

Notions of Scale in Geosciences

Wolfgang Förstner

Institute for Photogrammetry
University of Bonn, Germany. E-mail: wf@ipb.uni-bonn.de

Abstract. ¹ The paper discusses the notion *scale* within geosciences. The high complexity of the developed models and the wide range of participating disciplines goes along with different notions of scale used during data acquisition and model building. The paper collects the different notions of scale shows the close relations between the different notions: map scale, resolution, window size, average wavelength, level of aggregation, level of abstraction. Finally the problem of identifying scale in models is discussed. A synopsis of the continuous measures for scale links the different notions.

1 Preface

The notion scale regularly is used in geosciences referring to the large variety of size in space and time of events, phenomena, patterns or even notions.

This paper is motivated by the diversity of notions of scale and the diversity of problems related to scale which hinder a seamless communication among geoscientists. The problem is of scientific value as any formalization establishing progress needs to clarify the relation between old and new terms used.

Each notion of scale is model related as it is just another parameter and as nature is silent and does not tell us its internal model. The notion scale therefore only seemingly refers to objects, phenomena or patterns in reality. Actually it refers to models of these objects, phenomena or patterns and to data acquired to establish or prove models about objects, phenomena or patterns.

Each formalized model needs to *adequately* reflect the phenomena in concern, thus a mathematical model used in geoscience has no value in its own. The diversity of notions of scale therefore is no indication of conflicting models but of different views of reality. Integrating or even fusing models of the same phenomenon but referring to different notions of scale requires a very careful analysis of the used notion *scale*.

As there is no a priori notion of scale, integrating models with different notions of scale may result in new models with a new notion of scale, only being equivalent to the old notions in special cases. This effect is nothing new and well known in physics, where e.g. the transition from the deterministic Newtonian model of gravity to the probabilistic model of quantum physics lead to new notions of space and time.

¹ appeared in: Neugebauer, H. J. & Simmer, C. (Eds.): Dynamics of Multi-Scale Earth Systems, Lect. Notes in Earth Sc., Vol. 97, 2003, Springer, Heidelberg, pp. 17–39

We distinguish two notions of the term *model*:

1. *model of the first type*: The notion model may mean a *general model*, e.g. a mathematical model, a law of nature, thus refer to a *class of phenomena*, such as the models in physics. This type of model is denoted as *model of the first type* or mathematical or generic model, depending on the context.
2. *model of the second type*: the notion model may mean a *description* of a phenomenon being in some way similar to reality, thus refer to a *specific phenomenon*, such as a digital elevation model. This type of model is denoted as *model of the second type* or specific model.

Modeling means building a model. It therefore also has two meanings: building a model of the first type, thus finding or combining rules, relations or law of natures, e.g. when specifying a process model; and building a model of the second type, thus developing a description of a situation, e.g. the layered structure of sediments in a certain area.

Scale may directly appear in general models, e.g. a wavelength. It seems *not* to appear in descriptions, thus models of the second type, e.g. in a digital elevation model consisting of a set of points $z(x_i, y_i)$. However, it appears in the structures used for describing a phenomenon or in parameters derived from the descriptions, e.g. the point density of the points of a digital elevation model.

As the notion scale is model related we describe the different notions of scale as they are used in the different steps of modeling in geosciences and try to show their mutual relations. The paper is meant to collect well known tools and notions which may help mutual understanding within neighbored disciplines. The paper tries to formalize the used notions as far as possible using a minimum of mathematical tools.

We first want to discuss crisp and well defined notions of scale all referring to observations. Later we will discuss notions which are semiformal or even without an explicit numerical relation to the size of the object in concern, which however, are important in modeling structures. Finally we will discuss the problems occurring with the notion scale in process models.

2 Continuous Measures of Scale

There exist a number of well definable continuous numerical measures of scale, which we want to discuss first. They all describe data, thus models of the second type. They refer to different notions of scale. But there are close functional relationships, which we collect in a synopsis.

2.1 Scale in Maps

Scale in cartography certainly is the oldest notion. Other notions of scale can be seen as generalizations of this basic notion. The reason is the strong tendency of humans to visualize concepts in space, usually 2D space, e.g. using maps,

sketches or in other graphical forms for easing understanding and communication.

Maps are specific models of reality and comparable to observations or data.

The *map scale* refers to the ratio of lengths l' in the map to the corresponding lengths l in reality:

$$s = \frac{l'}{l} [1] \quad (1)$$

It is a dimensionless quantity. A large scale map, e. g. 1 : 1,000 shows much detail, a certain area in the map corresponds to a small detail in reality. Small scale maps, e. g. 1 : 1,000,000 show few details, a certain area in the map corresponds to a large area in reality.

The *scale number*

$$S = \frac{1}{s} [1] \quad (2)$$

behaves the opposite way.

In maps of the earth, e. g. topographic or geological maps, the scale is approximately constant only in small areas. This is due to the impossibility to map the earth's surface onto a plane with constant scale.

One therefore has to expect scale to vary from position to position. Therefore scale in general is *inhomogeneous* over space. The spatial variation of scale may be described by a scale function:

$$s = s(x', y') = \frac{dl'(x', y')}{dl(x, y)}$$

This implicitly assumes the ratio of lengths to be invariant to direction. This invariance w. r. t. direction is called *isotropy*. Maps with isotropic scale behaviour are called *conformal*. Conformal maps of a sphere can be realized and are called. The Gauss-Krüger-mapping and the stereographic mapping of the earth are conformal, in contrast to the gnomonic mapping.

In general scale is not only inhomogeneous thus spatially varying but also varies with the direction. Then scale is called *anisotropic*. The ratio of a small length in the map to its corresponding length in reality then depends on the direction α

$$s = s(x', y', \alpha') = \frac{dl'(x', y', \alpha')}{dl(x, y, \alpha)}$$

e. g. the map of the geographical coordinates (λ, ϕ) directly into (x, y) -coordinates by

$$x = \lambda \quad y = \phi$$

shows strong differences in scale in areas close to the poles as there the y -coordinate overemphasizes the distance of points with the same latitude λ (cf. fig. 2 left).

The four possible cases are shown in fig. 1.

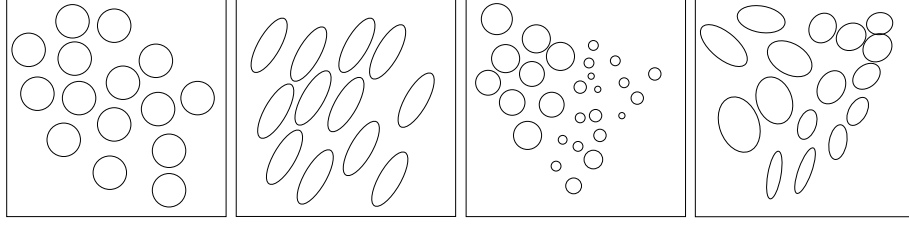


Fig. 1. Homogeneous isotropic scale, homogeneous anisotropic scale, inhomogeneous isotropic scale and inhomogeneous anisotropic scale

We will show how to formalize the concept of anisotropy and inhomogeneity when discussing scale as window size during analysis.

