34

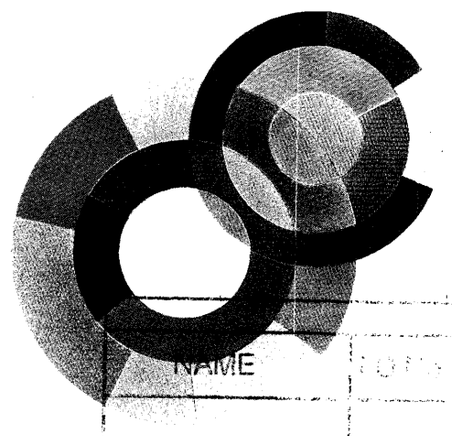
C.E. SACLAY
DSM

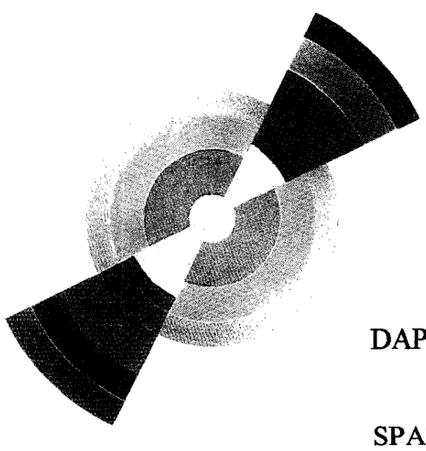
SERVICE D'ÉLECTRONIQUE ET D'INFORMATIQUE

DAPNIA-SEI-98-08

FERMILAB
FEB 12 1999
LIBRARY

Fermilab Library
0 1160 0068299 1





NAME

DAPNIA/SEI 98-08 December 1998

SPATIAL REPRESENTATION OF ECONOMIC AND FINANCIAL MEASURES USED IN AGRICULTURE VIA WAVELET ANALYSIS

M. Morehart, F. Murtagh, J.L. Starck

DAPNIA

Submitted to International Journal of Geographical Information Science

Couverture : Sylvie Cabrit et Marc Sauvage

Le DAPNIA (**D**épartement d'**A**strophysique, de physique des **P**articules, de physique **N**ucléaire et de l'**I**nstrumentation **A**ssociée) regroupe les activités du Service d'Astrophysique (SAp), du Département de Physique des Particules Elémentaires (DPhPE) et du Département de Physique Nucléaire (DPhN).

Adresse : DAPNIA, Bâtiment 141
CEA Saclay
F - 91191 Gif-sur-Yvette Cedex

Spatial Representation of Economic and Financial Measures Used in Agriculture Via Wavelet Analysis

December 5, 1998

Mitchell Morehart[1], Fionn Murtagh[2] * and Jean-Luc Starck[3]

[1]Economic Research Service, USDA, 1800 M Street NW, Washington DC 20036, USA

Tel. +1 202 694-5581, Fax +1 202 694-5758, Email morehart@mailbox.econ.ag.gov

[2]Faculty of Informatics, University of Ulster, Northern Ireland BT48 7JL, UK

Tel. +44 1504 375453, Fax +44 1504 375489, Email fd.murtagh@ulst.ac.uk

From end January 1999: Department of Computer Science, The Queen's University, Belfast BT7 1NN, Northern Ireland

[3]CEA/DSM/DAPNIA, 91191 Gif-sur-Yvette cedex, France

Tel. +33 1 69.08.57.64, Fax +33 1 69.98.65.77, Email jstarck@cea.fr

Abstract

A foundation is set forth for use of the wavelet transform as a spatial analysis tool for modelling the geographic representation of economic and financial measures used in agriculture. This provides a framework from which to estimate a smooth nonparametric function which

*Author for correspondence. This results of this analysis and any conclusions herein represent the views of the authors, and do not reflect those of the USDA

describes complex, multivariate relationships embedded in spatial data, with the resulting maps conveying large amounts of information in a familiar format. We illustrate this approach for tasks which include the graphical presentation of information, density estimation and wavelet-based nonparametric regression. A redundant wavelet transform is used, and we detail the properties which make it particularly appropriate for these objectives.

