

## MULTIFRACTAL CHARACTERIZATION OF SPATIAL INCOME CURDLING: THEORY AND APPLICATIONS

FRANCISCO ROSALES<sup>\*,†</sup>, ADOLFO POSADAS<sup>\*,†,§</sup> and ROBERTO QUIROZ<sup>\*,¶</sup>

\**Production Systems and the Environment Division,  
International Potato Center, Av. La Molina 1895, Lima 12, Peru*

†*Facultad de Ciencias Físicas, DAFI,  
Universidad Nacional Mayor de San Marcos,  
Av. Venezuela Cdra 4, Lima 1, Peru*

‡*l.rosales@cgiar.org*

§*a.posadas@cgiar.org*

¶*r.quiroz@cgiar.org*

Received 25 September 2007

Revised 14 April 2008

A spatial curdling index ( $\mathcal{C}$ ) is presented for studying income concentration in two-dimensional fields originated by a cascade process, resembling resource allocation maps. Computational simulations of spatial distributions show that the relation between  $\mathcal{C}$  and measures of income inequality such as the Gini coefficient ( $\mathcal{G}$ ) forms a concave area. Acknowledgment of this fact brings additional information for formulating geopolitical strategies for income inequality reduction, and for comparison purposes in a set of countries or across time within a country. With the aim of showing the importance of index  $\mathcal{C}$ , a prototype of the empirical application is shown based on the Peruvian and Swedish income maps.

*Keywords:* Resource allocation maps; cascade processes; Gini coefficient.

### 1. Introduction

This study presents a framework for policy analysis that accounts for spatial curdling when analyzing income distribution [5, 6, 13, 15], which is different from regular entropy indicators such as the Theil and Atkinson indexes [2]. This is important, because these indicators of income inequality and the well-known Gini coefficient [8] fail to discriminate between different spatial structures.

To conduct this analysis it was necessary to develop a theory for a spatial curdling index ( $\mathcal{C}$ ) from a sociophysics point of view [3, 14]. This theory considers that a small number of laws can characterize the emergence of certain patterns in social phenomena. The analysis is supported by a geographic information systems

<sup>†</sup>Corresponding author. Address: Av. La Molina 1895, PO Box 1558, Lima 12, Peru.

(GIS) perspective [1] to build income maps for the multifractal analysis [17–19] of the cascade process [9, 10].

The main finding of the study of index  $\mathcal{C}$  is its nonlinear generation with the Gini coefficient and the relation of a concave area when analyzing both indicators in the  $\mathcal{C}$ - $\mathcal{G}$  plane. This fact indicates that index  $\mathcal{C}$  bring additional information for formulating geopolitical strategies for income inequality reduction, and for comparison purposes between countries or across time within a country.

This paper is organized as follows. Section 2 reviews some basic concepts of multifractals. Section 3 presents a comparative analysis between the Gini coefficient ( $\mathcal{G}$ ) and the proposed spatial curdling index ( $\mathcal{C}$ ) to introduce a two-dimensional picture of inequality for its use in policy analysis. Section 4 presents a prototype of the empirical application studying the Peruvian and Swedish income maps. Finally, Sec. 5 lists concluding remarks.

## 2. Cascade Processes

The origins of multifractal theory can be tracked back to the works by Kolmogorov [9, 10]. Following Turiel [19], under conditions of fully developed turbulence, the velocity of the local dissipation of energy varies sharply between neighbor locations and should be regarded as random quantities. Let  $M_\delta(x)$  be the local dissipation of energy at the point  $x$  over a neighborhood of radius  $\delta$ :

$$M_\delta(x) = \frac{1}{|B_\delta(x)|} \int_{B_\delta(x)} dx' \sum_{ij} [\partial_i v_j(x') + \partial_j v_i(x')]^2,$$

where  $v_i$  are the components of the velocity vector and  $B_\delta(x)$  is the ball of the radius  $\delta$  centered around  $x$ . Kolmogorov's intuition was that the energy is transmitted from larger ( $\Delta$ ) to smaller scales ( $\delta$ ) by means of an injection process defined by a variable  $\eta_{\delta\Delta}$  which depends only on the ratio  $\delta/\Delta$ , as

$$M_\delta = \eta_{\delta\Delta} M_\Delta. \quad (1)$$

Originally the energy injection variable  $\eta_{\delta\Delta}$  was represented as a fixed value,  $\eta_{\delta\Delta} = (\frac{\delta}{\Delta})^{-\alpha}$ , from which it could be deduced that the order- $p$  moments of  $M_\delta$  are related to those of  $M_\Delta$  by

$$\langle M_\delta^p \rangle = \left( \frac{\delta}{\Delta} \right)^{-\alpha p} \langle M_\Delta^p \rangle \sim \delta^{-\alpha p},$$

and all dependence in  $\delta$  of the order- $p$  moment of  $M$  is concentrated in the power law  $\delta^{-\alpha p}$ . That is to say,  $\langle M_\delta^p \rangle \sim \delta^{\tau(p)}$ , which is known as the property of self-similarity. However, to model the factual nonlinear character of  $\tau(p)$  [19],  $\eta_{\delta\Delta}$  in Eq. (1) should be interpreted as a random variable independent of  $\Delta$ . This can be done assuming the existence of local scale-invariant laws; namely, that at any

