Supplementary Material to *Early Warning Signals of Complex Critical Transitions in Deterministic Dynamics*

Methods

Regime boundary detection algorithm

In order to determine where regimes such as a chaotic regime end and start, the periodicity of the timeseries needs to be determined for each value of the bifurcation parameter. Ideally, the exact moment of transition is found analytically, yet for complex bifurcations whose timing depends sensitively on the initial condition and rate of parameter change, this is not possible [86]. We thus resort to numerical methods.

Given that most EWS studies into complex bifurcations rely on only one or a few simulations [62, 65, 66], no systematic way of finding regime boundaries seems to be used in the literature. For some bifurcations it is possible to find the critical transition point analytically, but this will not correspond to the simulated transition under intrinsic noise. Empirical studies into transitions may rely on change point analysis [147, 148], which detects a shift in statistical properties such as the mean or variance. Though this may work well for simple bifurcations, these methods are not well suited for complex bifurcations, where the difference between regimes may be subtle (e.g. period-doubling bifurcations) and differences within the regime are substantial. Given the lack of guidance in the literature, we thus developed our own algorithm for detecting regime boundaries.

Algorithm 1: Regime boundary detection

Data: Timeseries with step-wise changes in bifurcation parameter range s
Result: Regime boundaries

Find peaks and troughs in timeseries
for *each step* s_i *in the bifurcation parameter range* s **do**
  Get peaks and troughs $p = p_j, \ldots, p_N$ corresponding to step s_i
  for *period* $k = 1, \ldots, K$ **do**
    Group p according to group indices $g = j \mod k$ with $j = 1, ..., N$
    Compute Euclidean distance $D_j = \|p_j\|$ for each group g_j
    Set fit f_k as maximum distance $\max(D)$ across groups
  end
  if *Distance between fits is less than* δ_{node} **then**
    Set behaviour as node
  else
    Choose best k based on best fit f^* weighted by δ_k
    if *Best fit* $f^* > \delta_{ampl}$ *&* $f^* > \delta_{idx}$ **then**
      Set behaviour as chaotic
    else
      Set behaviour as period k
    end
  end
end
Smooth over exceptions in behaviour (δ_{smooth})
Flag steps which touch basin boundary $[x_{\min}, x_{\max}]$
Flag chaotic behaviour where the spread of peaks and troughs covers an area larger than δ_{band}
Find regimes by grouping consecutive sequences in behaviour which are of length $l \geq \delta_l$

Rather than analysing the complete timeseries, a more efficient method finds the periodicity by using the amplitude and timing of the timeseries' peaks and troughs [149]. A first intuitive approach to finding the timeseries' periodicity may be to count the number of distinct peak coordinates using a histogram. However, this only works for simple oscillations with clearly separable peak coordinates which do not repeat. For instance, an oscillation with a pattern $[a, b, a, c]$ is not the same as $[a, b, c]$, yet a histogram will depict both as a period-3 oscillation. The order of the peaks thus needs to be included to accurately identify the period, which a histogram discards.

Incorporating a temporal aspect, a second approach to determining periodicity may be to find peaks in the autocorrelation function. However, repeated peak coordinates in the oscillation again make such an approach difficult. The first local maximum in the autocorrelation does not correspond to the period in patterns such as $[a, b, a, c]$, where the autocorrelation will peak at lag 2 and more strongly at lag 4. Oscillations with a long period and subtle differences between peak amplitudes are again difficult to pick up.

The present paper used a third, more computationally expensive approach because of its greater accuracy over the former two approaches. The period k was determined by assessing which k had the best fit when the peaks and troughs are partitioned according to their position in the period. For instance, to test whether the timeseries is of period $k = 3$, its peaks and troughs y_i are grouped using $i \mod k$, resulting in a grouping sequence $g = 1, 2, 3, 1, 2, 3, \ldots$ which correspond to indices in an oscillation of period $k = 3$ (Algorithm 1, line 5). For each group g, the (Euclidean) distance between peaks is computed (Algorithm 1, line 6). If the timeseries is indeed period-3, the distance for all groups $y_{g=1}, y_{g=2}, y_{g=3}$ should be minimal. The maximum distance across groups is taken as an indicator of fit (Algorithm 1, line 7), such that the best fit (i.e. the smallest maximum distance) across all $k = 1, \ldots, K$ yields the period of the signal. This procedure can be completed for both the amplitudes of the peaks and troughs (i.e. coordinates) as well as their timing (i.e. indices).

Completing this procedure for all timeseries per step in the bifurcation parameter gives the evolving order of periodicity across the bifurcation range. Some parameters were added to account for several weakness. Firstly, a timeseries was identified as a node if all fits were the same within some margin of error δ_{node} (Algorithm 1, line 10). Secondly, as multiples of the true period will lead to a better fit simply because each group g will have fewer peaks resulting in smaller distances, a penalty δ_k is added for higher k (Algorithm 1, line 12). This ensures that a higher k needs to result in a better fit than lower k. Thirdly, as the algorithm will always find a best-fitting k even when the timeseries is chaotic, behaviour was defined as chaotic if the best fit exceeded the maximum thresholds $\delta_{\text{ampl}}, \delta_{\text{idx}}$ for the allowed maximum distance in the amplitude and timing of the peak (Algorithm 1, line 13). Finally, the maximum period K to look for is defined as the maximum period which still repeats at least once given the length of the timeseries N: $K = \text{floor}(N/2)$.

To add more granularity to the regimes, we distinguish between regimes which touch the basin boundary and those that do not, as well as between different types of chaos. More fine-grained behaviours help to identify more critical transitions. To help detect boundary crises, the algorithm flags timeseries of which a chosen variable reaches a specified minimum $x_{\min}$ and maximum $x_{\max}$ (Algorithm 1, line 21). These correspond to the edges of the basin that are touched right before a boundary crisis happens, which are $x_{\min} = 0$ and $x_{\max} = 1$ for variable x_1 and x_2, in our case. Secondly, to help detect interior crises, the algorithm separates chaotic timeseries that have well separated chaotic bands and those that cover the entire area between the lowest and highest value (δ_{band}, Algorithm 1, line 22).

The result of this procedure yields the type of behaviour corresponding to each step in the bifurcation parameter. To finally identify regime boundaries across the parameter range, we defined regimes as a consecutive sequence of the same behaviour of a minimal length δ_l which may be interrupted by behaviours of length less than δ_l (Algorithm 1, line 23). This ensured that brief outliers given some flaws in the algorithm or the parameter settings were not seen as transition points, which may also be smoothed using δ_{smooth} (Algorithm 1, line 20). Though the algorithm of course has its weaknesses (e.g. too high periodicity is classified as chaos if only short timeseries are available), it detected regime boundaries quite

well as assessed visually (Supplementary Figures S1-S3). Even for limit cycles with growing amplitude across the bifurcation range, in which case peak coordinates do not match, the algorithm identified the whole stretch as belonging to the same cycle.

Simulation details

Default parameter settings for the regime boundary detection algorithm which yielded good performance as found via experimentation may be found in Table 5. Some deviations from the defaults needed to be made in order to either obtain the desired type of critical transition, or simply to save computational power and storage space (Table 6). For instance, as the default method was not able to generate a Boundary Crisis, the time step and solver were changed to $\Delta_t = .1$ and Euler integration only for this bifurcation type.

