ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Effects of a Parametrized Neighborhood Family on the Quality of Self-Organizing Maps

César Cárdenas[1] and Erzsébet Merényi[1,2] *

1- Rice University - Dept of Statistics
2- Rice University - Dept of Electrical and Computer Engineering
6100 Main Street, Houston, Texas, U.S.A.

Abstract. We present a parametrized family of neighborhood functions and experiments to analyze the isolated effects of their shapes on the quality of SOM representations. There appears to be an emerging trend in the relationship between topological and quantization errors, and Shannon entropy. We highlight the trade-offs among these SOM quality measures.

1 Motivation, Background and Objectives

Kohonen's Self-Organizing Map (KSOM) is a powerful unsupervised neural algorithm for learning the structure of an n-dimensional input space $M \subset \mathcal{R}^n$ and representing it on a lower- (usually 2-)dimensional lattice of neurons A [1]. Learning strives to distribute the n-dimensional weight vectors (prototypes) $w_j, j = 1, \cdots, P$ in M according to the data distribution, and simultaneously indexes the prototypes in A in a topology-preserving fashion. At each learn step t, a winner neuron (best-matching unit, BMU) is selected that minimizes the data-space distance $d_M(.)$ to the input vector $x(t)$ at t: $c(x(t)) = argmin_j(d_M(w_j(t), x(t)))$. Then the prototypes are updated as

$$w_j(t+1) = w_j(t) + \alpha(t)h(c(x(t)), j)(x(t) - w_j(t)) \quad (1)$$

where $\alpha(t)$ is a (time-decreasing) learning rate, $h(c, j)$ is the neighborhood function that prescribes the update strength for prototype w_j as a function of the lattice distance $d_A(c, j)$ between neurons c and j. The resulting lattice representations (whether visualized or not) are exploited for the interpretation of the manifold structure, specifically for cluster extraction. The success of this depends on the quality of the learned SOM. Quantization error QE (characterizing embedding) and topology preservation error, TE, are major quality criteria defined and studied by previous research, ranging from relatively costly but detailed measures ([2],[3] and references therein) to simpler but cost-effective ones [4]. These and many other works characterize TE and/or QE in a learned SOM regardless of what influenced the final configuration. A BMU selection approach to influence trade-off between QE and TE is offered in [5]. Obviously, the *shape* of the neighborhood function can also significantly influence SOM quality (beyond the customary good practice of gradually decreasing a starting neighborhood radius of about half the lattice witdth). In this work we offer some insight to this. Previous studies of neighborhood effects on SOM representation

*Partial support by the Cancer Prevention & Research Institute of Texas, grant RP2500090.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

focus on QE and mainly compare the use of Gaussian vs. box neighborhood functions. For example, [6] shows the Gaussian neighborhood to produce lower QE than the box function. A Gaussian neighborhood with locally-adaptive radius is proposed in [7]. There it is shown through an example that the locally-adaptive Gaussian neighborhood improves the TE while worsening the QE compared to the standard Gaussian neighborhood and that the box neighborhood produces the largest QE and TE of the three neighborhoods. Neither works portray the effect on the structure detection potential. Here we propose a parametrized family of neighborhood functions with systematically varying shapes that we generate by a modification of the Gaussian neighborhood in the KSOM learning (eq 1), and evaluate the isolated effect of these neighborhood shapes on SOM quality in terms of TE and QE, and support for structure extraction.