point  $x$

$$M_\delta(x) = \delta^{\alpha(x)}$$

holds and all dependence on the scale parameter  $\delta$  is conveyed by the power law factor  $\delta^{\alpha(x)}$ . The exponent  $\alpha(x)$  is called the singularity exponent of point  $x$ , and can then be arranged in sets  $\mathcal{S}_\alpha = \{x : \alpha(x) = \alpha\}$ , called singularity components  $\mathcal{S}_\alpha$ . To close the model, the singularity components should be of fractal character. The singularity spectrum associated with the multifractal hierarchy of fractal components is the function  $f(\alpha)$ , defined by the Hausdorff dimension of each component  $\mathcal{S}_\alpha$ , namely  $f(\alpha) = \dim_H(\mathcal{S}_\alpha)$ , where the maximum is reached when the derivative is zero and  $f(\alpha)$  equals the fractal dimension of the measure's support, such as 1 in the case of a one-dimensional signal. Another special point occurs when the derivative of  $f(\alpha)$  equals 1, i.e.

$$f(\alpha_S) = \alpha_S = S, \quad (2)$$

which is called the curdling dimension of the cascade process.

Under some assumptions on the homogeneity and isotropy of the statistics of local singularities [17], it is possible to derive a relation between self-similarity exponents  $\tau(p)$  and the singularity spectrum  $f(\alpha)$ . Furthermore,  $\tau(p)$  can be computed from a Legendre transform of  $f(\alpha)$ :

$$\tau(p) = \inf_{\alpha} \{\alpha p + d - f(\alpha)\}. \quad (3)$$

Therefore the singularity spectrum contains all the information about self-similarity, and about the multiplicative process [see Eq. (3)], i.e. it describes the statistics of changes in scale. For a better understanding of the previous formulations in economics, see Refs. 4, 7 and 16. In the following section we will employ the curdling dimension to construct an index for income inequality analysis across space.

### 3. Measures of Income Distribution

This section compares the Gini coefficient ( $\mathcal{G}$ ) income inequality measure with a multifractal index derived from the cascade process formalism reviewed in Sec. 2.

#### 3.1. The Gini coefficient $\mathcal{G}$

The Gini coefficient [8] is defined as a ratio of the form

$$\mathcal{G} = \frac{A}{A + B}, \quad (4)$$

where  $B$  is the area under the Lorenz curve of the distribution [12] and  $A$  is the area between this curve and the uniform distribution line. Furthermore, since

$A+B = 1/2$ , the Gini coefficient  $\mathcal{G} = 2A = 1 - 2B$ . If the Lorenz curve is represented by the function  $\mathcal{L}(x)$ , the value of  $B$  can be found by integration as

$$\mathcal{G} = 1 - \int_0^1 \mathcal{L}(x)dx. \quad (5)$$

The ratio  $\mathcal{G}$  in Eq. (4) is a measure of inequality with values ranging from 0 (perfect equality) to 1 (perfect inequality). Its interpretation when applied to income across geographic locations is the following: as  $\mathcal{G}$  approaches 1, all income is found concentrated in a small group of locations, and as  $\mathcal{G}$  reaches a value of 0, income is uniformly distributed among all locations.

It should be noted that this coefficient omits the spatial location of the income source, i.e. whether or not spatial clustering takes place, since it characterizes the income variable only by its distribution law.

### 3.2. Spatial curdling index $\mathcal{C}$

We introduce a spatial curdling index  $\mathcal{C}$ , a GIS-based method for analyzing clusters within the measure based on the intensity of the signal in a region by means of multifractal analysis.

As presented in Sec. 2, the curdling dimension captures the measure of multifractal sets where curdling occurs. That is when the singularity spectrum  $f(\alpha)$  is tangent to the  $45^\circ$  line [see Eq. (2)]. When applied to income maps, this becomes a measure of the income's curdling in multifractal sets. Following this interpretation, low-concentration systems will have a smaller curdling dimension,  $f(\alpha_S)$ , than those with high concentration. Therefore we can write

$$\mathcal{C} = 1 - \frac{f(\alpha_S)}{f(\alpha_0^E) - f(\alpha_0^T)}, \quad (6)$$

where  $f(\alpha_0^E)$  and  $f(\alpha_0^T)$  are the support dimensions of sets with perfect equality and perfect inequality respectively, and  $f(\alpha_S)$  is the curdling dimension. Notice that because  $f(\alpha_0^E) \leq f(\alpha_S) \leq f(\alpha_0^T)$ ,  $\mathcal{C}$  is a normalized measure of income's curdling ranging in  $[0, 1]$ , i.e. the same interval where the Gini coefficient operates, with a similar interpretation: as  $\mathcal{C}$  approaches 0, income is distributed perfectly across space, and as  $\mathcal{C}$  gets near 1, income clusters in a specific location.

To illustrate the application of the index in Eq. (6), let us construct the income map of an imaginary square shaped region formed by  $16 \times 16 = 256$  locations perfectly aligned across space so their borders take the form of a grid in the Cartesian plane, and axes  $X$  and  $Y$  represent arbitrary geographic coordinates of longitude and latitude. In the same manner, let us assign a gray scale corresponding to the income value in each region normalized to the interval  $[0, 1]$ .<sup>a</sup>

<sup>a</sup>In this and the following figures resembling resource allocation maps (see Figs. 3 and 7), the interpretation of axes in the Cartesian plane and the gray scale will be equivalent.