Keywords: nonparametric regression, wavelets, multiresolution, spatial analysis

1 Introduction

This paper explores the use of the wavelet transform as a spatial analysis tool for modelling the geographic representation of economic and financial measures used in agriculture. The principles of wavelet theory go back a number of decades (Haar developed what is now known as the Haar transform in the early years of this century, and Gabor amended the Fourier transform in the 1940s, applying it locally instead of globally). Numerous articles and books extending both theoretical aspects of wavelets and potential applications have been published since the late 1980s. This rapid pace of advancement is particularly reflected in signal processing, image processing, and in the fields of mathematics and statistics. Wavelet analysis has been shown to provide competitive and in some cases superior nonparametric density estimation when compared with traditional methods (Donoho *et al.* 1995). The appeal of wavelet methods stems from their ability to process noisy data with local structures and represent discontinuities such as jumps or peaks in a function. This is an important consideration for estimating geographically-referenced multivariate surfaces based upon economic or financial data where a high degree of spatial inhomogeneity is expected.

Two points are of particular importance here. First, that in contrast to other types of GIS studies which use physical data that have a more direct connection, the focus here is on economic or financial data which is noisy and has no direct expectations regarding geographic dependence. Second, many of the alternative statistical procedures can not handle data problems of significant size and

complexity nor do they adequately deal with the problems of identifying local structures and peaks and jumps in the data at the same time. The further benefits of this procedure are illuminated in the presentation and discussion of the results.

The next two sections present a brief overview of wavelet analysis with emphasis on application to nonparametric function estimation. This is followed by a discussion of the data source and application of a geographic information system. The fifth section provides an empirical framework for analysis and the underlying methodology. In the sixth section, spatial distribution problems for two agricultural financial indicators are used to illustrate nonparametric regression smoothing using wavelet methods. The first involves identifying specific areas of the U.S. where farms are most dependent on wheat as a source of income. The second problem is to determine the location of farm businesses experiencing the most extreme debt repayment problems. Concluding remarks focus on the potential of these techniques for public policy decision making and areas for further research and application.

2 Wavelets and Wavelet Regression

Wavelets are irregularly shaped functions that are specified to satisfy certain mathematical conditions. Wavelet analysis can be viewed as an alternative to Fourier analysis. The latter describes a function in terms of its frequency components. A wavelet basis instead describes the function in terms of a limited set of frequencies, and simultaneously in terms of a scale component which can be space-related (for spatial data) or time-related (for time series). Wavelet bases are therefore quite desirable in that they allow for localization in space and frequency. They are considered as being very useful analysis tools for investigation of localized phenomena, i.e. phenomena which are limited in extent relative to space or to time.

The wavelet transform involves initial choice of a wavelet function. Such a function satisfies different degrees of smoothness, compact support, and in some cases orthogonality. Perhaps most important

is their ability to provide very useful resolution views for data manifesting an extremely wide range of structures (Hall and Patil 1996). A more rigorous theoretical development of wavelet analysis is reported elsewhere (see for example, Daubechies 1988, Mallat 1989, and Chui 1992). Wavelets have found their way into a variety of applied fields including astronomy, music, image processing, nuclear engineering, signal analysis, magnetic resonance imaging, speech discrimination, earthquake-prediction, and many others. For purposes of this study, the discussion of wavelets will be confined to the statistical problem of nonparametric function estimation.