Here we only want to mention the so-called *indicatrix of Tissot* (cf. (Tissot 1881), pp. 20 ff and (Snyder 1987), pp. 20): It is the image of a differentially small circle. In the case of a mapping from one plane to another it can be determined from the mapping equations

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x'(x, y) \\ y'(x, y) \end{pmatrix}$$

with the Jacobian

$$\mathbf{J}(x, y) = \begin{pmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} \end{pmatrix} \bigg|_{(x, y)}$$

thus the differentially local mapping $(dx', dy')^\top = \mathbf{J}(dx, dy)^\top$. Then the differentially small circle in the local (dx, dy) -coordinate system $dx^2 + dy^2 = 1$ is mapped to the differentially small ellipse, namely Tissot's indicatrix, in the local (dx', dy') -coordinate system

$$(dx' \ dy') [\mathbf{J}(x', y')^\top \mathbf{J}(x', y')]^{-1} \begin{pmatrix} dx' \\ dy' \end{pmatrix} = 1$$

In case of map projections from a curved surface to plane or a curved surface the equations are slightly different. Examples for this detailed scale analysis for two classical map projections are shown in fig. 2.

We will observe inhomogeneity and anisotropy also when analyzing other notions of scale and formalize the concept of anisotropy and inhomogeneity when discussing scale as window size during analysis.

2.2 Resolution

Resolution often is used to refer to the scale of gridded data, representing models₂, either in time or space. It refers to the smallest unit of discourse or the least distinguishable unit.

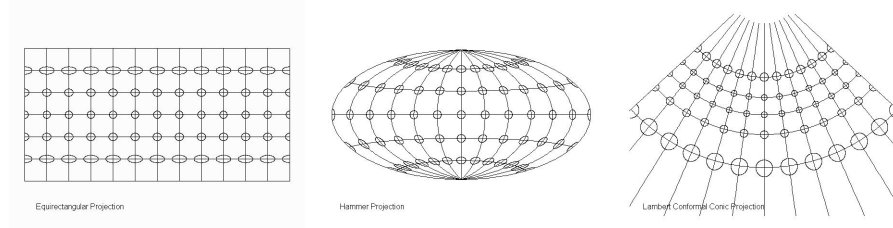


Fig. 2. Map scale for three map projections with Tissot's indicatrices indicating the local scale distortions. Left: equirectangular projection (anisotropic and inhomogeneous). Middle: Hammer projection (anisotropic and inhomogeneous). Right: Lambert conic conformal projection (isotropic and inhomogeneous), cf. <http://www.ftss.ilstu.edu/microcam/graphics.htm#tissot>. [Datum]

It therefore also can be generalized to non spatial concepts, as e.g. when resolving differences of species of plants during their classification.

Resolution of gridded data refers to the grid width in space and time:

$$\Delta x [\text{m}] \quad \text{or} \quad \Delta t [\text{s}] \quad (3)$$

The grid size Δx standing also for Δy or Δz in case of two-dimensional or three-dimensional data. Resolution obviously has a dimension. High resolution refers to small grid widths, low resolution to large grid widths.

This notion is also used in case of data given up to a certain digit, e.g. the height 12.3 m indicates the resolution to be $\Delta z = 1 \text{ dm}$.

Resolution in case of gridded data often is linked to the envisaged application. It often reflects the accuracy of the model behind the data or for which the data are used.

This implicit assumption leads to confusion or even to severe errors in case the meta information of the data, i.e. the purpose and the instrumentation of the data acquisition is not known or given. The resolution, in the sense of the last digit, therefore should reflect the underlying accuracy, 12.4 km obviously not saying the same as 12.40 km or even 12.400 km.

The sampling theorem inherently limits the reconstruction of the underlying high resolution signal from sampled data. If the underlying signal is band-limited, i.e. is smooth to some degree measured by the wavelength of the smallest detail, the grid width should be adapted to the smallest detail, namely being not larger than half the minimum wave length $\lambda_{\min}/2$. Practically one should take the so-called Kell-factor $K \approx 0.7 - 0.8$ into account, which states that the grid size should be $K\lambda_{\min}/2$ in order to capture the smallest details, This especially is important for signals with isolated undulations, which in case of homogeneous sampling dictate the grid width.

Again resolution may be inhomogeneous and anisotropic:

Geological models (Siehl 1993) usually have a much higher resolution Δz in z -direction than in (x, y) -direction ($\Delta x, \Delta y$) reflecting the different variability

of the phenomena in planimetry and height. They are highly anisotropic. This can also be seen from the usually small slope, i.e. the small height variations of the terrain w. r. t. to the planar grid size.

On the other hand if the data are not given in a regular grid they may show large differences in density e.g. where undulations in the surface are locally higher. Binary, Quad- or octrees (Samet 1989) or triangular irregular networks usually show a high degree of inhomogeneity.

2.3 Window Size for Smoothing

When characterizing phenomena by local features or properties being below the minimum size they either are

1. of no interest or
2. aggregated from a higher resolution.

Therefore the *window size* used for determining local properties of a signal often can be a measure for scale.

We want to take the example of smoothing to show one way of formalizing inhomogeneity and anisotropy already discussed before and depicted in fig. 1.

Averaging all data within a neighborhood with a representative radius r leads to a smoothed signal with an inherent scale r . Several weighting schemes are in use: unit weight for all observations leads to a rectangular type window for one-dimensional signals, or top box or cylindrical type window for two-dimensional signals. In case points further away from the center of the window one might use triangular or bell-shaped weighting functions in 1D or cone-shaped or bell-shaped weighting functions in 2D. We will derive a measure, denoted as

$$\sigma_x [\text{m}] \quad \text{or} \quad \sigma_t [\text{s}] \quad (4)$$

in space or time signals, which is more suitable for describing the scale of a window, as it refers to the *average radius* taking the weighting scheme into account.

In case the window is given by the weighting function $w(x)$ or the weights $w'(x)$ with the sum of all weights being 1, the scale can be computed from

$$\sigma_x = \sqrt{\int x^2 w(x) dx} \quad \text{or} \quad \sigma_x = \sqrt{\sum_x x^2 w'(x)} \quad (5)$$

depending on whether the window is continuous or discrete. Obviously the scale corresponds to the standard deviation of the function when interpreted as a probability density function.

When using a box filter, thus a filter with constant weight $1/r$ within a window $(-r/2, r/2)$ we obtain

$$\sigma_{\text{Box}} = \frac{1}{\sqrt{12}} r \quad (6)$$



Fig. 3. Windows for one and two dimensional signals having the same width, thus leading to the same scale during sampling. As smoothing with box-type and bell-shaped windows with the same width leads to the same degree of smoothing, we obtain the relation between resolution, representing a box-type window, and window-size. Observe, the radius of the window, where the weight is non-zero is not a suitable measure, as it theoretical would be infinite for Gaussian windows.