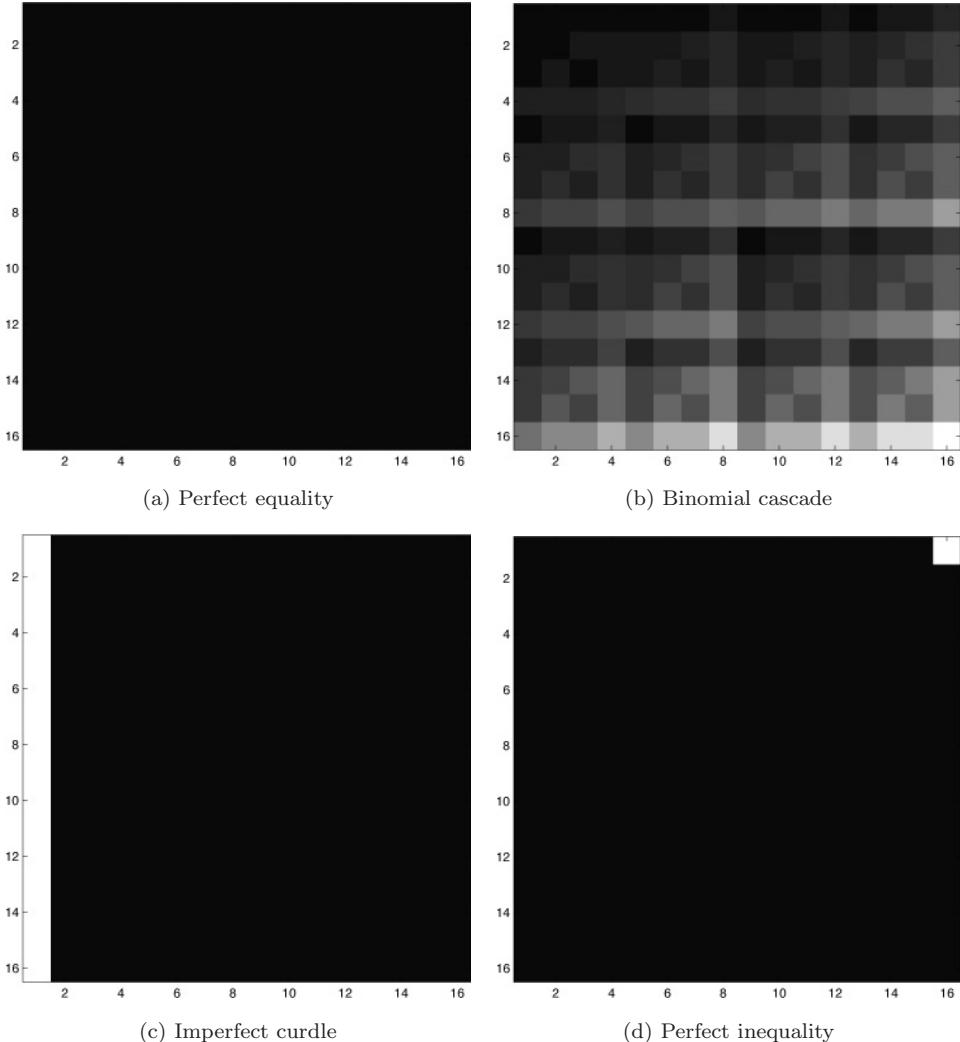


Fig. 1. Measure examples of  $\mathcal{C}$  extreme values.

Figure 1 shows four examples of income maps for the imaginary square-shaped region: (a) represents the case where all regions have the exact same income, (b) the distribution of a binomial cascade derived from a Cantor set [11], (c) a particular spatial curdling pattern where all income is located in a specific set of regions (which may be interpreted as regions close to a hypothetical capital), and (d) the perfect inequality case where all income is found in a single region.

To evaluate index  $C$  in the artificial maps of Fig. 1, one is required to compute the three parameters in Eq. (6) for each case. Figure 2 shows the curdling dimension for each of the examples in Fig. 1 along the dashed line [see Eq. (2)]: 2 in case (a),  $\approx 1.8$  in case (b), 1 in case (c) and 0 in case (d). As should be noted, the

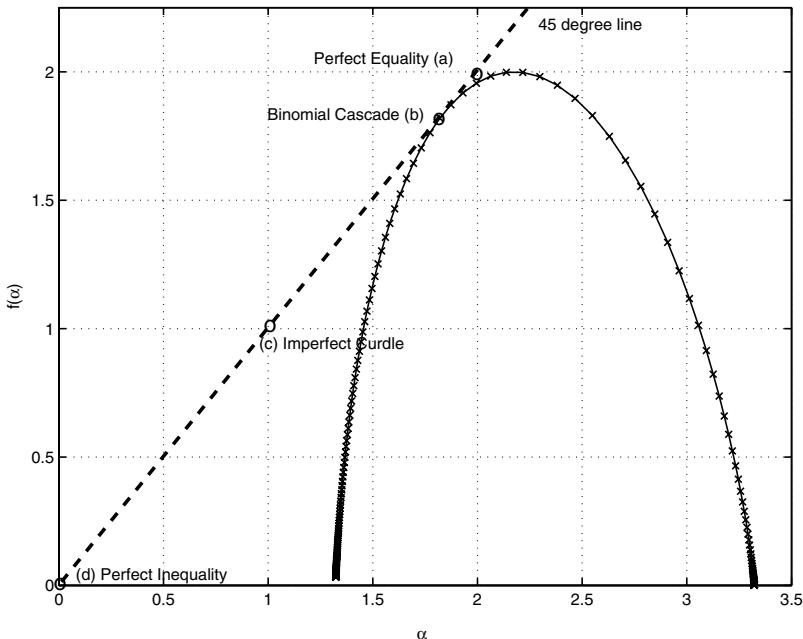


Fig. 2. Extreme values of  $\mathcal{C}$  in the singularity spectrum diagram. Cases (a)–(d) correspond to Fig. 1.