Table 5 Default parameter settings of regime boundary detection algorithm

Parameter	Default value	Meaning
δ_k	.1	Penalty term to fit for higher k
δ_{smooth}	0	Number of steps in bifurcation parameter to smooth regimes over
δ_{node}	.001	Maximum difference between fit for each k below which timeseries is classified as a node
δ_{ampl}	.025	Distance between peak amplitudes above which timeseries is classified as chaotic
δ_{idx}	2	Distance between peak indices above which timeseries is classified as chaotic
δ_{band}	.85	Maximum percentage of area covered between minimum and maximum value to be flagged as separated (cf. merged) chaotic bands
δ_l	.1	Minimum length for a consecutive sequence of the same behaviour to be classified as a regime
$x_{\min}$	0	Minimum edge of basin boundary
$x_{\max}$	1	Maximum edge of basin boundary

Table 6 Details of simulation parameters per bifurcation

s	Δ_s	Bifurcation	Changes to default settings
$[1.15, .9]$	-.0001	Saddle-node	$\delta_k = 1; T = 100, N_{\text{noise}} = 250; N_{\text{sim}} = 1$
$[.6, .85]$	.001	Hopf (Period 1 to 2)	$\delta_k = 1; T = 200; N_{\text{noise}} = 250; N_{\text{sim}} = 1$
$[.85, .985]$	.0001	Period 2 to 4	$\delta_l = 8$
		Period 4 to 8	
		Period 8 to 16	
		Mixed-Periodic to Chaotic	
$[.985, .8]$	-.0001	Chaotic to Mixed-Periodic	$\delta_k = .01; \delta_{\text{ampl}} = .01$
		Period 16 to 8	
		Period 8 to 4	
		Period 4 to 2	
		Reverse Hopf (Period 2 to 1)	
$[1.01, 1.02]$	.00005	Interior Crisis: Chaotic Expansion	$\delta_{\text{smooth}} = 5; \delta_l = 1$
$[1.03, 1.01]$	-.0001	Interior Crisis: Chaotic Reduction	$\delta_k = 1; \delta_{\text{smooth}} = 10; \delta_{\text{ampl}} = .01$
$[1.1, 1.3]$	.001	Boundary Crisis	$\Delta_t = .1; \delta_l = 20$; Euler simulation

Algorithm 2: Data simulation and EWS analysis

Result: Performance as measured by AUC per EWS per bifurcation type

for *each range* s **do**
  `// Simulate transition and null models and compute EWS`
  for *each simulation in* $1 : N_{sim}$ **do**
    Generate timeseries of a GLV forced through range s
    Apply regime boundary detection algorithm to find regime boundaries $[s_1, s_2]$ and initial condition x_0 of desired critical transitions
    for *each desired critical transition* **do**
      for *transition and null model* **do**
        Generate timeseries using $[s_1, s_2]$ and x_0
        for *each downsampling frequency* f_s **do**
          for *each observational noise intensity* σ_{obs} **do**
            for *each noise iteration in* $1 : N_{noise}$ **do**
              Distort timeseries by downsampling to f_s and adding observational noise σ_{obs}
              for *each step in* $[s_1, s_2]$ **do**
                Compute all EWS on distorted timeseries
              end
            end
          end
        end
      end
    end
  end
  `// Compute performance`
  for *each desired critical transition* **do**
    for *each metric in EWS* **do**
      for *each downsampling frequency* f_s **do**
        for *each observational noise intensity* σ_{obs} **do**
          Compute ROC curve by finding the number of false positives in null models and true positives in transition models for each critical cut-off value σ_{crit} until $\text{FPR} = 0$ and $\text{TPR} = 0$
          Integrate ROC curve to obtain AUC
          Find the optimal critical cut-off value σ^*_{crit} corresponding to the maximum Youden's J statistic, $J = \text{TPR} - \text{FPR}$
          Save warnings in transition models when setting σ^*_{crit}
        end
      end
      `// Summarise across conditions`
      Compute median timing of saved warning signals
      Assess whether saved warnings either all fell above, all below, or both above and below the confidence band
    end
  end
end

Note that to fairly evaluate which bifurcation types are able to be anticipated, we chose to have same amount of data for each bifurcation. However, as some bifurcations are inherently faster than others (e.g.

a period-4 to period-8 bifurcation is intrinsically shorter than a period-2 to period-4 bifurcation because of Feigenbaum's constant), the step size Δ_s in the bifurcation parameter could not be held constant. That is, the transition and null models of each bifurcation type had the same number of steps in the bifurcation parameter, the same number of timepoints for each step, the same time step, and the same sampling frequency, but had different Δ_s.

Bifurcation diagrams

The bifurcation diagrams of the all variables for $s \in [.6, 1.3]$, $s \in [1.3, .6]$ (reverse order of s), and of the saddle-node bifurcation are shown in Figures S1-S3.

Chaotic features: Intermittency

A typical feature within chaotic dynamics is *intermittency*: recurrent switches between two or more attractors. Intermittent switches may occur between a periodic and chaotic attractor (Supplementary Figure S4, left), or between two chaotic attractors (Supplementary Figure S4, right). Intermittency in a subduction is related to the traditional Pomeau–Manneville dynamics intermittency [101], whereas interior crises show crisis-induced intermittency [100].

Importantly, these intermittent switches change in character as a bifurcation point is approached (Supplementary Figure S4, bottom). For instance, the intermittency route to chaos contains short bursts of chaotic behaviour amidst regularity, which increase in length and frequency until the system fully occupies the chaotic attractor after the bifurcation point. Approached from the other direction, a dominant chaotic attractor may show brief bursts of regularity of increasing length and frequency as the bifurcation point is approached. Though the study of how these changes in intermittency express themselves in warning signs is beyond the scope of this paper, they offer a potential warning sign for chaotic bifurcations.

Results

Performance per condition (downsampling frequency $f_s \in [10, 1, .1]$ and observational noise intensity $\sigma_{\text{obs}} \in [0.0001, .02, .04]$) for each critical transition is shown in Figures S5-S10. The direction of EWS when setting an optimal cut-off value σ^*_{crit} using Youden's J statistic is shown in Figure S11. Finally, some examples of complex warning patterns are illustrated in Figure S12.