The use of wavelets in statistics is a relatively recent phenomenon that is rapidly developing. Several analysts have advanced wavelet theory in the context of nonparametric function estimation (Donoho 1993, Donoho and Johnstone 1992, Walter 1992, DeVore and Lucier 1992, Kerkyacharian and Picard 1992, and Kerkyacharian and Picard 1993). They have offered alternative approaches to traditional smoothing methods based on selective wavelet reconstruction or wavelet shrinkage. The idea is as follows. By design, a wavelet transform is invertible, i.e. from the wavelet transform we can exactly reconstruct our input data. Therefore we can pass from so-called direct space to transform space and back again. Carrying out some modification to the data in wavelet space is often found to work very well. By this we mean retaining wavelet levels such that more high-frequency components in the data are deleted. Alternatively, an effective denoising strategy is to assess all wavelet coefficients for their informativeness, – whether they are really signal or noise. Thus we see that we are recombining the data, based on modified wavelet coefficients. We have thereby made use of a linear or a nonlinear combination of wavelet functions to represent the data.

We can advance two reasons why it is effective to work on tasks such as denoising or, in general filtering, in wavelet transform space. Firstly, the wavelet transform is often found to be relatively very sparse. Many wavelet coefficients are zero or are small in value. Secondly, a wavelet transform provides us with a multiresolution decomposition of our data. If we are thereby accessing basic scale-related components in our data, then clearly it is good to base our filtering on these components.

The statistical application of wavelets is similar to parametric regression in the sense that the

underlying function is written in terms of some building-block functions, with projections onto these building-block functions being determined from the data. These building block functions are represented by what are known as the mother and father wavelets. The mother wavelet is good at representing the detail components of a function, while the father wavelet (also known as the scaling function) is noted for its ability to represent the smooth aspects of a function.

The general wavelet regression procedure is as follows. We use three primary steps: (1) applying the Discrete Wavelet Transform (DWT) to the data, (2) removing noise from the estimated coefficients, and (3) conducting diagnostics and applying the Inverse Wavelet Transform (IWT) to the modified coefficients to obtain the denoised data.

Suppose we want to model a random variable Y as a function of the variable X. These variables could represent any agricultural or economic dependencies. Given n independent observations $\{(X_i, Y_i)\}_{i=1}^n$ the standard nonparametric regression model involving an unknown function m is represented as

$$Y_i = m(X_i) + \epsilon_i. \tag{1}$$

where the residuals ϵ_i are assumed independent and identically distributed ($i.i.d.$) with mean zero and variance σ^2. The goal is to estimate the unknown mean response function $m(x)$ for the data points X_i. Later, below, we will relax the $i.i.d.$ requirement. We will consider all i so that the vector $m(X) = \{m(X_i) \mid i\}$ and likewise the vector $\epsilon = \{\epsilon_i \mid i\}$.

The orthonormal DWT can be represented in matrix notation by

$$w = Wm(X) + W\epsilon, \tag{2}$$

where W represents an $n \times n$ orthogonal matrix of wavelet filter coefficients associated with the selected wavelet system. The result of this procedure is a set of wavelet coefficients w, some of which contain significant signal and others which are mostly noise. Given an appropriate W which defines

the orthonormal DWT, we can write w as follows.

$$w = \begin{pmatrix} c_J \\ d_J \\ d_{J-1} \\ \vdots \\ d_1 \end{pmatrix} \tag{3}$$

The coefficients c_J generally represent the underlying smooth behaviour of the data at the coarse scale 2^J. They are created by the father wavelet, referred to above. The detail coefficients d_J, due to the mother wavelet, correspond to deviations from smooth behaviour at the coarse scale, while $d_{J-1}, \ldots d_1$ represent successively finer corrections to the smooth trend.

Many alternative methods have been offered to select what we may justifiably term a regression smoothing parameter given by the index J (Walter 1994, Pinheiro and Vidakovic 1995, Vannucci and Vidakovic 1996, and Hall and Nason 1996). The choice of J is an important consideration since increasing J tends to decrease the amount of smoothing in a manner that is analogous to using a smaller bandwidth in kernel estimation.