only case that generates a classical singularity spectrum is the binomial cascade.<sup>b</sup> This occurs because cases (a), (c) and (d) do not correspond to multifractal sets; however, one can still calculate their curdling dimension applying Eq. (2), knowing that they would be equal to their Euclidean dimensions. Furthermore, since curdling dimensions of cases (a) and (d) correspond to the perfect equality and perfect inequality situations respectively,  $f(\alpha_0^{\mathcal{E}}) = 2$  and  $f(\alpha_0^{\mathcal{T}}) = 0$ , making the computation of  $\mathcal{C}$  for the examples in Fig. 1 a matter of arithmetics (see Fig. 5 to visualize the computed values).

After clarification of the location of the  $\mathcal{C}$  critical values of Fig. 1, computational simulations of alternative measures are shown to portray how index  $\mathcal{C}$  accounts for spatial income concentration. Figure 3 illustrates this, presenting an artificial income map with values between 0 and 1 generated as a matrix of random numbers taken from a uniform distribution [see case (a) in the figure]. To force the concentration increase of income in the map, we artificially generate cross-shaped clusters by assigning a value of 0 to income in any region outside arbitrary selected clusters [see cases (b)–(d) in the figure].

<sup>b</sup>The computation of singularity spectra was performed using INRIA's FracLab freeware software, available for download at: <http://apis.saclay.inria.fr/FracLab//index.html>

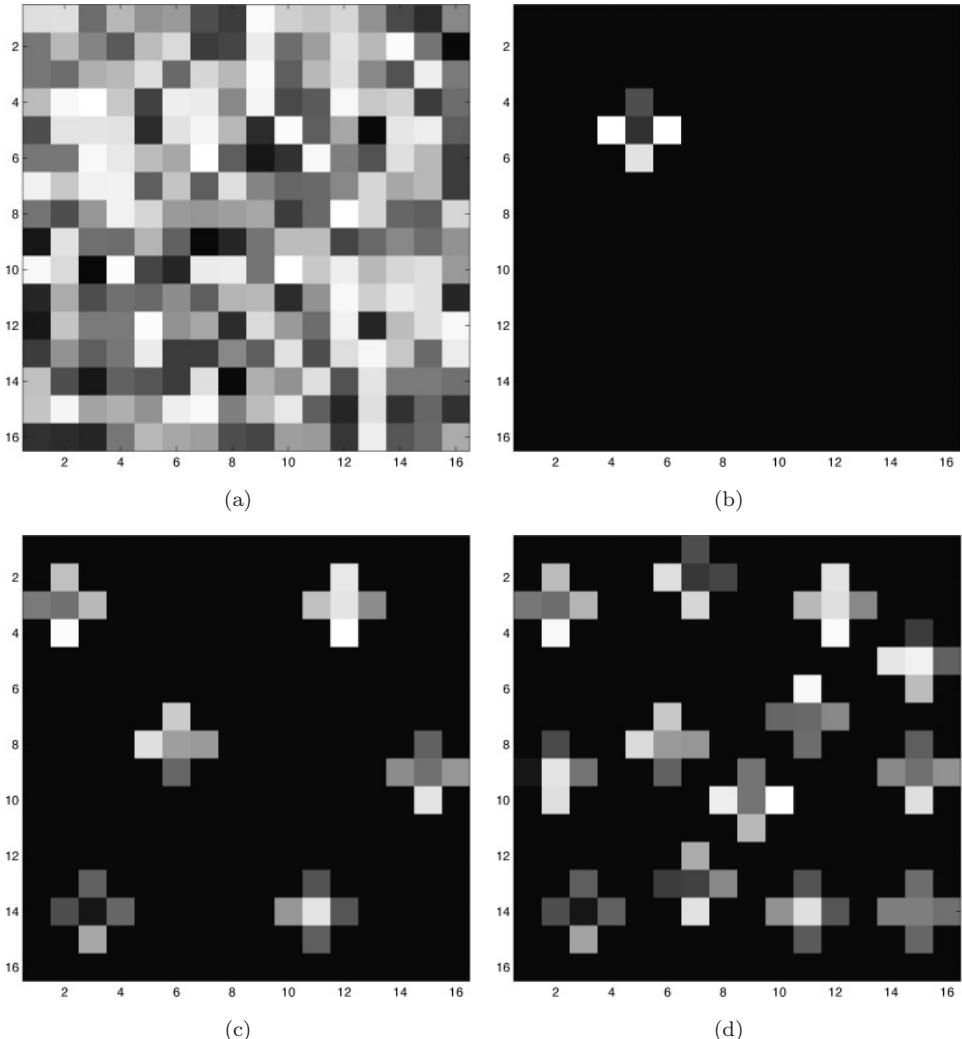


Fig. 3. Measure examples with different concentration intensities. Each map is normalized so that the aggregation of income in each case equals 1.