2 The Proposed Neighborhood Family

The often-used Gaussian kernel as $h(.)$ around the BMU c can be written as

$$h(\sigma(t), d_A(c,j)) = \exp\left[-\frac{1}{2}\left(\frac{d_A(c,j)}{\sigma(t)}\right)^2\right] = h(1, d_A(c,j)/\sigma(t)) \qquad (2)$$

where $\sigma(t)$ is the neighborhood radius at learning step t. Observe that when $\sigma(t)$ is decreased to 1 (toward the end of learning),

$$h(1,4) = e^{-8} = 3.35 \cdot 10^{-4} \quad \text{and} \quad h(1,5) = e^{-12.5} = 3.72 \cdot 10^{-6}.$$

I.e., the update weight h(.) is near-zero at $d_A = 5$ while it still has appreciable influence at $d_A = 4$. Therefore, we can consider $d_A = 4$ as the *effective neighborhood radius*, r_{eff}, of the Gaussian neighborhood in the late stages of learning (where we are interested to evaluate the effects). This inspires the construction of a parametrized family where all members have the same effective neighborhood radius, r_{eff} as the Gaussian, when $\sigma = 1$. Throughout this paper we will say that two neighborhood functions vary mostly in their shapes if they have the same r_{eff}. Formally, we define r_{eff} as the largest natural number d such that by eq (2), $h(1,d) \geq \exp(-8)$.

Uniform r_{eff} allows isolation of the effect of the different neighborhood shapes. We achieve uniform r_{eff} by defining new neighborhood functions using the relationship in eq (2) as follows:

$$h^p(\sigma(t), d_A(c,j)) = h^p(1, d_A(c,j)/\sigma(t)) = \exp\left[-\frac{1}{2^{2p-3}}\left(\frac{d_A(c,j)}{\sigma(t)}\right)^p\right] \qquad (3)$$

where the scaling constant $b_p = \frac{1}{2^{2p-3}}$ is such that $h^p(1,4) = \exp(-8)$. Obviously, $p = 2$ recovers the Gaussian neighborhood function. Fig. 1 gives a sampling of the shapes for different values of p. Unsurprisingly, the shape is exponential-like for $p < 2$, Gaussian-like for p around 2 and box-like for large p.

Notice that if $p_1 < p_2$ and $d_A \in [0,4]$ then $h^{p_1}(1,d_A) \leq h^{p_2}(1,d_A)$. This implies that p_2 induces higher attraction effects to the lattice-neighbors of the

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

BMU than p_1, within $r_{eff} = 4$. Therefore, as p increases we expect lattice-neighbor prototypes to be increasingly closer in input space. However, as with SOM learning in general, we expect that the distribution of the prototypes will still be similar to the distribution of the data. From the combination of these we intuit that the increased attraction effect is local rather than global. Thus, as p increases we expect groups of more concentrated prototypes in input space.

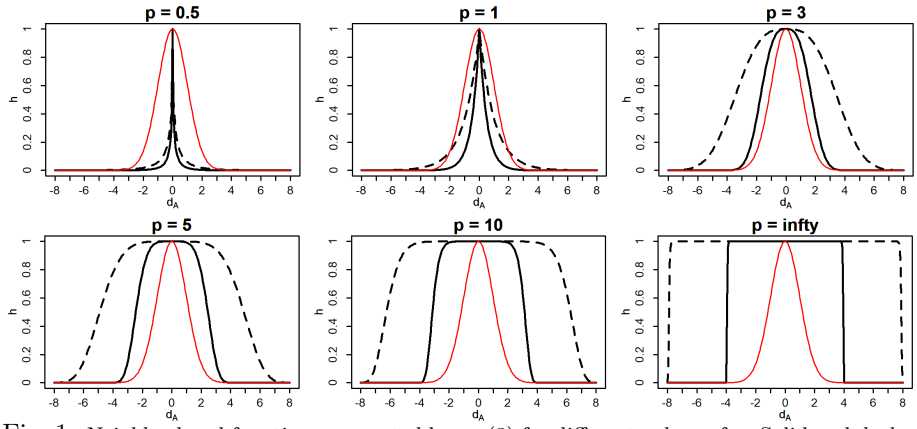

Fig. 1: Neighborhood functions generated by eq (3) for different values of p. Solid and dashed black lines, respectively, represent $\sigma(t) = 1$, $\sigma(t) = 2$ while the red curve shows the Gaussian neighborhood function with $\sigma(t) = 1$.