Other regression smoothing parameters are also of importance, related to where and how we modify wavelet coefficient values. After wavelet coefficients have been determined, a procedure referred to as (soft or hard) *thresholding* can be applied to separate coefficients which contribute significant signal from those that represent noise. Obviously, the treatment of coefficients has a substantial impact on the relative smoothness and mean square error performance of the estimated function. The choice of threshold and specific thresholding policy have undergone considerable study in recent years. The various methods that have been offered can be grouped into two categories. The first, known as *global thresholding*, involves applying a single-value threshold to all empirical wavelet coefficients. The second group involves some type of *level-dependent thresholding* where a different threshold value can be chosen for each wavelet level (Donoho and Johnstone 1994). There also are methods of

selecting level-dependent thresholds which can be described as data dependent. These alternative procedures are summarized by Nason 1995 with emphasis on cross-validation, while a host of others are summarized in Ogden 1997. After thresholding is completed, the estimated regression curve $\hat{m}(x)$ is obtained through the inverse transform of remaining or adjusted wavelet coefficients.

3 A Wavelet Transform for Geographic Data Analysis

Wavelets are inherently tied to the concept of multiresolution analysis. Using wavelets, a function can be depicted as a coarse overall shape, plus details that range from broad to narrow. This ability to zoom in or zoom out that is characteristic of multiresolution analysis permits an image to be interpreted as a sum of details which appear at different resolutions. Furthermore, each scale of resolution may pick up different types of structure in the image.

The à trous DWT (Shensa 1992, Holschneider 1989, Starck, Murtagh, and Bijaoui 1998) differs from the DWT methods considered so far in that *decimation* or subsampling is not carried out at successive resolution levels, in spite of the fact that the information carried by these resolution levels decreases. We have not dealt on this aspect of the orthonormal DWT, namely that it is usual practice to subsample coefficient values, taking one value from every two, at successive resolution levels. This procedure is automatically catered for, when the matrix W used in the previous section is appropriately defined. In the à trous DWT, the usual practice is for detail coefficients to remain as numerous as the input data, at the successive resolution levels. This has great benefits for cartographic representation, as we will see below. This sort of redundancy also aids enormously in feature detection and exploration in the succession of resolution views of the data, which are provided by the detail coefficients. For given data x, and using the notation used for equations (2) and (3), the à trous DWT gives us the following additive decomposition:

$$x = m + \epsilon = d_1 + d_2 + \ldots + d_J + c_J \tag{4}$$

Each term here is a vector of length p, or an image of dimensions $n \times p$, or even a data cube. The sum is simply carried out element-wise, or pixel-wise. The additive decomposition is one where the components are the succession of detail coefficients which are dependent on the resolution level, plus the underlying smooth version of the data which is given by c_J.

Dilated (i.e., expanded in scale, but keeping the same overall shape; furthermore the scale expansion is usually, in practice, a two-fold one) versions of a wavelet function are used to produce d_1, d_2, etc. A *scaling function* is used to produce c_J. These are, respectively, the mother and father wavelet function referred to earlier. The scaling function has the effect of smoothing the data. These functions are the mother and father wavelet functions referred to earlier. The à trous DWT has the nice property that dilated versions of a smoothing function are used to produce a set of successively smoothed versions of the input data, $c_1, c_2, \ldots, c_J$ and then the wavelet coefficients can be defined straight away from these just by subtraction: $d_1 = x - c_1, d_2 = c_1 - c_2, \ldots, d_J = c_{J-1} - c_J$. The scaling function used in this work is like a Gaussian function, but with limited support (i.e. no wings extending to infinity). It is in fact a B_3 spline (Strang and Nguyen 1996), known for its good interpolation properties which extend to irregularly sampled data. (In passing we note that the data which we are studying are irregularly sampled: the discrete convolution with the scaling function – whose shape and support determine the scale-related resolution – provide mappings of our data onto a regular grid.)