Figure 4 presents the singularity spectra of the cases displayed in Fig. 3, showing how the curdling dimension increases as a function of the concentration of income in space.

### 3.3. The relation between $\mathcal{G}$ and $\mathcal{C}$

The comparison between  $\mathcal{G}$  and  $\mathcal{C}$  is performed by simulations based on the imaginary square-shaped region from which we discussed deriving the relation between  $\mathcal{C}$  and the spatial clustering presented in Subsec. 3.2. Therefore each pixel represents

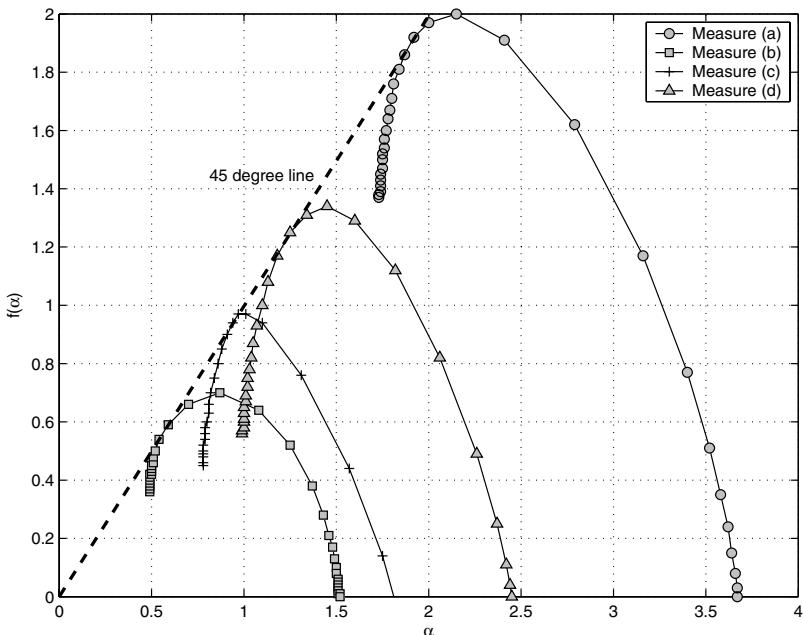


Fig. 4. Singularity spectra of measure examples. Cases (a)–(d) correspond to Fig. 3.

a subregion with its own characteristics, as described by the hypothetical cases in Figs. 1 and 3.

Figure 5 shows the relation between  $\mathcal{G}$  and  $\mathcal{C}$  of previous examples locating  $\mathcal{C}$  in the  $\mathcal{C}$ – $\mathcal{G}$  plane. The resulting graph shows that the relation between these two indicators is far from being linear, even though they coincide at the extreme values of perfect equality (perfect curdle) and perfect inequality (perfect income ubiquity). A more insightful scrutiny will show that not only is the relation not linear, but there is an area relating these two indicators. This happens because for a given Gini coefficient one can find spatial distributions with many different curdling patterns, and vice versa.<sup>c</sup>

To derive the shape of the area relating  $\mathcal{G}$  and  $\mathcal{C}$  presented in Fig. 6, we built thousands of artificial income maps of  $16 \times 16 = 256$  regions with income values generated as a matrix of random numbers taken from the Pareto–Lévy distribution laws with varying parameters [15], and computed the pair  $(\mathcal{C}, \mathcal{G})$  in each case. The generation of a concave area is explained because for a given Gini coefficient one can find spatial distributions with very different curdling patterns. Therefore, the unified analysis in the plane  $\mathcal{C}$ – $\mathcal{G}$  could be useful when one is analyzing the evolution of

<sup>c</sup>One should note that this does not happen in the extreme cases  $\mathcal{G} = \mathcal{C} = 0$  and  $\mathcal{G} = \mathcal{C} = 1$ , and in any monofractal set for that matter.

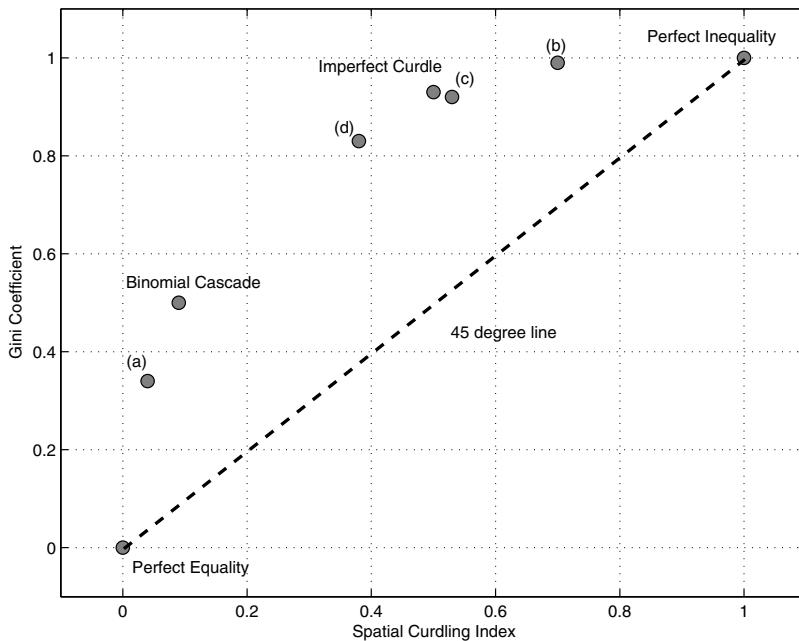


Fig. 5. Comparison between  $\mathcal{G}$  and  $\mathcal{C}$ . Cases (a)–(d) correspond to Fig. 3.

