a time delay of 1 day (similar to $t = 1$ for ECA). The color indicates the time difference with September 15, 2008 (big black dot), which we consider the starting point of the 2008 crisis, from dark blue (long before) to dark red (long after) and white at the crisis date. The data spans more than twelve years: the EUR data from 1998-01-12 to 2011-08-12 and the USD data from 1999-29-04 to 2011-06-06. Mutual information is calculated using a sliding window of $w = 1000$ days; the 200 windows partially overlap and are placed uniformly over the dataset, where the first and last window include the first and last day of the dataset, resp.

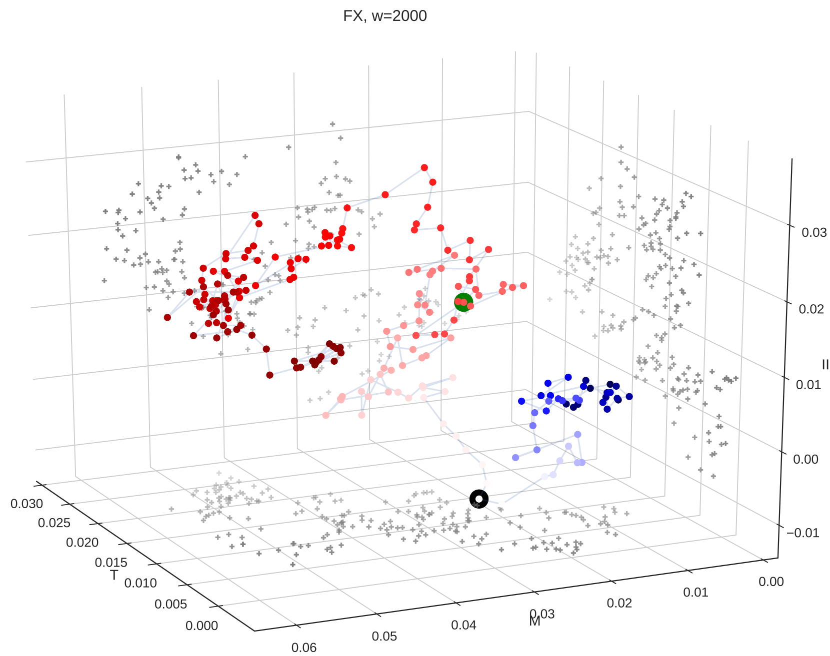

Figure 7: 200 time points showing the progression of the three information features memory (M), transfer (T), and synergy (S) computed with a time delay of 1 day (similar to $t = 1$ for ECA). The color indicates the time difference with September 15, 2008 (big black dot), which we consider the starting point of the 2008 crisis, from dark blue (long before) to dark red (long after) and white at the crisis date. The data spans from 1999-01-01 through 2017-04-21; the large green dot is the last time point also present in the IRS data in 2011. Mutual information is calculated using a sliding window of $w = 2000$ days; the 200 windows partially overlap and are placed uniformly over the dataset, where the first and last window include the first and last day of the dataset, resp.