When using a Gaussian bell shaped weighting function $G_\sigma(x)$ as filter the scale is identical to the parameter σ of the Gaussian function (cf. fig. 3, left).

In two dimensions the circular symmetric Gaussian is given by $G_\sigma(x, y) = G_\sigma(x)G_\sigma(y)$. It can be represented by a circle with radius s . Usually a homogeneous weighting is used to derive smoothed versions of a signal, e.g. when smoothing height or temperature data.

Inhomogeneous windowing can be achieved by spatially varying $\sigma(x, y)$. This allows to preserve finer structures by locally applying weights with smaller windows.

As soon as the window needs to be anisotropic, i.e. elliptic there is not one scalar describing the local scale. In case of different scales in x - and y -direction one may use the Gaussian: $G_{\sigma_x \sigma_y}(x, y) = G_{\sigma_x}(x) G_{\sigma_y}(y)$ which shows just to use different window sizes in x - and y -direction. This type of anisotropic smoothing only is adequate if the spatial structure is parallel to one of the two coordinate systems. It can be represented by an ellipse. Because the anisotropy usually is not parallel to the reference coordinate system one in general needs a rotational term.

This can be achieved by generalizing the scalar value of the scale to a 2×2 matrix Σ representing an ellipse: $(x \ y) \Sigma^{-1} (x \ y)^T = 1$ which in case the scale matrix $\Sigma = \text{Diag}(\sigma_x^2, \sigma_y^2)$ contains the squares of the scales in x - and y -direction reduces to the special case above. Because of symmetry the scale matrix is defined by three values:

1. the larger scale σ_1
2. the smaller scale σ_2
3. the direction α in which the larger scale is valid

We then have

$$\Sigma = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}^T \quad (7)$$

with the rotation matrix indicating the azimuth α of the most dominant scale.

We now can explicitly represent anisotropic inhomogeneous scale by a spatially varying *scale matrix*: $\Sigma = \Sigma(x, y)$. This makes the visualization of the different situations of inhomogeneity and anisotropy in fig. 1 objective.

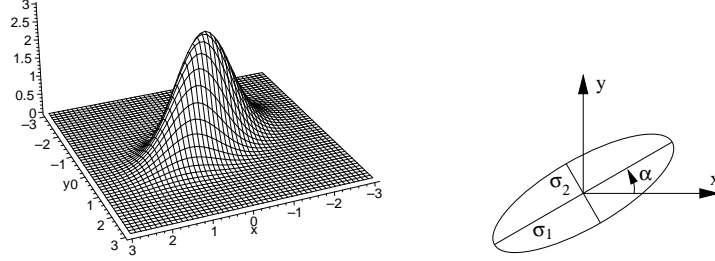


Fig. 4. Left: An anisotropic Gaussian window. It may directly related to the local form of a surface indicating its orientation and its two scales in the two principle directions. Right: Parameters for describing the anisotropic scale.

An example for an anisotropic window is given in fig. 4. Observe the form to be similar to a small hill having different width in the principle directions and not being oriented parallel to the coordinate system. This enables to link the window size and form with the intended scale of the analysis, e.g. when analyzing landforms.

2.4 Average Wavelength

The notion scale often is used to indicate the average size of objects or phenomena in concern, both in space and time. Though the average size usually is not a very precise value and depends on the crispness of the context the value as such can be defined well and thus related to other concepts of scale.

In case we have onedimensional signals with a dominant wavelength we may identify the scale of that signal with the average wavelength

$$\bar{\lambda} \text{ [m]} \quad \text{or} \quad \bar{\lambda} \text{ [s]} \quad (8)$$

Fig. 5 and 6 show two simulated signals with dominant wavelength, one with average wavelength $\bar{\lambda} = 40$ the other with $\bar{\lambda} = 160$.

As the wavelength λ [m] or λ [s] of a periodic signal $z(x)$ or $z(t)$ is $\lambda = 1/f$, where f [1/m] or f [Hz] is the frequency we could also average the frequency. This should take the amplitudes of the periodic signals into account.

In case we take the quadratic mean, weight it with the squared magnitude $Z(f)$ of the amplitude spectrum², thus with the power spectrum $P_z(f) = |Z(f)|^2$

² We use the Fourier transform $F(u) = \int \exp(-j2\pi ux) f(x) dx$ according to (Castleman 1996).

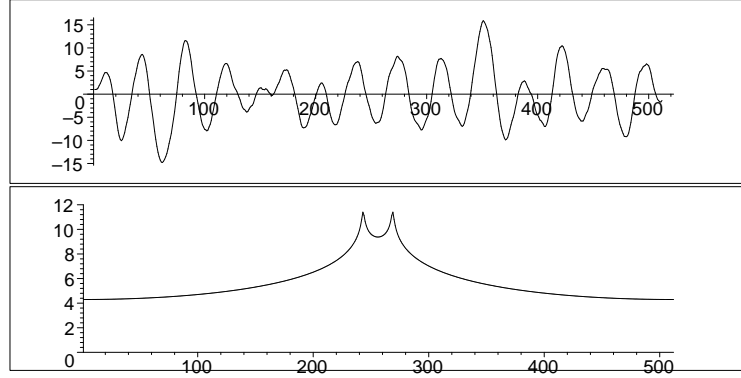


Fig. 5. A signal with short wavelength, $\bar{\lambda} = 40$ (above) together with its amplitude spectrum (below). It is a profile generated using an autoregressive process (cf. below sect. 4.1). The amplitude spectrum shows two peaks, symmetric with respect to the origin, which is in the middle of the graph. The distance between the peaks is large (compared to the following figure) indicating a large frequency, thus a short wavelength.

(Castleman 1996) and normalize, we obtain the *effective bandwidth* (Ryan et al. 1980) assuming independent and homogeneous noise

$$b = \sqrt{\frac{\int_{-\infty}^{+\infty} f^2 P_z(f) df}{\int_{-\infty}^{+\infty} P_z(f) df}}$$

It can easily be derived from the original signal *without performing a spectral analysis*:

$$b = \frac{1}{2\pi} \frac{\sigma_z}{\sigma_{z'}}$$

where

$$\sigma_z = \sqrt{\frac{1}{(N-1)} \sum_{i=1}^N (z(x_i) - \mu_z)^2} \quad \text{and} \quad \sigma_{z'} = \sqrt{\frac{1}{N} \sum_{i=1}^N (z'(x_i))^2}$$

are the standard deviations of the signal and its derivative, assuming the mean of the slope to be zero.

Thus we may derive an estimate for the *signal internal scale*, namely the average wavelength from

$$\bar{\lambda} = 2\pi \frac{\sigma_z}{\sigma_{z'}} \quad (9)$$

This is intuitive, as a signal with given amplitude, thus given σ_z , has longer average wavelength if it is less sloped. Of course this is only meaningful in case there is a dominant wavelength.