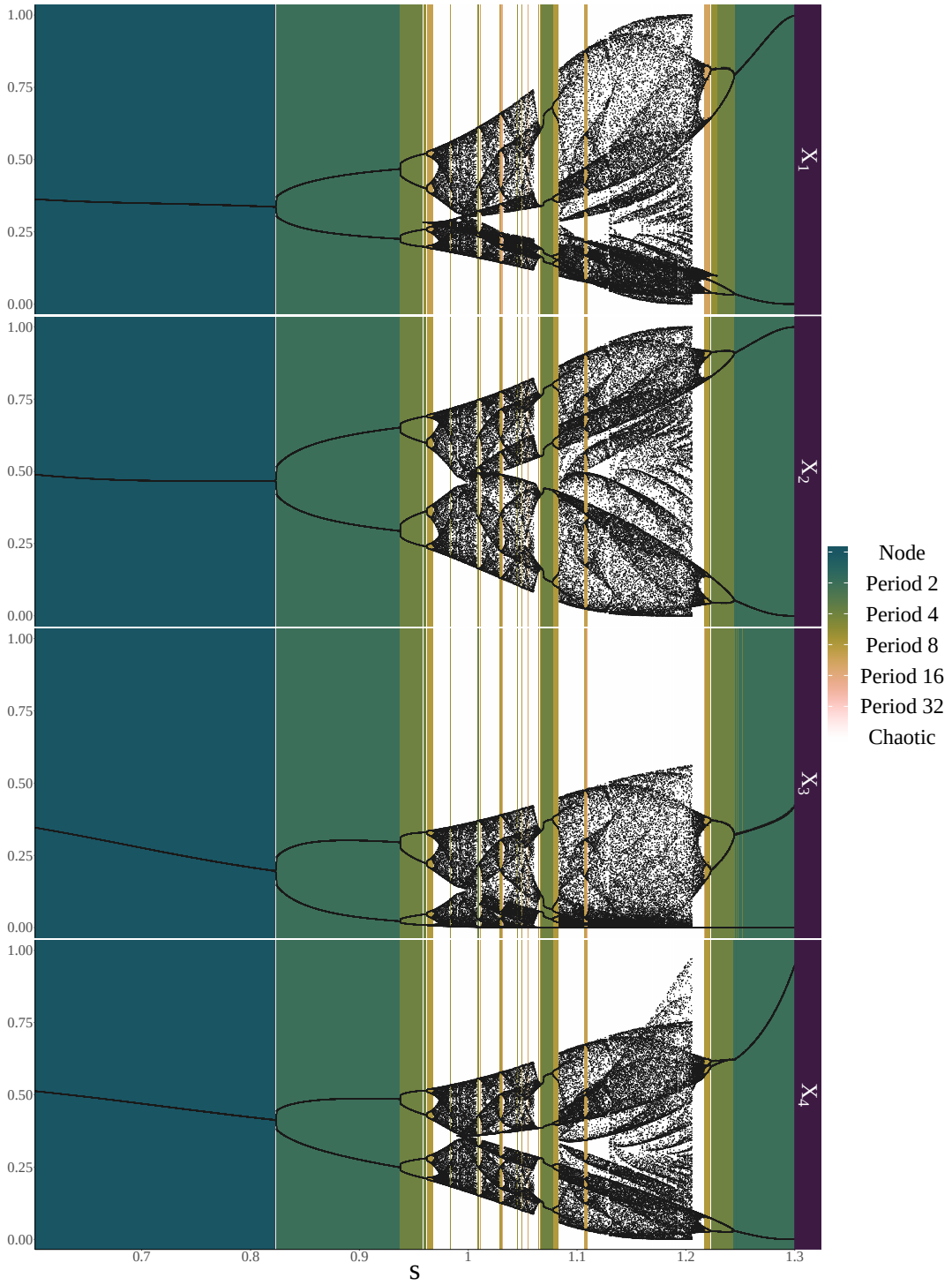

Fig. S1 Numerically obtained bifurcation diagram of all variables with range $s \in [.6, 1.3]$. Bifurcation diagram of all variables in the GLV showing the peaks and troughs (i.e. local minima and maxima, indicated in black) corresponding to each value of the bifurcation parameter s (x-axis). Coloured bars indicate the periodicity of the timeseries as found using our regime boundary detection algorithm, where white indicates chaotic regimes. Note that transients are included in the diagram to show the full transitional process

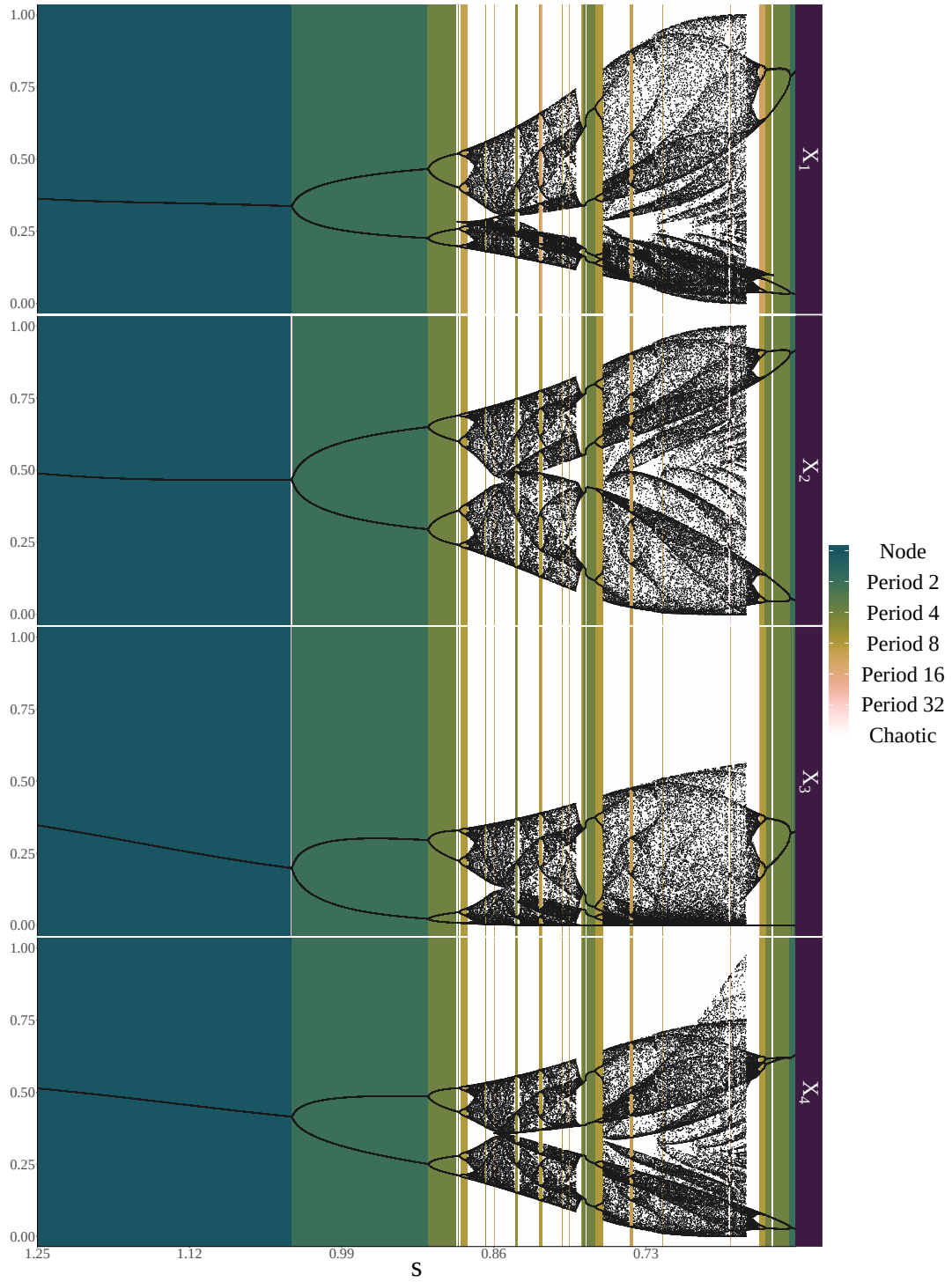

Fig. S2 Numerically obtained bifurcation diagram of all variables with range $s \in [1.25, .6]$. Bifurcation diagram of all variables in the GLV showing the peaks and troughs (i.e. local minima and maxima, indicated in black) corresponding to each value of the bifurcation parameter s (x-axis). Coloured bars indicate the periodicity of the timeseries as found using our regime boundary detection algorithm, where white indicates chaotic regimes. Note that transients are included in the diagram to show the full transitional process