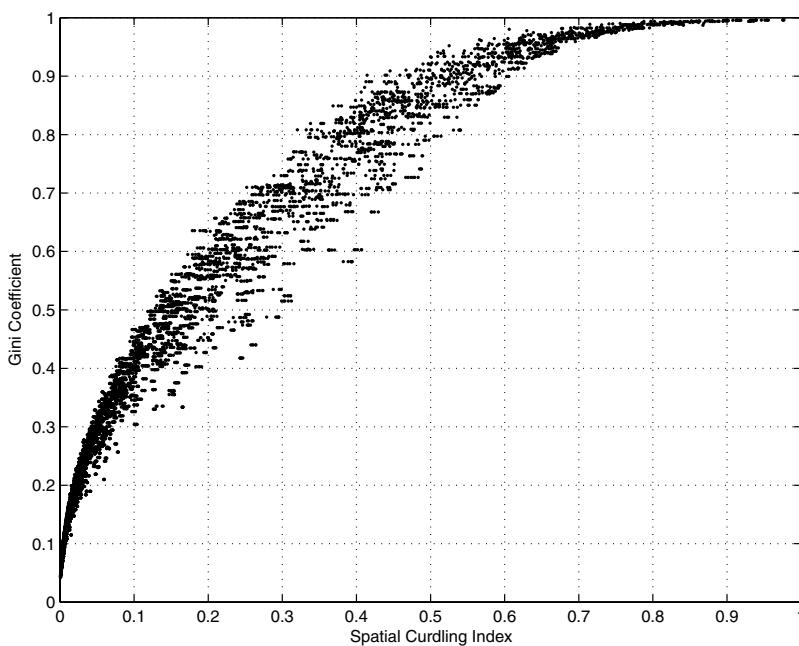


Fig. 6. Generation of the  $\mathcal{G}$ - $\mathcal{C}$  concave area.

income inequality considering a spatial indicator, and when comparing two countries with similar Gini coefficients, by showing the behavior of its spatial distribution.

It is important to note that Fig. 6 is an idealized representation of reality since in actual income maps each region corresponds to more than one pixel, and the GIS data processing may generate some distortions if the population density is not stable across regions. In the following section we will deal with these problems by showing empirical applications in Peru and Sweden.

#### 4. Empirical Applications

In this section we present the empirical application of the curdling index  $\mathcal{C}$  and the interpretation of the plane  $\mathcal{C}-\mathcal{G}$  for policy analysis. As prototypes Peruvian and Swedish income maps are studied. Computations are performed using the Multifractal Analysis and Scaling Systems (MASS) software developed at CIP<sup>d</sup> employing the moment method formalism [17, 18] to calculate singularity spectra. For a deeper review of the methodology, see Ref. 18.

To perform the empirical analysis of real income maps, data must be transformed from a list of regions with a corresponding wealth value, to a map where regions are represented by polygons in space. These polygons are generally available from the national statistics web page of each country, and their visualization can be computed using any specialized software, such ArcGIS or ESRI GIS (see Fig. 7). Under this representation income maps are built from  $n \times m$  matrices making two basic assumptions: (i) the aggregated income of each district is homogeneously distributed in each pixel, conforming to the district (basically due to the lack of information at a more detailed level); and (ii) there is a positive relation between the size of a district and its population, i.e. the population density is stable across districts, so that  $\mathcal{G}$  coefficients can be calculated at a pixel level.

Let us begin by analyzing the distribution of aggregated income at a district level in Peru for the year 2001.<sup>e</sup> As presented in the examples of Subsec. 3.2, the computation of  $\mathcal{C}$  is direct from Eq. (6) once the three parameters of such an index are calculated. Table 1 presents the results obtained for each parameter, and their corresponding error measures, using the MASS software. In the case of the Peruvian income maps, the support dimension of the perfect equality case  $[f(\alpha_0^{\mathcal{E}})]$  is less than 2 because of geographical anomalies such as lakes, and the perfect inequality case  $[f(\alpha_0^{\mathcal{T}})]$  equals 0 because the smallest district (La Punta) has only one pixel.

Regarding index  $\mathcal{G}$ , it is important to note that even though its computation seems to be straightforward from Eq. (5), there will be a region size bias in its estimation. This is the case because  $\mathcal{G}$  is computed directly from the income matrix processed by GIS software and the same value repeats itself as many times as the polygon of the district requires. This bias can be interpreted as a population weight in the Gini coefficient if assumption (i) is reasonable.

<sup>d</sup>This software is available at <http://inrm.cip.cgiar.org/vlab/tiki-index.php.f>

<sup>e</sup>Data was taken from the Institute of Statistics and Informatics (INEI) of Peru.