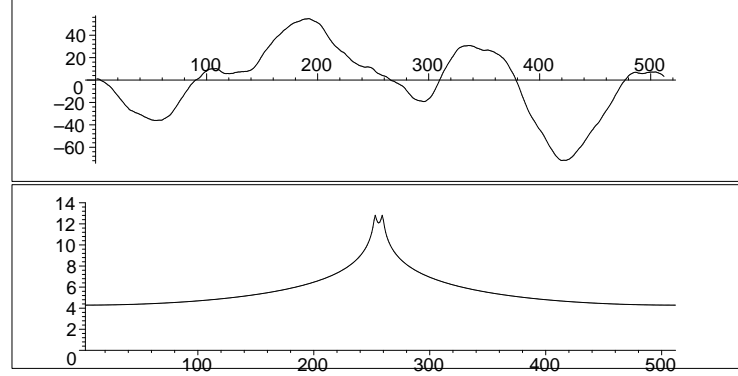


Fig. 6. A signal with long wavelength, $\bar{\lambda} = 160$ (above) together with its amplitude spectrum (below). It is a profile generated using an autoregressive process (cf. below). The amplitude spectrum shows two peaks, symmetric with respect to the origin, which is in the middle of the graph. The distance between the peaks is small (compared to the previous figure) indicating a small frequency, thus a long wavelength.

For regularly spaced data we may determine the standard deviation of the signal and the slope from:

$$\sigma_z = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (z_i - \mu_z)^2} \quad \sigma_{z'} = \sqrt{\frac{1}{N-2d} \sum_{i=d+1}^{N-d} \left(\frac{z_{i+d} - z_{i-d}}{2d} \right)^2} \quad (10)$$

showing no systematic error if N is large enough, say $N > 40$. One may need to smooth the data before calculating the derivative. The distance d for calculating the slope can be used to apply some (crude) smoothing.

For two-dimensional data $z = z(x, y)$ we can also derive the data internal scale, which now may be anisotropic, as we have discussed above. We obtain a 2×2 -matrix for the average squared wavelength for two-dimensional data

$$\bar{\mathbf{A}} = 4\pi^2 \sigma_z^2 \boldsymbol{\Sigma}_{z'z'}^{-1} \quad (11)$$

with the matrix

$$\boldsymbol{\Sigma}_{z'z'} = \begin{pmatrix} \sigma_{z_x}^2 & \sigma_{z_x z_y} \\ \sigma_{z_x z_y} & \sigma_{z_y}^2 \end{pmatrix}$$

containing the variances $\sigma_{z_x}^2$ and $\sigma_{z_y}^2$ of the slopes $z_x = \partial z / \partial x$ and $z_y = \partial z / \partial y$ and their covariance $\sigma_{z_x z_y}$. The wavelength in general will be anisotropic, thus depend on the direction. The maximum and the minimum wavelength show in two perpendicular directions. Their value results from the square-roots of the eigenvalues of $\bar{\mathbf{A}}$. The direction of the largest wavelength can be determined in analogy to (9) from

$$\alpha = \frac{1}{2} \text{atan2}(2\bar{\mathbf{A}}_{12}, \bar{\mathbf{A}}_{11} - \bar{\mathbf{A}}_{22})$$

being an inversion of (7).

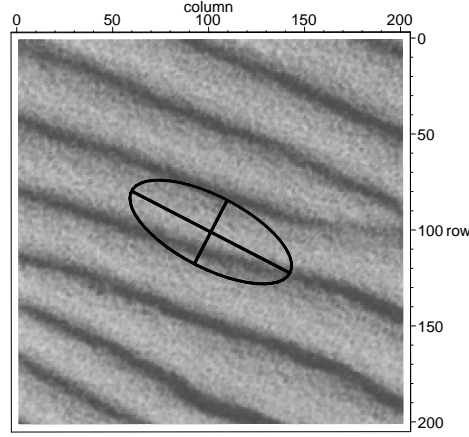


Fig. 7. Wavelengths and orientation estimated from an image of a ripple structure. The derivatives are calculated similar to (10) with $d = 5$. The wavelength of the ripples corresponds to the *shorter diameter* of the scale-ellipse. It locally is a bit larger, as it has been calculated from the complete image shown.

As an example, the data inherent scale matrix has been calculated from the ripple image fig. 7. The ellipse visualizes the scale in the different directions. The diameters indicate the the estimated wavelength in that orientation. The wavelength of the ripples is approximated by the *shorter* diameter, which appears to be a fairly good approximation. The wavelength in the direction along the ripples would only be infinity, in case the ripples would be perfectly straight, thus the directional variation of the ripples leads to a finite, but long estimated wavelength.

2.5 Smoothness

In contrast to the previous notions of scale which are related to the extension of the object in space or time we also use the notion scale related to the smoothness of the object.

Objects loose crispness in case they are viewed from further away in space or are further in the past. We use the notion of a coarser scale if we talk about objects, in case we are not interested in the details, even if we do not change map scale or the grid size when presenting the object. Objects then appear generalized.

In case smoothness is achieved by an averaging process we may formalize this by linking it to the used window size Δx or Δt .

Assume a surface is rough or noisy with a certain standard deviation σ_n . Smoothing the data with a window size of width σ_x or σ_t reduces the roughness of the noise by a *factor*

$$k = \sqrt{12} \frac{\sigma_x}{\Delta x_0} \quad \text{or} \quad k = \sqrt{12} \frac{\sigma_t}{\Delta t_0} \quad (12)$$

Thus the smoothing factor k is the larger the larger the smoothing is. The degree of smoothing is directly related to the scale parameter induced by the size of the smoothing window. Obviously this is a relative measure and requires to define a basic unit in which σ_x or σ_t is measured. In case of gridded data this is the basic resolution Δx_0 or Δt_0 of the data. Smoothing then would lead to data which have less resolution, thus could be represented with a coarser grid with grid width Δx or Δt .

This relation easily can be derived from analyzing the effect of smoothing gridded data $z(x, y)$ with a box filter. Assuming the data to have random errors and have the same standard deviation σ_z , then smoothing with a $r \times r$ -window leads to a standard deviation of the smoothed data, assuming $\Delta x_0 = 1$ for this derivation

$$\bar{z}(x, y) = \frac{1}{r^2} \sum_{u=x-r/2}^{x+r/2} \sum_{v=y-r/2}^{y+r/2} z(x-u, y-v)$$

which is the average of r^2 values which therefore has standard deviation

$$\sigma_{\bar{z}} = \frac{1}{r} \sigma_z = \frac{1}{\sqrt{12} \sigma_x} \sigma_z$$

where the window-size $\sqrt{12} \sigma_x$ of the box filter from (6) is used. Thus the noise is reduced by a factor $k \approx 3.5 \sigma_x$.

The relation only holds for noise in the data. In case the data are already smooth and are further smoothed the relations are more complicated.

2.6 Synopsis of Continuous Scale Measures

We now want to find the relations between these continuous measures.