3 Experiments and Results

Data: We study neighborhood effects on three data sets of varying complexity.

i) The 6d-synthetic 20-class data is a simulated, 128 x 128-pixel spectral image with a 6-dimensional feature vector characterizing each of 16,384 pixels. 16 "large" spectral classes occupy 32 x 32-pixel quadrants, and four small, (1-, 16-, 128-, 256-pixel) classes are inserted into some of the large ones. Pixel signatures in each class were generated by adding Gaussian noise to challengingly close predefined class means. See details in [3].

ii) Ocean City is a real remote sensing spectral image cube of Ocean City, MD, where each of 262,144 pixels is represented by a 8-dimensional vector. Previous work extracted 28 verified clusters including a variety of roof materials, road, paint, vegetation [3]. These serve as reference in this work.

iii) Iris is the well-known Fisher data set [8]. It includes 150 4-dimensional feature vectors characterizing 50 labeled samples of each of three iris flower classes. One of the classes is easy to extract, the others pose a known challenge for delineation. The heterogeneity of this small data set can also lead to the extraction of four (instead of three) clusters. We use the Iris data also to study the neighborhood effects for a very small SOM.

Processing: We use 20x20 SOMs with 1 million learning steps for the 6d synthetic 20-class data, 40x40 SOMs with 4 million steps for Ocean City, and 7x7 SOMs with 10 thousand steps for Iris. In each case, we keep initialization and sampling order of input data the same for a given data set while sweeping the

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

values of p shown in Table 1. We use similar step functions as decay schedules of α and σ for the three data sets after accounting for sample size differences.
Evaluation: Topology preservation errors (different components of TE) in rows 2 – 4 of each data section of Table 1 are derived from the CONN matrix [3]. CONN portrays the connectivity (weighted Delaunay graph) of the input space M and shows *forward* topology violations (or absence thereof) in an SOM lattice by line segments connecting non-lattice-neighbor cells. Line width indicates the severity of the violation as visualized in Fig. 2. # CONN Connections refers to the overall number of connected pairs of prototypes. We separate all violations to global and local ones. Global ones occur at folding lengths k beyond the *radius of local violation*, r_l, as defined in [3]. $r_l = 1, 2, 3$ for Iris, 20-class and Ocean City data, respectively. Local violations have negligible effect on the interpretation of the manifold structure thus not considered "real" violations. Quantization Error is computed as the average reconstruction (distortion) error for a sample. We show Shannon Entropy of each map, normalized to lattice size. The higher the entropy, the more faithfully the map represents the data distribution. In Figs 2 – 4 the mU-matrix expresses the relative winning frequency of the prototypes (intensity of the red color) which in turn informs of the entropy. The more uniform the red intensity the higher the entropy. The brightness of the "fence" between two cells is proportional to the data-space distance of the respective pair of lattice-adjacent prototypes, thus it may alert to backward topology violation.

Data	Quality Measure	p = 0.5	p = 1	p = 1.5	KSOM	p = 3	p = 5	p =10	p=infty (box)
20-class	# CONN Connections	1489	1593	1305	1041	878	712	620	506
	# Topology Violations	758	720	462	298	180	120	112	175
	# Global Violations	251	311	203	147	57	37	27	65
	# Local Violations	507	409	259	151	123	83	85	110
	Quantization Error	3.71	3.9	4.18	4.36	4.57	4.78	4.91	5.12
	Shannon's Entropy	0.9089	0.9334	0.9331	0.9223	0.9049	0.8937	0.8511	0.7661
Ocean City	# CONN Connections	11591	9777	8220	7743	7196	6436	6126	7037
	# Topology Violations	6670	4662	3456	3273	2854	2224	2184	3090
	# Global Violations	2289	2189	2053	2023	1786	1369	1150	885
	# Local Violations	4381	2473	1403	1250	1068	855	1034	2205
	Quantization Error	153.82	157.55	165.32	168.14	174.22	182.65	192.34	197.58
	Shannon's Entropy	0.9544	0.9766	0.9793	0.9817	0.9807	0.9777	0.9688	0.9619
iris	# CONN Connections	50	53	51	49	53	47	41	19
	# Topology Violations	9	3	3	0	4	2	7	2
	# Global Violations	9	3	3	0	4	2	7	2
	# Local Violations	0	0	0	0	0	0	0	0
	Quantization Error	0.09	0.10	0.11	0.11	0.12	0.14	0.16	0.17
	Shannon's Entropy	0.7895	0.8349	0.8321	0.8388	0.8393	0.8174	0.7073	0.5700