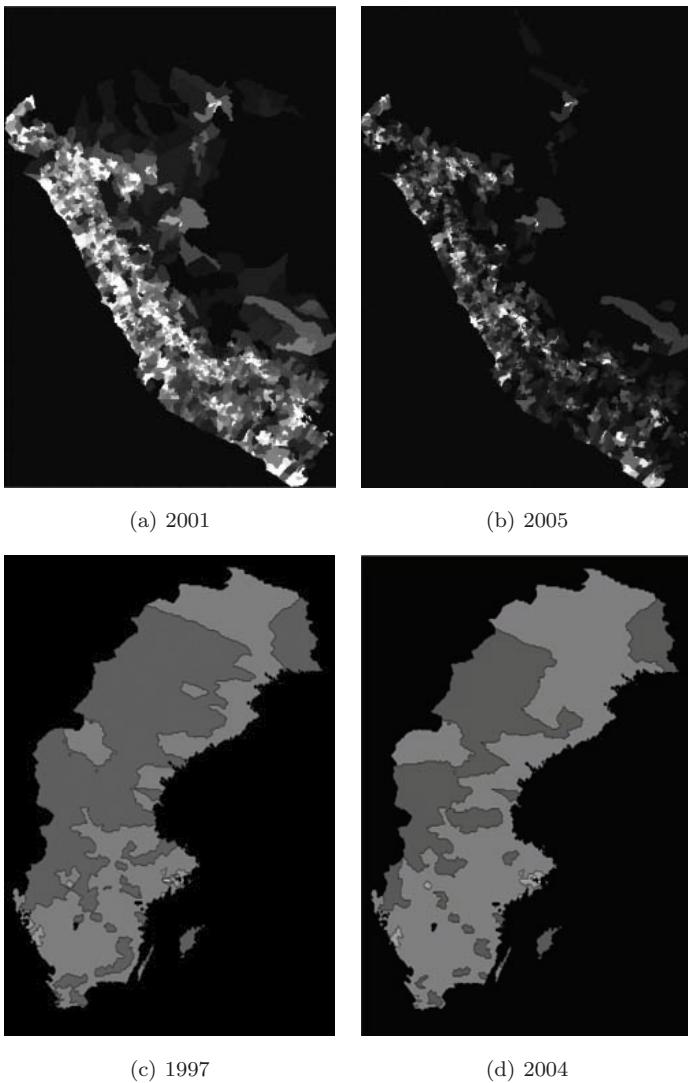


Fig. 7. Income maps from Peru [(a) and (b)] and Sweden [(c) and (d)].

Table 1 presents a comparison between  $\mathcal{C}$  indexes resulting from repeating the previous steps in the Peruvian income map of 2005 and the Swedish income maps of 1997 and 2004.<sup>f</sup> Furthermore it specifies the parameter values and the errors associated with the calculations.

To analyze inequality reduction in a country such as Peru or Sweden, one should simply locate the inequality situation in two years in the  $\mathcal{G}$ - $\mathcal{C}$  plane (see Fig. 8), and observe whether or not the trajectory of its displacement approaches the perfect

<sup>f</sup>Data was taken from the authority of official statistics of Sweden.

Table 1. Parameters in Multifractal Analysis

Parameter	Value	$\alpha$ error range	$f(\alpha)$ error range
Peru			
$f(\alpha_S)^a$	1.584	$\pm 0.027$	$\pm 0.027$
$f(\alpha_S)^b$	1.525	$\pm 0.028$	$\pm 0.028$
$f(\alpha_0^{\mathcal{E}})$	1.94	$\pm 0.001$	$\pm 0.001$
$f(\alpha_0^{\mathcal{T}})$	0	$\pm 0.000$	$\pm 0.000$
Sweden			
$f(\alpha_S)^c$	1.841	$\pm 0.025$	$\pm 0.025$
$f(\alpha_S)^d$	1.832	$\pm 0.027$	$\pm 0.027$
$f(\alpha_0^{\mathcal{E}})$	1.843	$\pm 0.025$	$\pm 0.025$
$f(\alpha_0^{\mathcal{T}})$	0	$\pm 0.000$	$\pm 0.000$

The superscripts a, b, c and d refer to maps in the years 2001, 2005, 1997 and 2004, respectively.

The range of scales ( $\delta$ ) and of moments ( $p$ ) employed for each country are [8–200] and [0.2–1.1] respectively.