We want to distinguish:

1. Map scale s as dimension-less quantity indicating the zooming factor usually being much smaller than 1, thus indicating a reduction.
2. Map scale number S as dimensionless quantity indicating the reduction factor usually being much larger than 1.
3. Resolution Δx or Δt as spacing of gridded data also the smallest distinguishable unit in the data.
4. Window-size σ_x or σ_t as average radius of function for averaging or smoothing data. The value is identical to the standard deviation of a Gaussian window.
5. Average wavelength $\bar{\lambda}_x$ or $\bar{\lambda}_t$ for signals with a repetitive structure.

6. Relative smoothness k of the data referring to smoothness of original data.

Having six measures we need at least five relations in order to be able to derive one from another. We have the following relations, which only hold in certain contexts, which are indicated in the equations:

1. Map scale and map scale number (table 1, row 3, col. 3)

$$S = \frac{1}{s}$$

This relation is generally valid.

2. Resolution and window size (table 1, row 4, col. 4)

$$\sigma_x \stackrel{s}{=} \sqrt{\frac{1}{12}} \Delta x \approx 0.28 \Delta x$$

A rectangular window with width Δx has the same effect in smoothing as a Gaussian window with parameter σ_x . The notion resolution Δ_x refers to the width of a rectangular window whereas the notion window size refers to the parameter of a Gaussian window of equivalent smoothing effect.

Obviously, the *relation* between resolution and window size is made *via the effect of smoothing*, indicated by the s above the equal sign.

3. Map scale and resolution can be linked in case we define the basic resolution Δx_{map} of a map. This may e.g. relate to the printing resolution, e.g. 600 dpi, leading to a resolution of $\Delta x_{\text{map}} = 25.6 [\text{mm}]/600 \approx 40 \mu\text{m}$. As this resolution is on the printed map, thus at map scale we have the relation (table 1, row 5, col. 5)

$$s = \frac{\Delta x_{\text{map}}}{\Delta x}$$

where Δx now relates to the resolution in object space, e.g. for a map of scale 1:10,000 printed with 600 dpi we have $\Delta x \approx 40 \text{ cm}$.

This relation is generally valid.

4. Average wavelength $\bar{\lambda}_x$ and window-size σ_x can be linked the following way: In case we smooth *a pure noise signal* the resulting signal will be smooth. Its average wavelength is (table 1, row 6, col. 6)

$$\bar{\lambda}_x \stackrel{sw}{=} 2\sqrt{2} \pi \sigma_x \approx 9 \sigma_x$$

The reason is: The Gaussian window function $G_\sigma(x)$ has effective bandwidth $b = 1/2\sqrt{2} \pi \sigma_x$. The graph in fig. 8 indicates this relation between window size and average wavelength to be plausible.

Obviously, the *relation* between wavelength and window size is made *via the effect of smoothing white noise*, indicated by the sw above the equal sign.

5. Relative smoothness k and window size σ_x can be linked if we refer to some *basic resolution* Δx_0 . It is defined as that resolution, where the data show

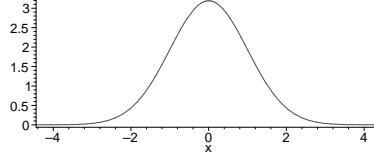


Fig. 8. Gauss function $8 \cdot G_1(x)$ with $\sigma = 1$ in the range $[-\sqrt{2\pi}, \sqrt{2\pi}] \approx [-4.5, 4.5]$. It appears as one wave with length $\lambda = 2\sqrt{2\pi} \approx 9$.

most detail, thus show *highest roughness* or have original uncorrelated measurement errors. Then we have an increase k of smoothness or a reduction $1/k$ of noise. Thus we have (table 1, row 7, col. 6)

$$k \stackrel{sw}{=} \sqrt{12} \frac{\sigma_x}{\Delta x_0}$$

The definition of this notion of scale refers to gridded data, but can also be used for non gridded data, e.g. when reducing the density of data, then indicating the ratio of the average point distance after and before the data reduction.

Obviously, the *relation* between relative smoothness and window size is made *via the effect of smoothing white noise*.

Obviously we have three scale measures being proportional, namely resolution, window size and average wavelength, related by

$$\bar{\lambda}_x \stackrel{sw}{=} 2\sqrt{2} \pi \sigma_x \stackrel{s}{=} \sqrt{\frac{2}{3}} \pi \Delta x \quad \text{or} \quad \bar{\lambda}_x \stackrel{sw}{\approx} 9 \sigma_x \stackrel{s}{\approx} 2.56 \Delta x \approx 1.3 (2 \Delta x)$$

The factor 1.3 indicates, that waves of length λ do not need 2 samples but by a factor 1.3 more (table 1, row 6, col. 5). This corresponds to experience that the actual resolution Δx_a of gridded data is by a factor K less, thus $\Delta x_a = \Delta x / K$ than the Nyquist theorem (Castleman 1996) in case of band limited signals would say, namely Δx , the Kell factor K which is between 0.7 and 0.8.

Using these five relations we can derive all possible relations between the numerical measures for scale. They are collected in the following table. The guiding relations are shown in bold face letters.

Comments: We want to discuss the relevance of the other relations between the different measures of scale, this will make their dependency on the context explicit:

- map scale/window size $s \stackrel{sw}{=} 0.28 \Delta x_{\text{map}} / \sigma_x$ (table 1, row 2, col. 6):

In order to visualize the details in the map at the required resolution Δx_{map} and in case one has taken averages of the objects details with a window of size σ_x , then one should take a map scale of approx. s . The map resolution in this case needs not be the printing resolution, but the minimum size of details in the map to be shown separately.

Table 1. Relations between numerical notions of scale. Each row contains the name, the symbol and the dimension of each measure of scale. In the 6 last columns each measure is expressed as a function of the others, the relations either hold generally, refer to smoothing $^{(s)}$, or refer to smoothing white noise or detail of highest roughness $^{(sw)}$. Constants: $c_1 = 1/\sqrt{12} \approx 0.288$, $c_2 = 2\sqrt{2}\pi \approx 8.89$, thus $c_1^{-1} \approx 3.45$, $c_2^{-1} \approx 0.112$, $c_1 c_2 \approx 2.56$, $c_1^{-1} c_2^{-1} \approx 0.390$