Table 1: Summary statistics of SOM quality measures for the tested values of p

From Table 1, opposing trends seem to emerge for TE (the total # of Topology Violations) and QE as a function of p. Figs 2–4 show selected illustrations. While TE decreases with p QE becomes worse (increases) with p. Some increase in TE occurs at $p = \infty$ for both the 20-class and Ocean City data. This may be caused by the box neighborhood shape that induces a strong concentration of the prototypes (similar to what we see in Fig. 4 for the Iris data). Notice, however, that a significant portion of the TE increase is contributed by local violations in both cases. While the same trend of QE holds for the Iris data, it is less clear whether the fluctuations of TE are real or due to either boundary

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

effects in a small SOM or small sample size.

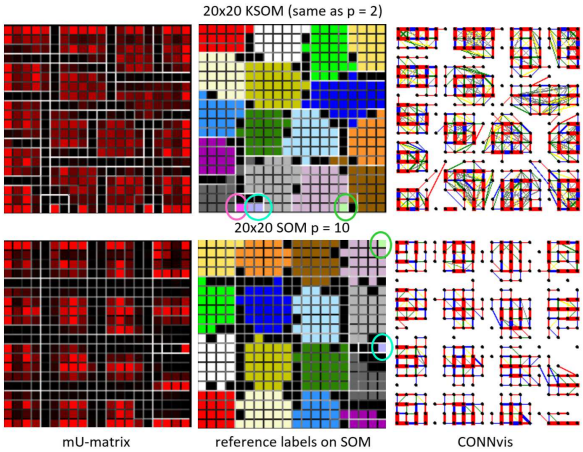

Fig. 2: SOMs representations of the 6d-synthetic 20-class data set for KSOM (top) and $p = 10$ (bottom). Lower entropy for $p = 10$ is evident from the mU-matrices, as is the decrease in TE from the CONN visualization. However, two clusters became connected indicating less cluster separability. (For $p = \infty$, not shown here, more clusters become connected.) This is also illustrated by reference labels over the SOMs: KSOM resolves three of the four rare classes (hot pink, turquoise, light green ovals) while the $p = 10$ SOM only resolves two.