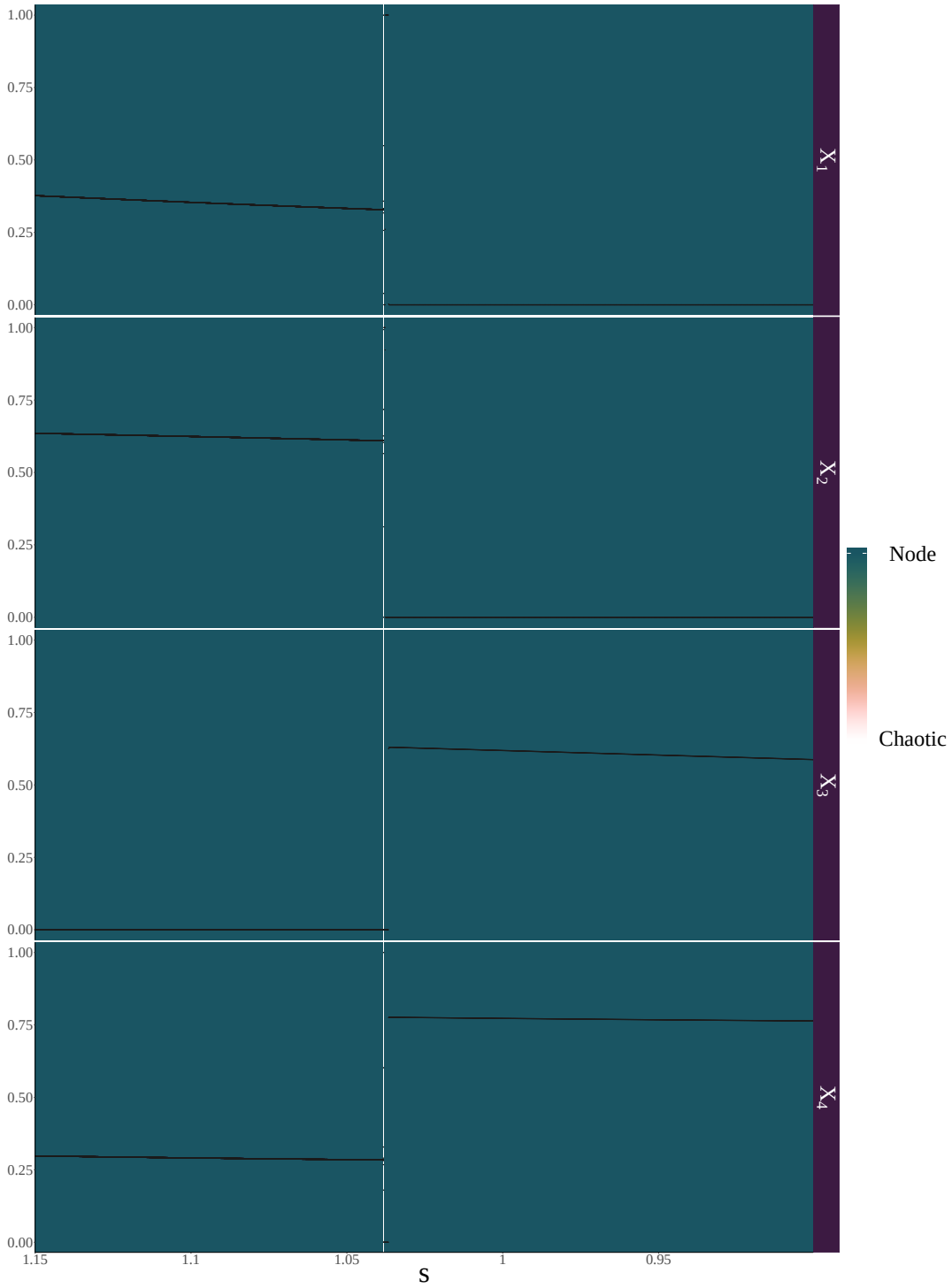

Fig. S3 Numerically obtained bifurcation diagram of all variables with range $s \in [1.15, .9]$. Bifurcation diagram of all variables in the GLV showing the peaks and troughs (i.e. local minima and maxima, indicated in black) corresponding to each value of the bifurcation parameter s (x-axis). Coloured bars indicate the periodicity of the timeseries as found using our regime boundary detection algorithm, where white indicates chaotic regimes. Note that transients are included in the diagram to show the full transitional process

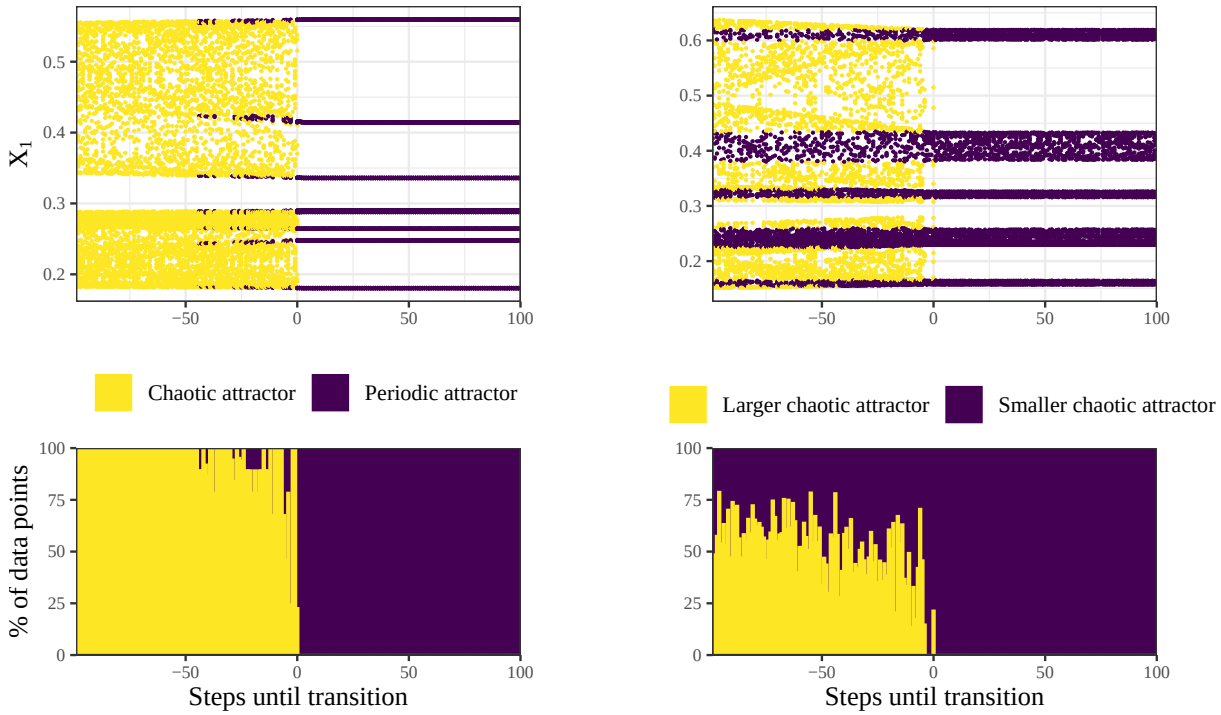

Fig. S4 Illustration of intermittency between a periodic and chaotic attractor (*top, left*), or between two chaotic attractors (*top, right*). The new attractor (purple, after the transition) is already embedded within the current attractor (yellow and purple, before the transition) before the bifurcation occurs. The time spent in the new attractor increases approaching the bifurcation (*bottom*)