Data dependent noise modeling may be based on the noise model posed for the input data, such as (in the additive noise model case) $x = \{X_i + \epsilon_i\}_{i=1}^n$ where ϵ need not be *i.i.d.* Based on such a noise model, significance levels can be determined at successive resolution levels using the classical hypothesis testing approach. In figures 7 and 8 below, we assume a simple (and approximate) stationary Gaussian model for ϵ. For function estimation, or wavelet regression, in figures 9 and 10, we relax entirely this assumption of homoscedasticity. We take the noise as not necessarily stationary, and of unknown distribution, modeling it as locally Gaussian, only, independently at the varying resolution levels.

The MR/1 multiresolution image and data analysis package (MR/1 1998, Starck, Murtagh, and Bijaoui 1998) was used in this work. This package handles (raster) image and point pattern data. An extensive range of wavelet and other multiscale transforms are supported (à trous, Mallat, Feauveau, Haar, and others, in various pyramidal and non-pyramidal versions; Laplacian, median and morphological transforms are also available). Data dependent noise models include Gaussian and Poisson, additive and multiplicative, stationary and non-stationary. Supported are such tasks as visualization, filtering, deconvolution, compression, feature and structure finding, and many other operations.

4 Data Source

Analysis of the spatial distribution of economic and financial indicators for agriculture is based on financial statements prepared from USDA's (United States Department of Agriculture) Agricultural Resource Management Study (ARMS). These data are the only information gathered nationally which provide a comprehensive perspective on the financial performance of farm businesses. The ARMS is a personally enumerated survey, conducted since 1984 by the National Agricultural Statistics Service (NASS) and the Economic Research Service (ERS) of the U.S. Department of Agriculture. The ARMS is a probability-based multi frame, stratified survey that uses multiple questionnaire versions to collect information on farm production expenses, capital purchases, income, production practices, and other farm operating characteristics.

Local geographic references for sample points are often omitted from surveys designed to collect economic and financial information. In this instance, some form of data integration is required to associate these attributes with geographic references that are more specific than the usual political boundaries such as State or County. The simplest form of data integration involves attaching geographic references to sample points from survey data. The most common geographic reference is the longitude/latitude coordinate system. For this study, geographic references were obtained for

each of the 3,100 county population centroids. The projection applied was the Albers-Equal-Area (AEA).

An additional component of data preparation for this study involved binning. Binning methods originated with development of the histogram as a means to graphically display univariate data. More recently, binning has been used to reduce computational difficulties associated with applying nonparametric regression techniques for smoothing large data sets (Scott 1992, Härdle and Scott 1992, and Wand and Jones 1994). In this context, binning involves reducing the original data to a relatively small number of estimates at equally spaced intervals. This treatment of sample data has been used to support various kernel-type estimators such as the Average Shifted Histogram (Scott and Whittaker 1996), bivariate local linear regression (Werthenbach and Herrmann 1998), and Fast Fourier Transform (FFT) computation methods (Fan and Marron 1994). Donoho *et al.* 1995 provide examples for density estimation based on wavelet thresholding, while Antoniadis and Pham 1995 incorporated binning directly into a nonparametric wavelet regression estimator.

In addition to computational efficiencies, binning offers other important benefits to practical problems associated with spatial modeling of complex survey data. First, the binning algorithm developed by Scott allows the complex, probability-based sample characteristics to be incorporated in the analysis (Scott and Whittaker 1996). The second advantage of this approach is that the selection of bin width or grid spacing allows the practitioner some flexibility in controlling the degree of resolution given the statistical reliability constraints of the sample data. Results by Hall and Wand 1996 for kernel density estimation suggested that the accuracy of binning and the choice of grid size are dependent on the relative smoothness of the underlying density and sample size. Finally, as noted by Antoniadis and Pham 1995, binning can be used to compensate for several restrictions that pertain to widely used wavelet estimators (section 2) such as fixed equidistant design, homoscedasticity, and sample sizes that are a power of two.

In the context of the à trous DWT introduced in section 3, binning arises naturally as the discrete convolution of a scaling function with the given input data. We use, below, a heteroscedastic noise

model with this wavelet transform.