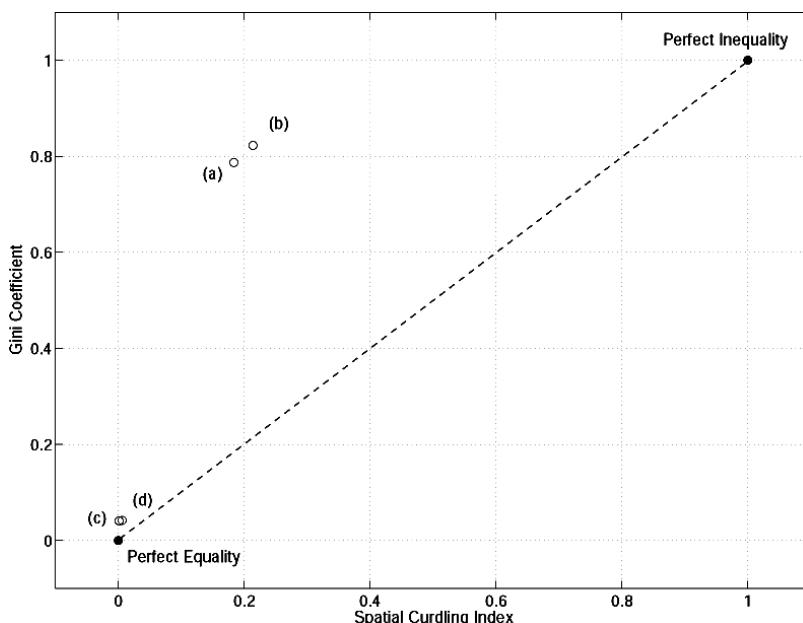


Fig. 8. Inequality transitions in Peruvian [(a) to (b)] and Swedish [(c) to (d)] income maps across the  $\mathcal{C}$ - $\mathcal{G}$  plane.

equality goal  $(0, 0)$ . On one hand, the Peruvian case presents a transition from the year 2001  $(0.1837, 0.7870)$  to the year 2005  $(0.2141, 0.8226)$ , showing that inequality reduction policies are not working not only when employing the traditional Gini coefficient  $\mathcal{G}$ , but also when a spatial entropy measure such as  $\mathcal{C}$  is taken into

the analysis. On the other hand, a different behavior is observed in the Swedish case, where a movement from (0.0011, 0.0408) to (0.0060, 0.0416) evidences a slight deterioration in the Gini coefficient, while index  $\mathcal{C}$  does not vary significantly (see the errors for the Swedish maps in Table 1).

## 5. Concluding Remarks

The spatial curdling index ( $\mathcal{C}$ ) presented in this study adds a new dimension to the traditional measures of inequality by accounting for income curdling when analyzing its distribution. Computational simulations of spatial distributions show that the relation between  $\mathcal{C}$  and traditional measures such as the Gini coefficient ( $\mathcal{G}$ ) forms a concave area, which indicates that policy decisions should account for the spatial structure of a country when one is formulating goals for income inequality reduction.

Empirical applications have shown that better data is needed to get sharper results and avoid assumption (i) (see the second paragraph of the empirical section). On the other hand, it should be noted that even though assumption (ii) may not be totally accurate and the population's density can actually be heterogeneous among districts, we believe that is a small price to pay for accounting for spatial cluster formation when analyzing income inequality.

## Acknowledgments

We thank Prof. Jacques Lévy-Véhel for the transference of the FracLab software, which was instrumental in Sec. 3 of this paper.

## References

- [1] Appleby, S., Multifractal characterization of the distribution pattern of the human population, *Geog. Anal.* **28** (1996) 147–160.
- [2] Atkinson, A., On the measurement of inequality, *J. Econ. Theory.* **2** (1970) 244–263.
- [3] Galam, Y. and Shapir, Y., Sociophysics: A mean behavior model for the process of strike, *J. Math. Soc.* **9** (1982) 1–13.
- [4] Calvet, L. and Fisher, A., Multifractality in asset returns: Theory and evidence, *Rev. Econ. Stat.* **84** (2002) 381–406.
- [5] Champernowne, G., A model of income distribution, *Econ. J.* **63** (1953) 318–351.
- [6] Champernowne, G., A comparison of measures of inequality of income distribution, *Econ. J.* **84** (1974) 787–816.
- [7] Di Mateo, T., Multi-scaling in finance, *Quant. Fin.* **7** (2007) 21–36.
- [8] Gini, G., Variabilitá e mutabilitá, in *Memorie di Metodologia Statistica*, eds. Pizetti, E. and Salvemini, T. (Rome Libreria Eredi Virgilio Veschi, 1955).
- [9] Kolmogorov, A., The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers, *Dok. Akad. Nauk SSSR* **30** (1941) 299–303.
- [10] Kolmogorov, A., Dissipation of energy in the locally isotropic turbulence, *Dok. Akad. Nauk SSSR* **32** (1941) 16–18.
- [11] Feder, J., *Fractals* (Plenum, 1988).
- [12] Lorenz, M., Methods of measuring the concentration of wealth, *Pub. Am. Stat. Assoc.* **9** (1905) 209–219.

- [13] Lydall, H., The distribution of employment incomes, *Econometrica* **27** (1959) 110–115.
- [14] Mimkes, J., Society as a many-particle system, *J. Therm. Anal.* **60** (2000) 1055–1069.
- [15] Mandelbrot, B., The Pareto–Lévy law and the distribution of income, *Int. Econ. Rev.* **1** (1960) 79–106.
- [16] Mandelbrot, B., *Fractals and Scaling in Finance* (Springer-Verlag, 1997).
- [17] Parisi, G. and U. Frisch, On the singularity structure of fully developed turbulence, in *Geophysical Fluid Dynamics*, eds. Ghil M., Benzi, G. and Parisi, G. (Amsterdam, 1985).
- [18] Posadas, A. *et al.*, Characterization of soil particle-size distributions, *Sci. Soc. Am. J.* **65** (2001) 1361–1367.
- [19] Turiel, A., Pérez-Vicente, C. and Grazzini, J., Numerical methods for the estimation of multifractal singularity spectra on sampled data: a comparative study, *J. Comput. Phys.* **216** (2006) 362–390.