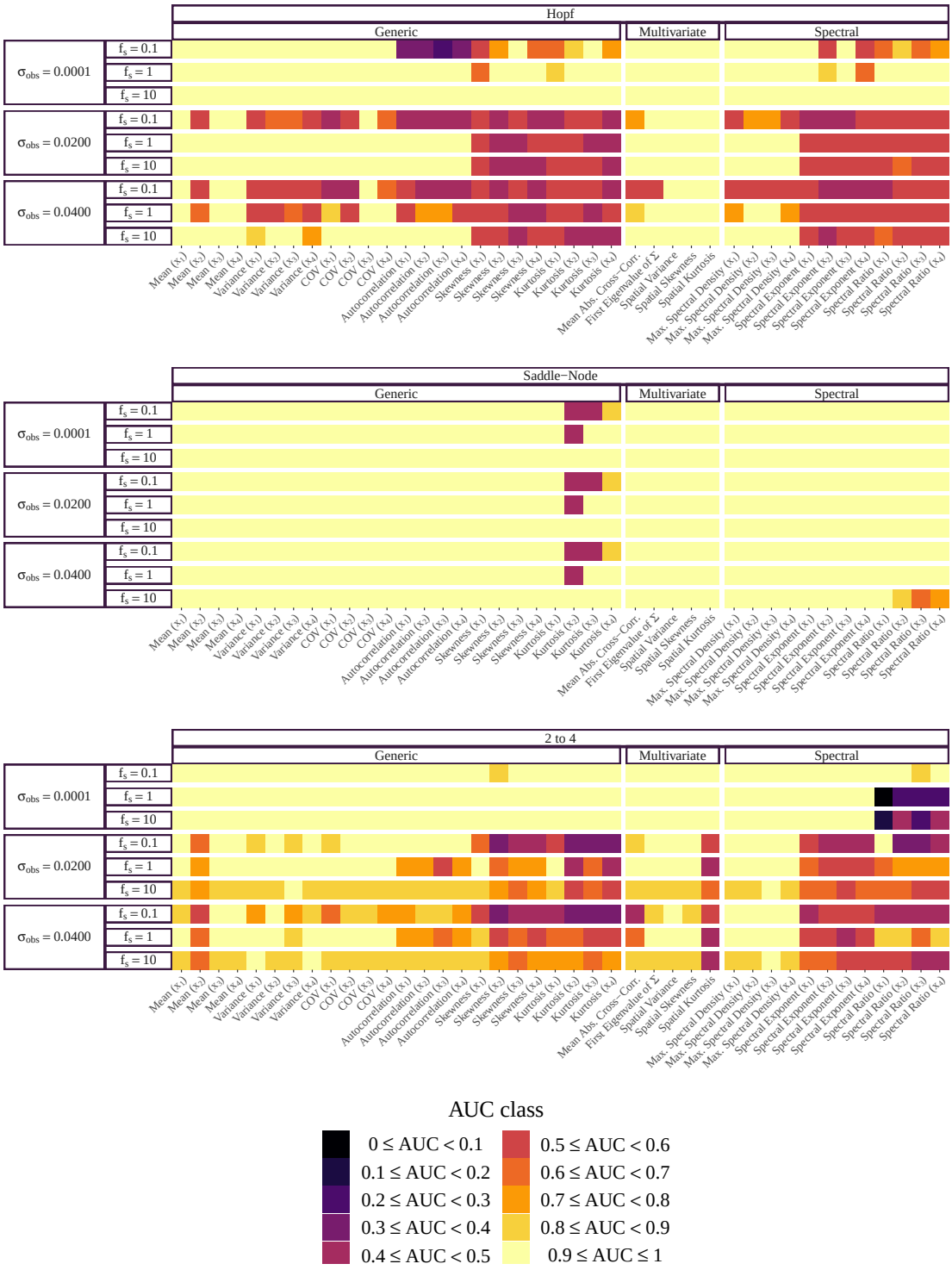

Fig. S5 AUC per condition (brighter, yellow colours indicate better performance, whereas darker, purple colours indicate worse performance)