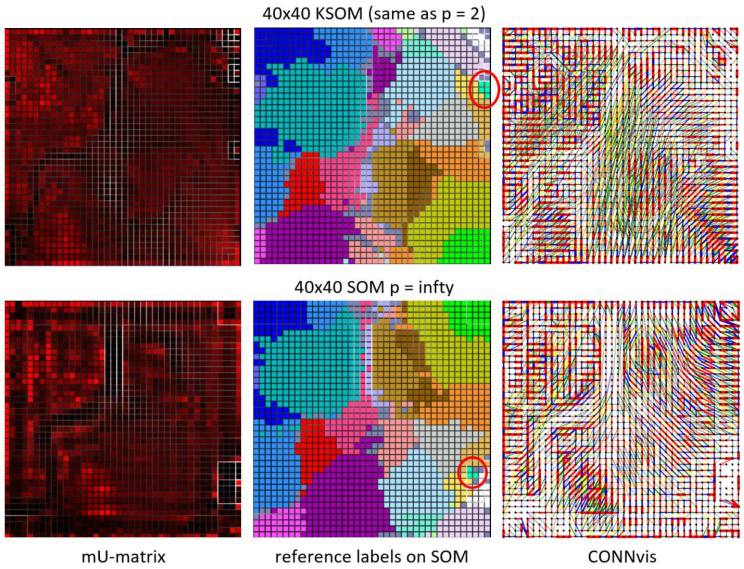

Fig. 3: SOMs representations of Ocean City for KSOM (top) and $p = \infty$ (bottom). As in Fig. 2, mU-matrices and CONNvis confirm lower entropy and TE at $p = \infty$. Reference labels indicate long-range folding in the KSOM, "twisting" the magenta and hot pink classes into two (top center and bottom left). The long violations are confirmed by CONNvis. The separated classes are reunited in the lower map and long violations disappear. However, the $p = \infty$ map has reduced the representation of rare classes (e.g., neon green class in red oval.)

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

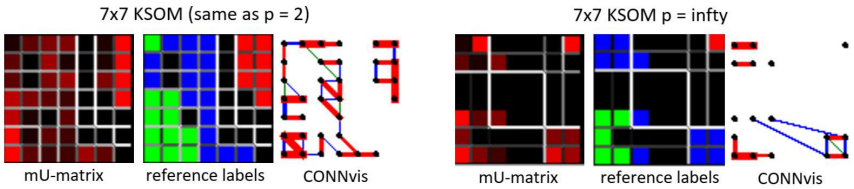

Fig. 4: SOM representations of the Iris data set, for KSOM (left) and $p = \infty$ (right). For the extreme value of $p = \infty$ the concentration of prototypes is very high, reducing the entropy (as seen from the mU-matices) and introducing strong global (long) violations while the number of connections dramatically drops compared to the KSOM. From the reference labels, we also see that strong prototype concentration can pull apart a class. Here, the blue class has been split into separate regions in the $p = \infty$ lattice.

The entropy —which has its highest values in the mid-range of p— implies that Gaussian-like neighborhoods produce more faithful representations of the data distribution than exponential- or box-like neighborhoods. Therefore, they may support cluster extraction better than maps at the extreme values of p since low entropy favors the merging of clusters in the representation.

4 Conclusion

We propose a family of SOM neighborhood functions and present preliminary assessment of the shape effects. Our experiments suggest systematic trends, as well as trade-offs, of TE, QE, and Shannon entropy, all of which bear on the maps' support for structure extraction. This, and observed deviations from the general trend motivate more complete (e.g., more resolution in p between 10 and ∞) and more refined evaluation (e.g., quantifying the quality of independently extracted clusters), on more varied data sets in follow-up work.

References

[1] T. Kohonen, *Self-Organizing Maps*, 2nd ed. Berlin, Germany: Springer-Verlag 1997.

[2] T. Villmann, R. Der and M. Herrmann, Topology preservation in self-organizing feature maps: exact definition and measurement, *IEEE Trans. Neural Networks*, 8:256-266, 1997.

[3] K. Tasdemir and E. Merényi, Exploiting Data Topology in Visualization and Clustering of Self-Organizing Maps, *IEEE Trans. Neural Networks*, 20:549-562, 2009.

[4] L. Hamel, SOM Quality Measures: An Efficient Statistical Approach. In: *Advances in Self-Organizing Maps and Learning Vector Quantization (WSOM 2016).* pp 49-59, 2016.

[5] J. S. Kirk, and M. Zurada, Algorithms for improved topology preservation in self-organizing maps, *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics*, pages 396-400, vol 3, Tokyo (Japan), 1999.

[6] P. Stefanovic and O. Kurasova, Influence of Learning Rates and Neighborhood Functions on Self-Organizing Maps. In: J. Laaksonen, T. Honkela eds. *Advances in Self-Organizing Maps. WSOM 2011.* Lecture Notes in Computer Science, vol 6731, 141–150. Springer.

[7] K. Kiviuloto, Topology Preservation in Self-Organizing Maps, *Proc. International Conference on Neural Networks (ICNN'96)*, vol 1, 294-299, Washington, DC, (USA), 1996.

[8] R. A. Fisher, The use of multiple measurements in taxonomic problems. *Ann. Eugenics.* pages 178-188, vol 7, 1936.