5 Model Specification and Methodology

Having established a spatial data structure from which to examine multivariate relationships, the next step involves implementation of the wavelet transform in a two-dimensional setting. The major goal of this process is to produce a sequence of versions of the binned approximation of the original data at more and more coarse resolutions separating noise from salient features in the data.

Extension of the nonparametric regression principles for wavelet analysis to two dimensions is based on estimating the unknown mean response function:

$$\hat{m}(x, y) = E(Z|X = x, Y = y), \tag{5}$$

where (x, y) represents the centre of one of the geographically referenced bins and Z represents the variable of interest. The preconditioned $\tilde{Z}$ matrix is constructed by calculating the weighted average value within each of the bins as

$$\hat{z}_{j,l1,l2} = \frac{\sum_{i=1}^{n} w_i Z_i}{\sum_{i=1}^{n} wi}, \tag{6}$$

where n is the number of Z_i's in $B_{j,l1,l2}$ and w_i represents the corresponding sampling weights for each observation. The indexes $j, l1, l2$ used here serve to remind us of the weighting relative to the variable, which are referred to as the sampling weights in a statistical survey, and in addition the mapping into the discrete bin indexed by $l1, l2$. In many cases, the estimate will be mathematically undefined or zero. The first situation occurs when there are no sample observations in the bin, while the latter reflects no positive response for the variable of interest.

The redundant à trous method has a range of useful properties for the handling of sparse spatial data. Compared to other wavelet transforms, we may note the following:

Boundary treatment The handling of data boundaries in the DWT always requires attention. Implementations of DWT methods which use decimation usually require that the input data dimensions be integer powers of two; or else that the input data be padded to be of such pre-specified dimensions. In the case of the redundant à trous method, we do not have such a constraint. In practice, we usually use reflection in the boundary (the "virtual" value at position $n+1$ is taken as the known value at n, position $n+2$ is given by $n-1$, and so on), which makes the convolutions with the scaling function well-defined at all stages.

Stationary transform The redundancy involved in the succession of applications of the scaling function, yielding $c_1, c_2, \ldots$, means that we ought not take each translated value into account when constructing the following resolution level: in fact, a sampling scheme based on gaps of $\{2^i\}_{i=0}^{J-1}$ unused values between values used in the convolution does the job admirably. Such a sampling scheme is compatible with what would have been the case, had we decimated. (Note that decimation can be carried out with the à trous wavelet transform, if so desired.) This works well all the more so since we are using a scaling function with excellent interpolation properties. It is for this reason that the à trous ("with holes" or gaps) method bears this name.

Computational efficiency The à trous DWT is of linear computational complexity, i.e. $O(n)$ for n input values (and is thus more efficient than, e.g. the FFT).

Due to redundancy, the à trous transform provides intuitively understandable visualization. The noise, or other, filtering of *sub-sampled* data can be entirely avoided. And as noted in section 3, we have at our disposal a wide range of powerful data dependent noise models.

6 Results

Even though wheat is produced in nearly every state, production is geographically concentrated. In 1995, Kansas, Montana, North Dakota, and Washington accounted for 43 percent of total U.S. production (NASS, Crop Production 1996). Perhaps a logical extension of this information would

be to assert that these are areas of the country where farm business and farm household incomes are most affected by changes in the price of wheat. As a test of this supposition, the first application explores the spatial distribution of the economic dependence on wheat. Table 1 shows the estimated average income statement for all farms that produced wheat in 1995 and for three categories of wheat dependence. As a group, wheat producers were not very dependent on wheat. Wheat was the primary commodity for one in five wheat producers and represented at least 60 percent of the total value of production. The income of these farms would be extremely sensitive to changes in the price of wheat.

Insert Table 1 about here.

Debt repayment capacity utilization is computed as the ratio of actual debt to maximum debt repayment capacity, and measures the extent of farm businesses or operators' use of their potential credit repayment ability. Debt repayment capacity, measuring the amount of debt that the maximum loan payment could support, is a function of the loan payment, the interest rate, and the term of the loan. Construction of the debt capacity utilization ratio (DCUR) assumed an average interest rate of 7.5% and a 10-year loan term.