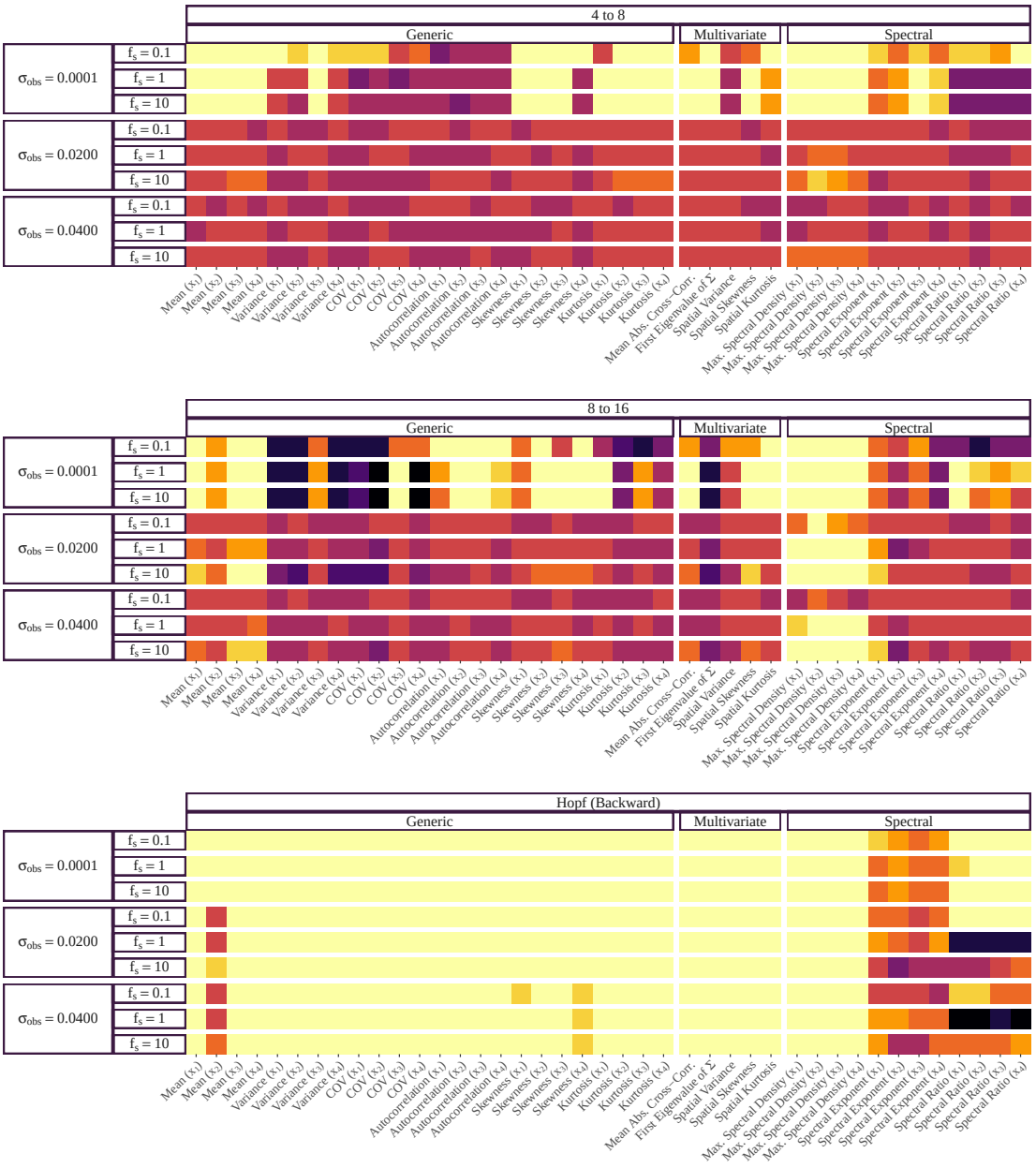

Fig. S6 AUC per condition (brighter, yellow colours indicate better performance, whereas darker, purple colours indicate worse performance)

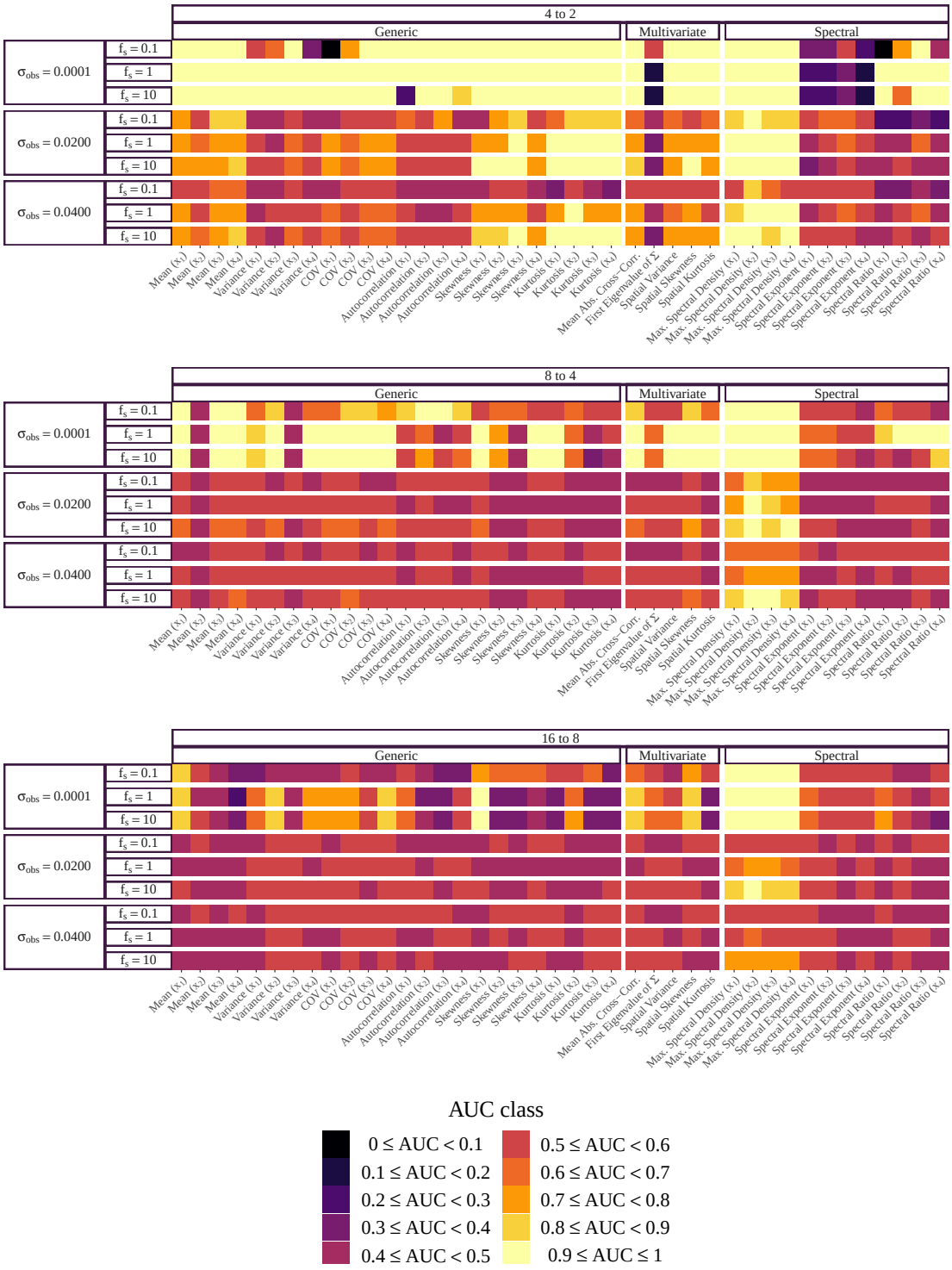

Fig. S7 AUC per condition (brighter, yellow colours indicate better performance, whereas darker, purple colours indicate worse performance)