	1	2	3	4	5	6	7	8
1	notion	dim.	s	S	Δx	σ_x	$\bar{\lambda}_x$	k
2	map scale s	1	—	$\frac{1}{S}$	$\Delta x_{\text{map}} \frac{1}{\Delta x}$	$c_1 \Delta x_{\text{map}} \frac{1}{\sigma_x}$	$c_1 c_2 \Delta x_{\text{map}} \frac{1}{\bar{\lambda}_x}$	$\frac{\Delta x_{\text{map}}}{\Delta x_0} \frac{1}{k}$
3	map scale number S	1	$\frac{1}{s}$	—	$\frac{1}{\Delta x_{\text{map}}} \Delta x$	$\frac{1}{c_1 \Delta x_{\text{map}}} \sigma_x$	$\frac{1}{c_1 c_2 \Delta x_{\text{map}}} \bar{\lambda}_x$	$\frac{\Delta x_0}{\Delta x_{\text{map}}} k$
4	resolution Δx	$[m]$, $[s]$	$\Delta x_{\text{map}} \frac{1}{s}$	$\Delta x_{\text{map}} S$	—	$c_1^{-1} \sigma_x$	$c_1^{-1} c_2^{-1} \bar{\lambda}_x$	$\Delta x_0 k$
5	window size σ_x	$[m]$, $[s]$	$c_1 \Delta x_{\text{map}} \frac{1}{s}$	$c_1 \Delta x_{\text{map}} S$	$c_1 \Delta x$	—	$c_2^{-1} \bar{\lambda}_x$	$c_1 \Delta x_0 k$
6	average wavelength $\bar{\lambda}_x$	$[m]$, $[s]$	$c_1 c_2 \Delta x_{\text{map}} \frac{1}{s}$	$c_1 c_2 \Delta x_{\text{map}} S$	$c_1 c_2 \Delta x$	$c_2 \sigma_x$	—	$c_1 c_2 \Delta x_0 k$
7	relative smoothness k	1	$\frac{\Delta x_{\text{map}}}{\Delta x_0} \frac{1}{s}$	$\frac{\Delta x_{\text{map}}}{\Delta x_0} S$	$\frac{1}{\Delta x_0} \Delta x$	$\frac{1}{c_1 \Delta x_0} \sigma_x$	$\frac{1}{c_1 c_2 \Delta x_0} \bar{\lambda}_x$	—

E. g. if the map should show a resolution of $\Delta x_{\text{map}} = 1$ mm and the smoothing window size at object scale is $\sigma_x = 10$ m one should choose a map scale $s \approx 0.281/10,000 \approx 1 : 35,000$.

Vise versa, a map showing details at a separation distance of Δx_{map} can be assume to have been generated by averaging object details with a window of size σ_x

- map scale/average wavelength $s \stackrel{sw}{=} 2.56 \Delta x_{\text{map}} / \bar{\lambda}_x$ (table 1, row 2, col. 7):
This relation is similar to the previous one: If details have wavelength $\bar{\lambda}_x$ and one wants to visualize them in the map at the required resolution Δx_{map} , then one should take a map scale of approx. s .
- map scale/smoothing factor $s \stackrel{sw}{=} \Delta x_0 / \Delta x_{\text{map}} k$ (table 1, row 2, col. 8):
In case one filters a signal with basic resolution Δx_0 with a smoothing factor k and wants to show it in a map with a resolution Δx then one should choose a scale s .

Vices versa (table 1, row 6, col. 3), if one wants to show a signal with basic resolution of Δx_0 in a map at scale s with a resolution in that map of Δx one should apply a smoothing with faktor k .

- resolution/smoothing factor $\Delta x \stackrel{sw}{=} \Delta x_0 k$ (table 1, row 4, col. 8):
This important relations states, that a signal with resolution Δx_0 that is smoothed with a smoothing factor k , can be resampled with Δx . Vice versa (table 1, row 7, col. 5) in case one wants to decrease resolution from Δx_0

to Δx one needs to smooth with a smoothing factor k , being realizable by a smoothing filter with window size $\sigma_x = 0.28\Delta x_0 k$ (table 1, row 5, col. 8).

Obviously, the relations hold under specific conditions, thus one needs to take the individual context of the derivation of the relations into account.

3 Discrete Measures of Scale

Taking the average size of an object as measure for scale already contains uncertainty in its definition. Though the notion of a window size is well defined the average size of objects usually is not. Moreover there are applications where even the average size of a set of phenomena is not adequate to represent the *scale* on which the phenomena are handled. Generalization of river networks or the interpretation of sets of bore holes in no way can be *adequately* characterized by the above mentioned measures. Therefore there exist other notions of scale used in geosciences. Two are collected in the following.

3.1 Level of Aggregation

One of the main properties of spatial objects is the ability to partition them into parts or aggregate them into large units. In object oriented modeling these relations are represented in aggregation or is-part-of hierarchies.

Examples are manifold: The aggregation level of spatial political units as states, counties, towns or of time domains in geology.

In hydrology e. g. the scale of the modeling might be reflected in the size of catchments. Taking an average size of catchments would correspond to the window size mentioned above. However, there is no unique way to average. Taking the minimum size of the catchments as characteristics reflects the resolution. Again, there is a conceptual difference as the compartment sizes are not multiples of the minimum size, unless the basis for the definition are raster data. Also maximum sizes or the size range may be used to characterize scale of modeling.

3.2 Level of Abstraction

An even more general way to perceive scale is the property of an object to be a special case of another one or to generalize a set of objects. In object oriented modeling these relations are represented in specialization or is-a hierarchies.

Examples again are manifold: species on biology are classified in a is-a-hierarchy, phenomena in meteorology or geophysics are also classified in specialization hierarchies. Obviously referring to a more general class of objects is using a coarser scale in reasoning, while referring to very specific objects is using a fine scale in reasoning. This type of scale selection often is used to guide modeling and may lead to specific types of mathematical relations, e. g. different types of partial differential equations.

As the different scale types — aggregation and abstraction — interfere and directly relate to the used ontology, one could use the term *ontological scale*

to indicate the scale depending on the different types of hierarchies. Though it is a notion of scale with quite some degree of uncertainty in its definition, it covers all legends of maps including the abstraction hierarchies of different object types and at the same time the spatial extent of the shown objects reflected in the different signatures, enabling a differentiated description: e.g. large cities are shown with an approximation of their boundary, small are shown with a circle as symbol, principal cities are given an underlined name, other cities not underlined.

The relations between the very different notions of scale are by far not understood enough to guarantee clear specifications for data acquisition, to allow formalization of simulation models or to enable generalization for generating maps on demand.

4 Scale Defining Parameters in Models

The notions discussed so far all refer to description of reality either in observations or data or in models of the second type. Scale also appears in models of the first type aiming at explaining observations. Fourier techniques are the classical tool to represent and analyze scale.

The goal of this subsection is to show the difficulty of predicting scale in models.

Mathematical models, thus models of the first type, contain variables, such as pressure or velocity, which change over space and time, and parameters which are constant. Moreover, boundary conditions heavily influence the behavior over space and time. The notion of scale of a phenomenon is depending on the parameters and the boundary conditions. Thus without a specification of the parameters or the boundary conditions a scale of the phenomenon cannot be told. In non-linear models even small changes of the parameters may change the type of the generated process. Unfortunately there is no canonical way how a scale of the phenomenon is related to the parameters or the boundary conditions. Moreover, the notion of scale in models shows the same variety as in data.

Two examples want to demonstrate the difficulty of identifying some scale in models: both are recursive processes, where the current state x_i depends only on one or a few, say K , previous values

$$x_i = f(\{x_{i-k}\}, p_j), \quad i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K$$

The J parameters p_j specify the underlying the process.

4.1 Parameters in Autoregressive Models

In the first example we only deal with a linear relationship, which allows to use the powerful tools of linear system theory. Autoregressive (AR) models are one of the best understood models for one dimensional signals, such as time processes or terrain profiles. For the underlying theory is found in (Box and Jenkins 1976), (Lücker 1980).