The DCUR provides a framework to identify farm businesses that are experiencing financial difficulties based on the inability of current income to support farm business debt. In 1995, it was estimated that 14 percent of all commercial farms faced extreme repayment problems having debt that exceeded their feasible debt by more than 2.5 times (table 2). These farms had average debt of $309,000, but their current income could only support payments associated with debt of $49,000. A logical extension of estimating and summarizing this measure is to determine where the most extreme debt repayment problems occurred in 1995.

Insert Table 2 about here.

One of the unique features of wavelet analysis is that no matter what the function of interest a number of alternative views that represent detail in the data are presented graphically. Visualization of the wavelet transform provides knowledge about the structure of the data and presents a convenient forum for conducting wavelet-based diagnostics prior to reconstruction. The multiresolution aspect of the transform also permits the detection and parsing of objects within an image.

Figures 1 and 2 show perspective plots of the wheat dependence and debt data as spatially mapped over the surface of the continental United States, from a due south direction. These provide wheat and debt "topographies" at varying resolution scale. Our original data, in both cases, is the sum of the 4 scales shown. The last scale is the most smoothed version retained of the data. The detail or wavelet coefficients are shown in figures 3 and 4, with positive values only being represented in these figures.

Insert Figures 1, 2, 3, 4 about here.

Regression smoothing, or nonparametric function estimation, is necessary for interpretation of the spatially represented data under consideration here. A family of smoothing methods avoids undue dependence on just one estimate (Marron and Chung 1998). Such a family of smooths provides not just a set of alternative smoothing possibilities, but the transition from one smooth to the next may be important for change point detection and picking out scale-related features in the data. A condition for this is that the analyst can "visually connect" (Marron and Chung 1998) successive members of the family. This possibility immediately follows from the à trous wavelet transform. Figures 5 and 6 show such a family of fits to the wheat and debt functions under consideration.

Insert Figures 5, 6 about here.

Simple visualizations of the data based on a stationary Gaussian model are portrayed in figures 7 and 8. There are various ways to analyze empirical wavelet coefficients that enhance statistical

applications such as nonparametric estimation. Graphic displays of wavelet coefficients provide good diagnostics for data analysis since the data or function of interest is completely captured in the wavelet coefficients. Time-scale (*scaleogram*) and time-frequency plots (*spectogram*) are useful for studying the tradeoff between time and scale localization (Bruce and Gao 1996). Image querying, multiresolution editing, and multiresolution plots provide alternative means of visually exploring complex relationships in the data. Other standard methods of graphical data analysis such as box plots and Q-Q plots can also be applied to wavelet coefficients. We now turn our attention to more detailed noise modeling, aiming at a definitive regression surface of our data.

Insert Figures 7, 8 about here.

Figures 9 and 10 show the results of wavelet regression based on a model of heteroscedastic noise, of the wheat dependence and DCUR data. We empirically investigated a range of noise models for this data. Our initial considerations were of a Poisson distribution, given especially the sparsity of the data. However the complex preprocessing of this data argues strongly against this model. Wavelet coefficients are necessarily of zero mean, and we therefore used histograms at different scales to check whether a Gaussian distribution looked reasonable, which did not prove to be the case. This led us to base our noise filtering on a local, scale-dependent noise model. Denoising of the last smooth version of the data was not carried out, since this provides a background continuum. Some "spottiness" in the resulting filtered view of the data was removed with minimal consequences by removing isolated pixel values in the detected signal versus noise support image, built up at each resolution level. Only significant positive structure was detected. The last-mentioned property of the processing has an effect on the values of the noise-filtered data. The wheat dependence original and denoised data had maximum, minimum and totalled values of, respectively, 100.4, 0.0, 14775.7, and 99.0, 0.0, 20124.0 The DCUR original and denoised data had maximum, minimum and totalled values of, respectively, 12.0, 0.0, 2661.5, and 11.8, 0.0, 3597.6.[1]

[1] The increase in totalled values of the data was due to removing negative values in the estimation process.