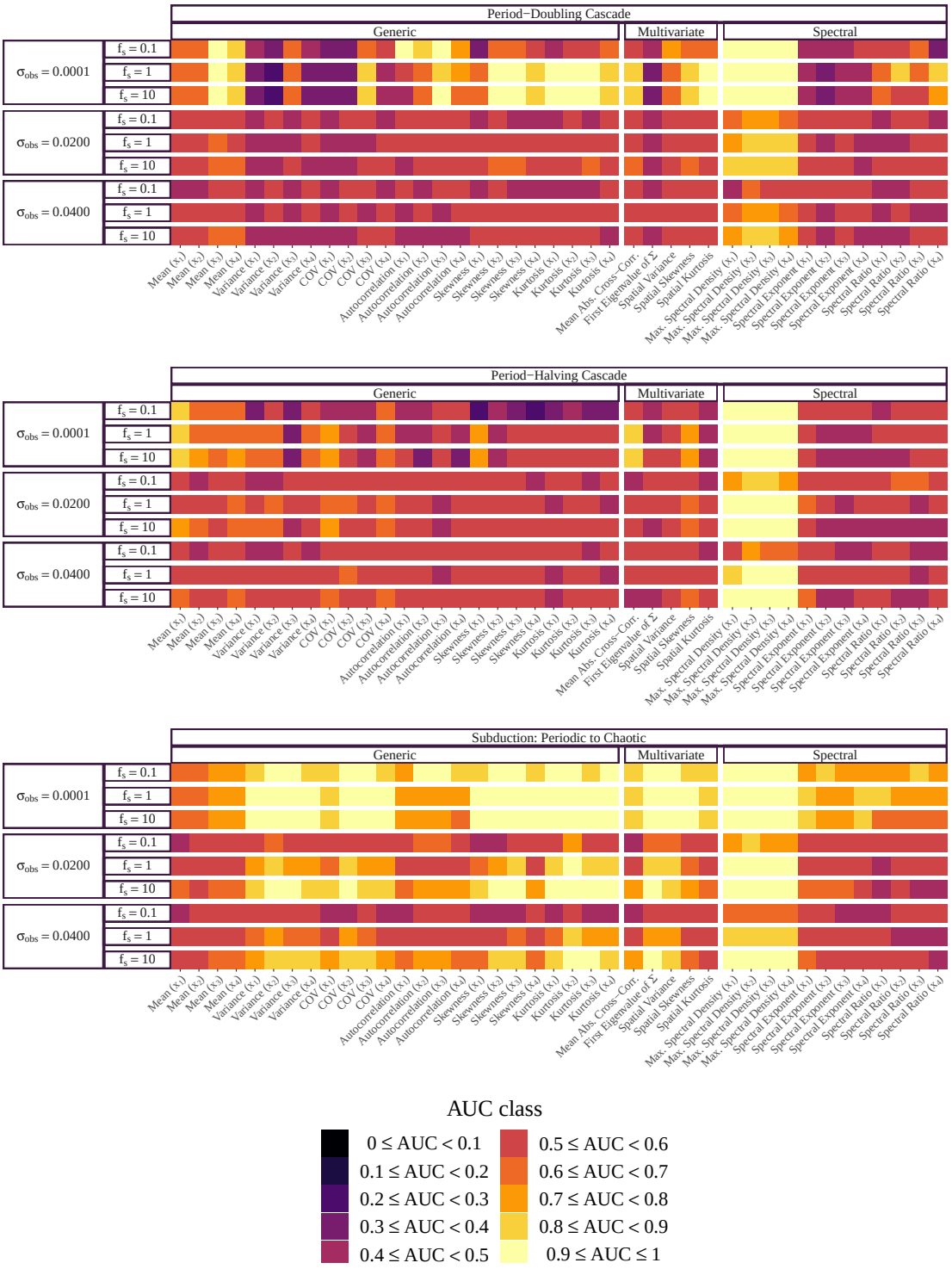

Fig. S8 AUC per condition (brighter, yellow colours indicate better performance, whereas darker, purple colours indicate worse performance)

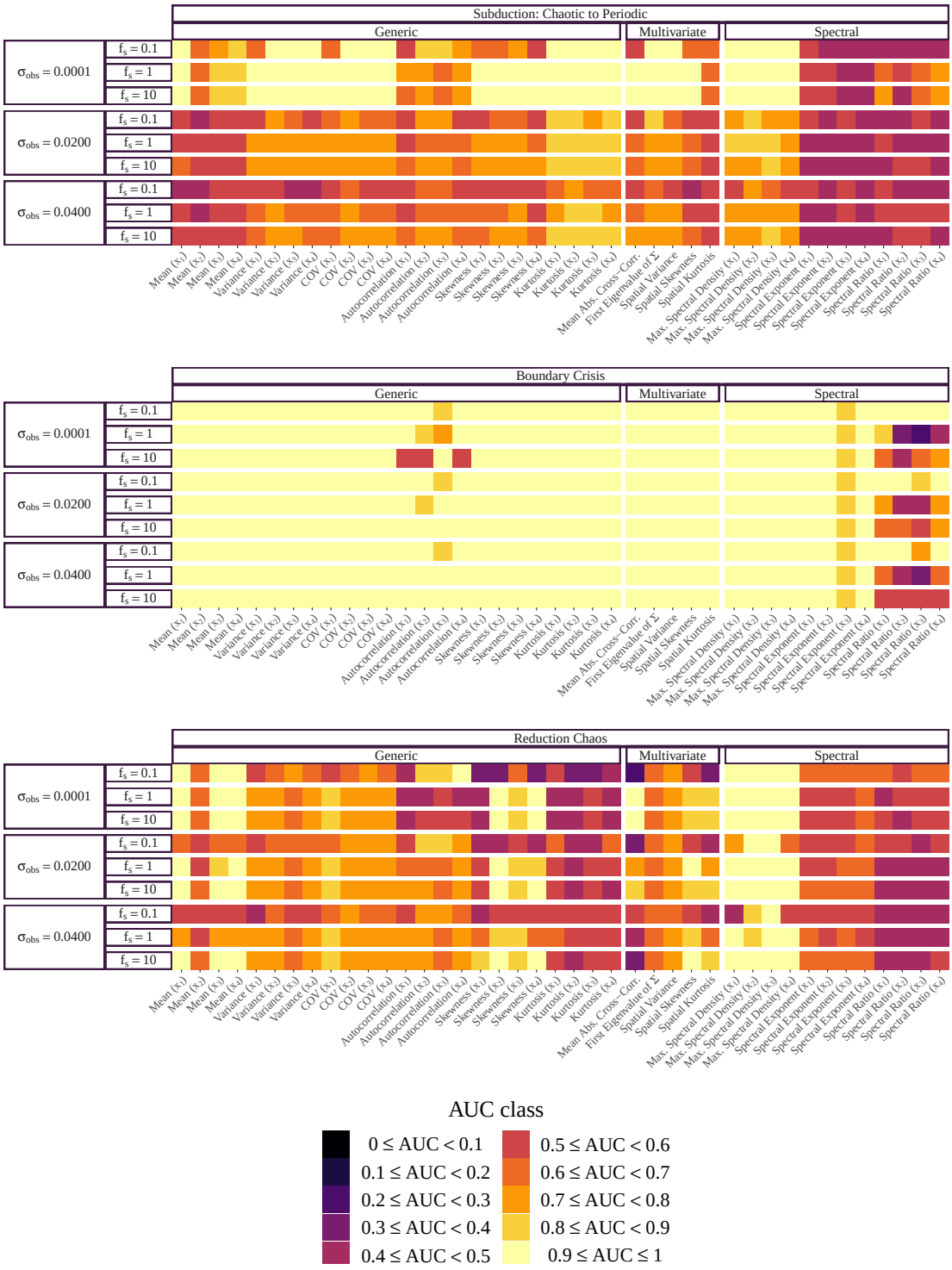

Fig. S9 AUC per condition (brighter, yellow colours indicate better performance, whereas darker, purple colours indicate worse performance)