We want to demonstrate the ability to generate models for time series with a given spectral content and the difficulty to interpret the parameters. The models shown here explicitly contain a stochastic component to generalize specific signals.

The most simple AR-model is given by the recursive generation:

$$x_i = ax_{i-1} + \epsilon_i$$

of the sequence $\{x_i\}, i = 0, 1, \dots$. The parameter a influences the general appearance of the process. The sequence $\{\epsilon_i\}$ of random values defines the actual form of the sequence. The distribution of ϵ may be characterized by its standard deviation σ_ϵ . Fig. 9 shows two samples of such AR-processes, one with $a = 0.95$ one with $a = 0.99$. Obviously the parameter a controls the appearance, larger a lead to longer dependencies, thus subjectively to a larger scale, thus to higher smoothness. However, the roughness of the signals does not allow to talk about a typical wavelength, i.e. representative value for the scale.

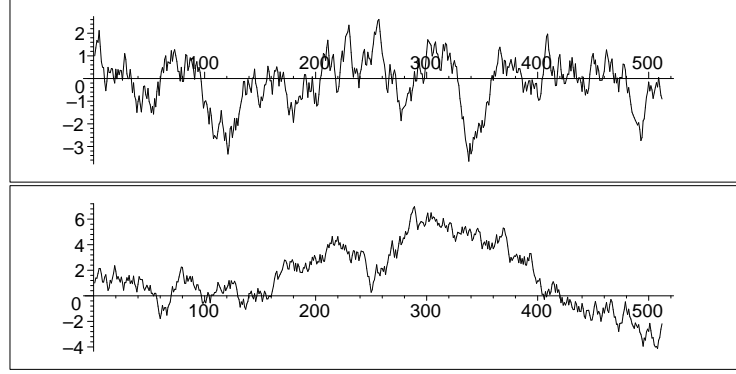


Fig. 9. Top: AR-process with $a = 0.95$. Bottom: AR-process with $a = 0.99$. In both cases $\sigma_\epsilon = 0.25$

Signals with a predominant wavelength can be achieved with higher order AR-process:

$$x_i = \sum_{k=1}^K a_k x_{i-k} + \epsilon_i$$

where K is the order of the AR-process. Again the parameters a_k control the behavior.

For second order processes we are able to choose parameters a_1 and a_2 which lead to signals with one predominant scale, i.e. wavelength. In the top of the three figures 5, 6, shown above, and 10 we observe three signals with typical wavelengths 40, 160 and 40 resp.

Obviously the first two profiles differ in their dominance of a specific wavelength, the last profile is more noisy than the first. The bottom of the three profiles shows the theoretical Fourier spectrum³. One can easily identify two peaks. Their distance is inversely proportional to the wavelength.

This shows, that it has been quite reasonable to identify the average wavelength $\bar{\lambda} = \frac{1}{\bar{u}}$ derived from the average frequency as natural scale of a signal. In the examples, shown so far, however, we are able to predict the scale from the parameters of the underlying process.

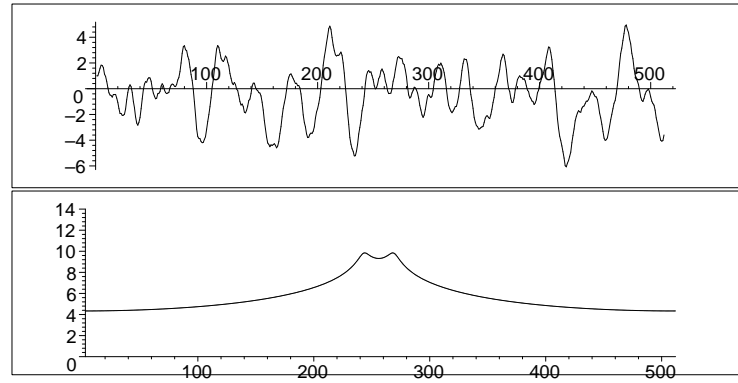


Fig. 10. Top: A rough AR-process with short wavelength. Bottom: Amplitude spectrum.

If we now have a look at the parameters a_1 and a_2 in the following table they do not immediately show us the structure of the signals. In this special case of a linear system, one however is able to tell the properties of the signal, namely its roughness and average periodicity.⁴

Profile	fig.	r	λ	a_1	a_2
1	5	0.99	40	1.9556229	-.98010000
2	6	0.99	160	1.9784734	-.98010000
3	10	0.95	40	1.8766078	-.90250000

We are also able to generate signals with *two* specific scales using higher order AR-processes. Figure 11 shows two signals each containing two different wavelengths, namely 10 and 40 (first two rows) and 6 and 80 (last two rows), together with their Fourier spectra. The two dominant wavelengths are clearly visible in the signals, and the Fourier spectra tell which wavelengths are present.

³ The theoretical Fourier or amplitude spectrum is only depending on the process parameters a and σ_ϵ and given by $\sigma_\epsilon / |1 - a \exp(j2\pi f)|$ (Lücker 1980). The empirical Fourier spectrum is given by the Fourier transform of the signal.

⁴ By looking at the roots of the polynomial $P(z) = 1 - a_1 z - a_2 z^2$.

Again the parameters a_1 , a_2 , a_3 and a_4 collected in the following table do not tell anything immediately about the structure. Linear systems theory allows to predict the behavior of the signal also in this case.⁵

Profile	fig.	r_1	λ_1	r_2	λ_2	a_1	a_2	a_3	a_4
1	11 top	0.99	40	0.99	10	3.5574765	-5.0928217	3.4866827	-.96059601
2	11 bot.	0.99	80	0.99	6	2.9684734	-3.9188887	2.9094008	-.96059601

The example intensionally is chosen such that there is *no unique natural* scale or average wavelength.

In this case theory is available to analyze the signal in detail and understand its behavior. This theory is restricted to one-dimensional signals generated by a linear system. The theory for two-dimensional signals already is quite difficult and cumbersome.

4.2 Scale in non-linear models

The second example wants to demonstrate the inability to predict the behavior of a nonlinear system, containing only one parameter.

The well known logistic map (Kneubühl 1995)

$$x_k = a x_{k-1}(1 - x_{k-1})$$

is fully deterministic. It shows all types of behavior, fixed points, periodic sequences as well as chaotic i.e. non predictable sequences for $a > 3.5699 \dots$ (cf. figure 12). Though this sequence is well studied it has the flavor of non linear systems which need to be studied individually and only in special cases allow prediction of their behavior without simulation studies. This is in contrast to one-dimensional linear systems, which are theoretically fully understood.

The examples of this section give a clear message: In case the system can be adequately modeled as a linear system, one should exploit the power of linear systems theory. This especially holds, if reality is modeled phenomenologically or only simple linear physical laws are involved. Even in case the approximation is crude, it helps making predictions. However, all real systems are likely to be nonlinear, which makes the theory based definition of a scale very difficult.

