# §37. Quantifying Self-Organization with Optimal Wavelets

Milovanović, M. (Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade),
Rajković, M. (Institute Vinča, Belgrade),
Tanaka, K., Watanabe, T.-H., Škorić, M.M.

Measuring organization quantitatively has been the subject of various studies in spite of the inherent difficulties to characterize complex systems in an accurate manner. The method presented here [1], is based on a parametric model for a wavelet tree distribution attributing hidden Markov (HM) variable to each node of the tree. The wavelet tree is considered as a self-organizing system by identifying hidden states of the wavelet coefficients with local causal states. The wavelet decomposition of the real-world data is sparse so that most of the energy is compacted into small number of large coefficients, which we call *yang*, while the remaining large number of small ones we label as *yin*. While the yang coefficients provide information on singularities, the yin coefficients carry information on smooth characteristics of the data.

**i) Statistical self-organization** We perceive the wavelet Hidden Markov model (HMM) from the viewpoint of self-organization giving the concept of self-organization specific physical interpretation. First, it is necessary to define the time axis. The interdependence of the nodes of the wavelet tree takes place vertically through the tree so we consider the time axis as dyadic frequency axis directed from the coarsest to the finest scale. We regard the signal domain as spatial even for temporal signals because the concept of time is replacing the frequency domain. Thus, by introducing *diffeomorphism invariance* the wavelet tree becomes the spatio-temporal tree. The direction of time is determined by the branching process representing information flow from parent to descendant coefficients. In the context of binary tree structure and the chosen time axis causality is defined by interdependence of the wavelet coefficients so it lies solely in the HM structure of the wavelet tree. Tying in the EM algorithm implies statistical stationarity (translatory invariance of the distribution) in the spatial domain. As a consequence of the persistence property causality is defined by the presence or absence of singularity in the spatial support of the wavelet coefficients. Recall that the EM algorithm estimates parameter vector $\theta$ by maximizing conditional expectation. This implies maximization of the incomplete likelihood $\log P_\theta(\mathbf{d})$ diminished by the conditional entropy $H_\theta\ (\mathbf{S}|\mathbf{d})$, i.e. $\log P_\theta(\mathbf{d}) - H_\theta\ (\mathbf{S}|\mathbf{d})$. In the course of maximization the second term is minimized, so we assume that it is close to zero, i.e. $H_\theta\ (\mathbf{S}|\mathbf{d}) \approx 0$. It follows that $\mathbf{S}|\mathbf{d} = f_\theta(\mathbf{d})$ with probability close to 1. The parameter vector $\theta$ is characteristic of the dynamic process but not of the particular realization $\mathbf{d}$, hence the implication is that $\mathbf{S} = f_\theta(\mathbf{D})$ i.e. the global causal state is statistic of the wavelet tree. The condition $H_\theta\ (\mathbf{S}|\mathbf{d}) \approx 0$ defines the property of *approximate unifilarity* so we say that the wavelet HMM tree determines the *wavelet machine* or *w-machine*. The knowledge of $\mathbf{S}$ is related to the optimal prediction because $\mathbf{D}$ in HMM depends on $\mathbf{S}$ only. The entropy of the wavelet tree may be expressed as

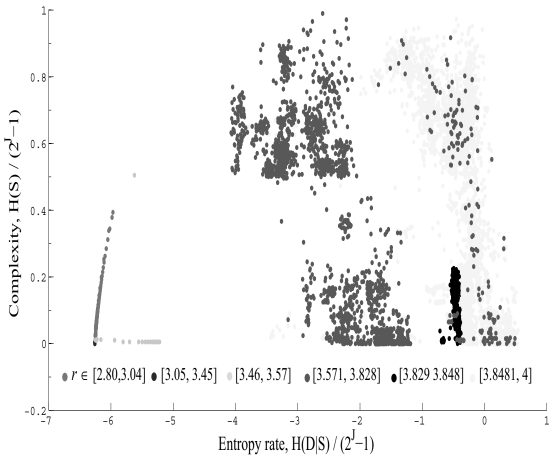


Fig. 1: Entropy rate and complexity pairs $(H(\mathbf{D}|\mathbf{S})/(2^J-1), H(\mathbf{S})/(2^J-1))$ for the logistic map. Shades correspond to different intervals of parameter $r$. Negative entropy values stem from the properties of the differential entropy.

$$H(\mathbf{D}) = H(\mathbf{D}, \mathbf{S}) = H(\mathbf{D}\,|\mathbf{S}) + H(\mathbf{S}), \qquad (1)$$

where $H(\mathbf{D})$ and $H(\mathbf{D}\,|\mathbf{S})$ are differential entropies of continuous random variables. The extensive term $H(\mathbf{D}\,|\mathbf{S})$ represents irreducible randomness that remains even after all correlations are subsumed. The addition of noise increases only this term while complexity $H(\mathbf{S})$ remains unaltered. The local complexity $C_i = H(S_i)$ has a specific physical interpretation: it is higher if the distribution of the hidden yang an yin states in the node is more uniform.

**ii) Global complexity** The local complexity $C_i = H(S_i)$ is a measure which guarantees that the information contained in the node is optimally preserved. The global complexity $C = H(\mathbf{S})$ fulfills that goal for the complete tree. The higher global complexity means that the yang states are more uniformly distributed within the tree allowing for more optimal preservation of the background information. So, we define the optimal representation of the data (signal) as the one which maximizes the global complexity of the tree. In Fig. 1 we present the complexity-entropy rate diagram for the logistic map.

1) M. Milovanović and M. Rajković, Europhysics Letters (2013) vol.102.