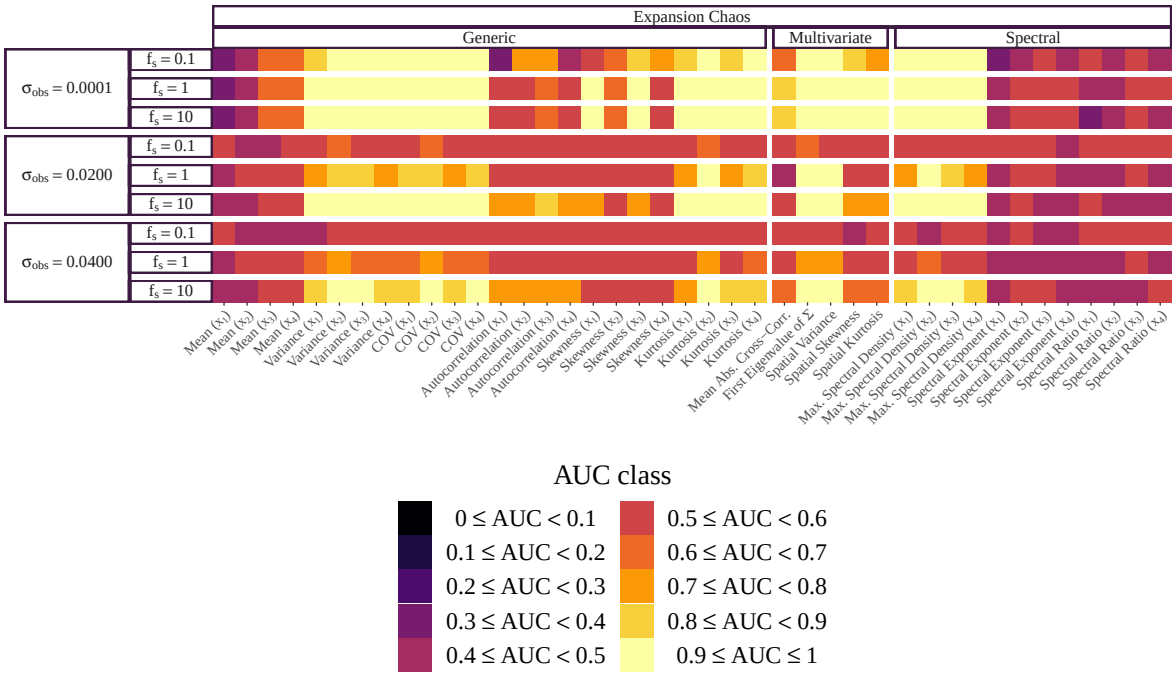

Fig. S10 AUC per condition (brighter, yellow colours indicate better performance, whereas darker, purple colours indicate worse performance)

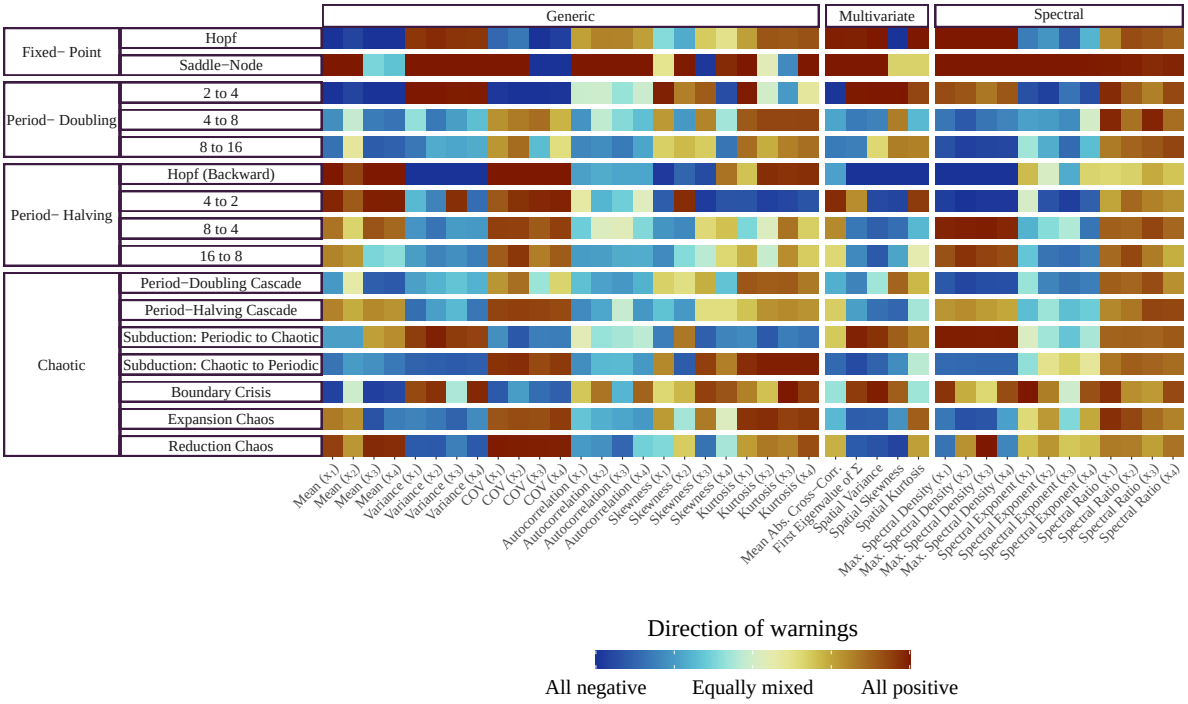

Fig. S11 Direction of Early Warning Signals (EWS) when setting an optimal cut-off value σ^*_{crit}. A warning may be given by a signal that peaks above (i.e. an increase) or below (i.e. a decrease) the confidence band. Across simulations, warnings may show a consistent pattern of either increasing (brown-red) or decreasing (blue), or both may occur (mixed, white)

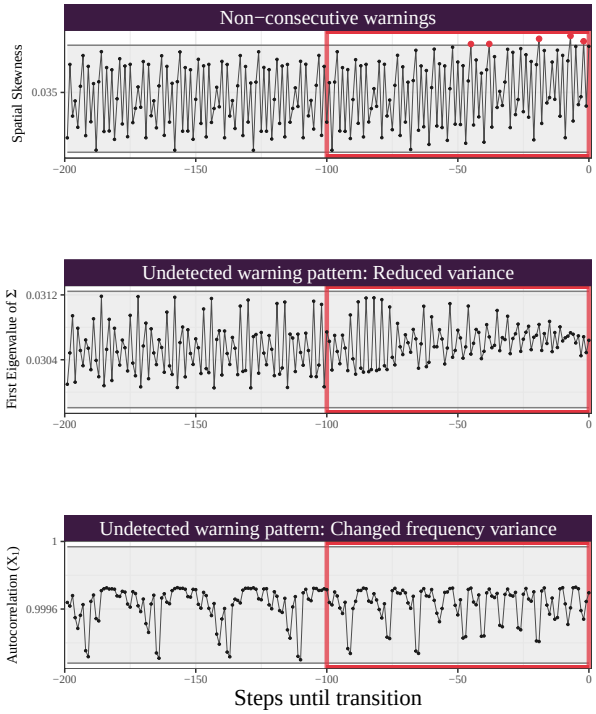

Fig. S12 Complex warning patterns that are not detected well by standard methods. Warning signs are shown for a period-doubling bifurcation (period-4 to period-8, top), period-halving bifurcation (period-4 to period-2, middle), and a period-halving cascade from chaos to periodic behaviour (bottom). The transition period in which s is changing is indicated in red, which is preceded by a baseline period. Confidence bands are constructed with $\sigma_{\text{crit}} = 2$