4.3 Omni-scale models

Real data ofte show no dominant scale, i.e. one dominant wavelength. They often show a continuum of wavelenths, i.e. the height $z(x)$ is composed of sinewaves of all frequencies:

$$z(x) = \sum_k h_k \sin(kx + p_k)$$

where the phases p_k and especially the amplitudes h_k depend on the frequency.

⁵ Again by analysing the roots of a polynomial $P(z) = 1 - \sum_k a_k z^k$.

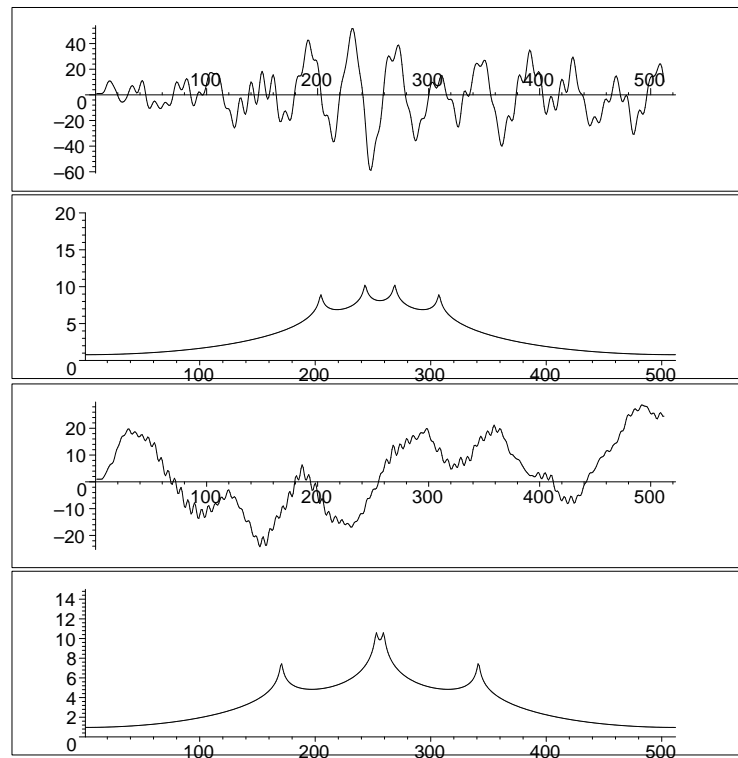


Fig. 11. Two profiles, each with two wavelengths and their theoretical Fourier spectrum

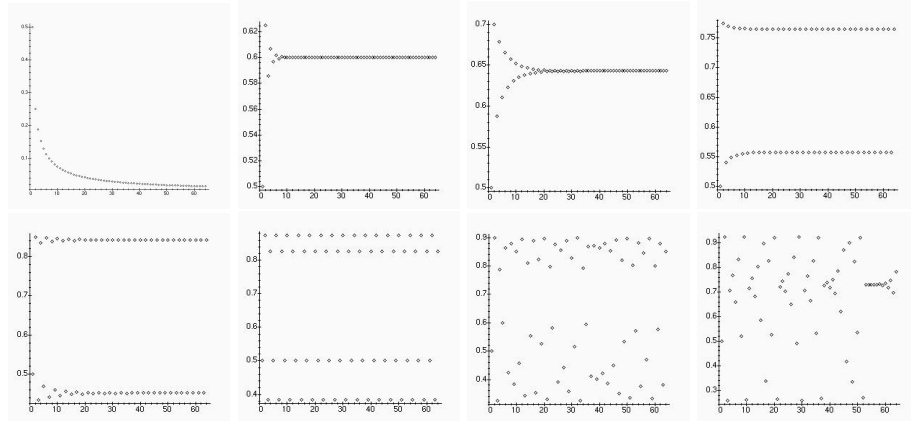


Fig. 12. Sequences of the logistic map with $a = 1, 2.5, 2.8, 3.1, 3.4, 3.5, 3.6$ and 3.7

Terrain data have been analysed by (Mandelbrot 1980). In case of terrain profiles (Frederiksen et al. 1985) often found a linear dependency

$$\log h_k = c - \alpha \log k$$

between the logarithmic amplitude h_k and the logarithmic frequency k . Thus often all frequencies are present in terrain profiles, there is no specific wavelength or scale in the data.

The slope α is large for smooth terrain, e.g. in the range of $\alpha \approx 2.5$, the slope is small for rough terrain, e.g. in the range $\alpha \approx 2$ or less.

5 Conclusions and Outlook

The paper discussed the various notions of scale resulting from the diversity of the disciplines within the geosciences. They refer to data and to models and appear in several degrees of crispness.

There seem to be basically two different crisp types of notions of scale in use:

1. scale as the ratio between model and reality. The only application of this notion is the scale in *map scale*.
2. scale as a measure of the extent of an object on space or time. This holds for nearly all notions, quantitative and qualitative.

The other notions qualitatively refer to the size of the data set in concern. There are quite some close relations between the crisp notions of scale, all referring in some way to linear systems theory. Their use definitely is promising and helps getting a first insight into geoprocesses.

In contrast to this coherence, however, there are only few crisp relations between the more qualitative notions, which hinders a smooth communication.

Taking the complexity of geoprocesses into account one needs to discuss how far a crisp notion of scale actually carries research and at which steps within modeling more detailed concepts need to support the description of the inherent structure of natural processes. These concepts definitely refer to the used qualitative notions of scale in some way, but they certainly will be more adequate.

References

- BOX, G. E. P.; JENKINS, G. M. (1976): *Time Series Analysis*. Holden-Day, 1976.
- CASTLEMAN, K. R. (1996): *Digital Image Processing*. Prentice Hall Inc., 1996.
- FREDERIKSEN, P.; JACOBI, O.; KUBIK, K. (1985): *A Review of Current Trends in Terrain Modelling*. *ITC Journal*, 1985-2, pp. 101–106.
- KNEUBÜHL, F. (1995): *Lineare und nichtlineare Schwingungen und Wellen*. Teubner Studienbücher Physik. Teubner, 1995.
- LÜCKER, R. (1980): *Grundlagen Digitaler Filter*. Springer, 1980.
- MANDELBROT, B. (1982): *The Fractal Geometry of Nature*. Freeman and C., San Francisco, 1982.
- RYAN, T. W.; GRAY, R. T.; HUNT, B. R. (1980): Prediction of Correlation Errors in Stereopair Images. *Optical Engineering*, 1980.
- SAMET, H. (1989): *The design and Analysis of Spatial Data Structures*. Addison-Wesley, Reading, Mass., 1989.
- SIEHL, A. (1993): *Interaktive geometrische Modellierung geologischer Flächen und Körper*. *Die Geowissenschaften*, 11(10-11):342–346, 1993.
- SNYDER, J. P. (1987): *Map Projections – A Working Manual*. US Government Printing Office, Wash., 1987.
- TISSOT, A. (1881): *Mémoire sur la Représentation des Surfaces et les Projections des Cartes Géographiques*. Gauthier-Villars, Paris, 1881.