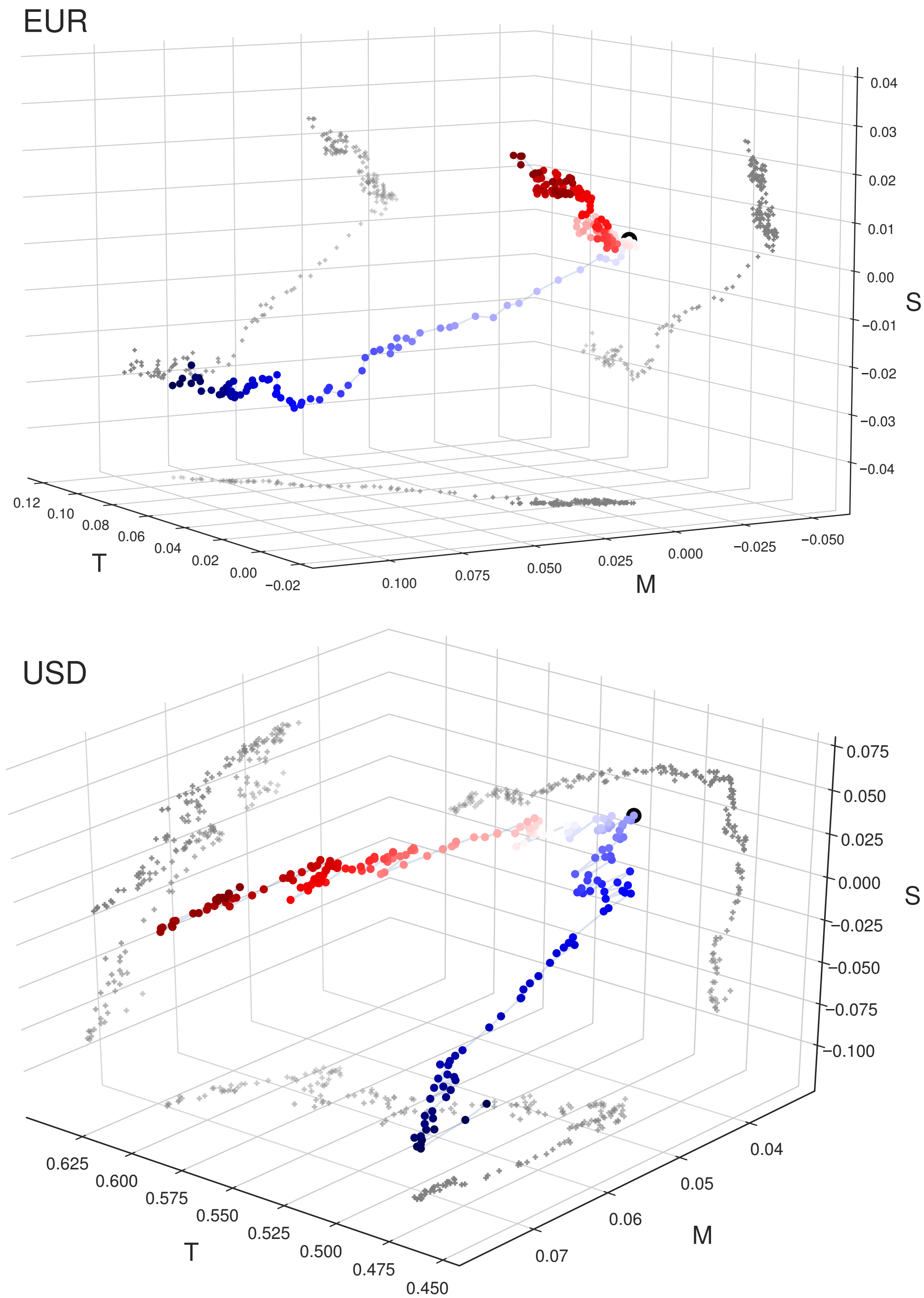

Figure 8: 200 time points showing the progression of the three information features memory (M), transfer (T), and synergy (S) computed with a time delay of 1 day (similar to $t = 1$ for ECA). The color indicates the time difference with September 15, 2008 (big black dot), which we consider the starting point of the 2008 crisis, from dark blue (long before) to dark red (long after) and white at the crisis date. The data spans more than twelve years: the EUR data from 1998-01-12 to 2011-08-12 and the USD data from 1999-29-04 to 2011-06-06. Mutual information is calculated using a sliding window of $w = 2000$ days; the 200 windows partially overlap and are placed uniformly over the dataset, where the first and last window include the first and last day of the dataset, resp.

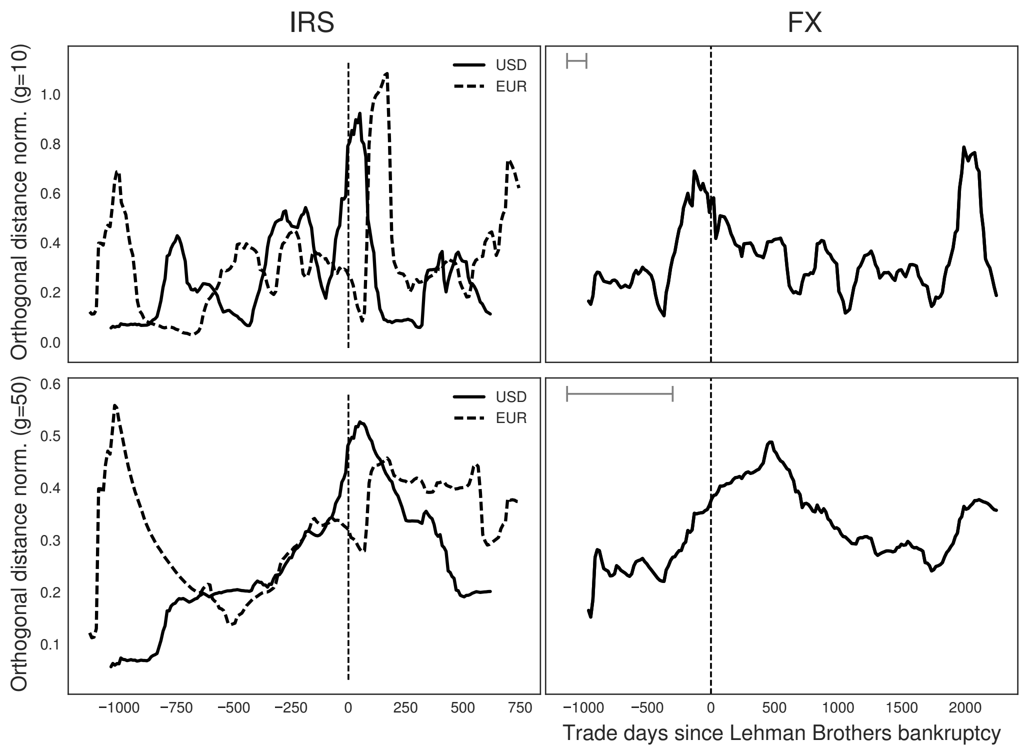

Figure 9: Indicator curves as in main text's Fig. 6 but using perpendicular distances to the trend vector as alternative to Eq. 9.