Insert Figures 9, 10 about here.

It is easy to see that economic dependence on wheat is invariant to state boundaries. The dependent variable is defined as the share of the total value of production represented by wheat. The peaks of the estimated regression surface, which indicate where farm business and farm household incomes are most affected by changes in the price of wheat, are most notable in Kansas and North Dakota Even in these states, however, economic dependence on wheat is confined to specific areas within the state. In contrast, there were relatively small portions of Montana and Washington estimated to contain farms that were economically dependent on wheat. Areas of high dependence were identified in other states which contributed much less to total U.S. production, such as Oklahoma, Colorado, Nebraska, Oregon, and Arkansas.

The nonparametric regression surface for the DCUR data generated using wavelet analysis is presented in figure 10. Although extreme debt repayment problems were found in many different locations, many of the estimated regression surface peaks appear to be concentrated in the area from northern Texas through southern Nebraska. The location of farms experiencing debt repayment problems in 1995 reflects either a short-term cash shortfall, which could be associated with adverse weather conditions, or a longer-term breakdown in financial management of the business. Most farms with short-term debt repayment problems can negotiate with their lender to restructure current loan terms or borrow additional funds, sell inventories or other liquid assets, or utilize off-farm income to mitigate the severity of the problem. The latter group, whose situation is further constrained by a relatively high debt load, faces substantial financial and business restructuring issues.

The approach used can handle spatial correlation very well. One of the data sets analyzed, the wheat dependence data, shows strong spatial correlation, while the other, the debt repayment data, is spatially dispersed. The approach we used works well on the two kinds of data, while many alternative methods are only appropriate for the former, and may actually be misleading in the

latter case.

7 Discussion

A foundation for the application of wavelets was put forth for the statistical problem of nonparametric function estimation. This provided a framework from which to estimate a smooth nonparametric function which describes complex multivariate relationships embedded in spatial data. The exploratory data analysis inherent in the wavelet transformation and resulting maps provide a means of visually conveying tremendous amounts of information. This portrayal of the complex relationship between economic attributes and demographic characteristics could make an important contribution to understanding the spatial implications of government policy.

The spatial analysis developed and demonstrated in this paper could be augmented by further integration of demographic characteristics and resource-based attributes. These types of applications might include: (1) identifying the proximity of farm businesses that are dependent on off-farm income to areas of rural and urban non-agricultural employment, (2) determining the extent to which highly leveraged farms are located in disaster prone areas such as flood plains, and (3) examining how these geographic dependencies change over time and the extent to which weather and other natural occurrences influence these changes. Implementation would involve either extending the nonparametric model to accommodate higher dimensions or a more complex post-processing activity focused on merging geographic representations.

In comparing density estimation methods (an adaptive kernel density estimator and the maximum penalized likelihood method) and the à trous wavelet transform using extensive one-dimensional simulations, Fadda et al. 1998 find that the wavelet transform performs well when there are local gaps in the data distribution, and when small substructures are superimposed on larger structures. (They also argue for the robustness of kernel methods.) To this we would add the wide range of possibilities for exploratory visualization, exemplified in this paper, when using the wavelet transform. Finally,

we would stress the importance of a multiscale transform approach whenever our data is taken as a set of superimposed or compounded resolution scale-related phenomena.

In regard to the applications used in this work, we have discussed above in this section how these types of results can be combined with other data which may or may not have some type of geographic reference. The two examples do not have any relationship to each other; they were chosen to illustrate the methods and address current policy topics in U.S. agriculture. The economic dependence on wheat example is a means to provide a quick, visual response to the question if wheat prices dramatically decline (as they did during the latter part of 1998) what parts of the country would be most impacted? One could further utilize this surface to draw out other interesting relationships. Have the areas which have been identified as most sensitive to changes in the price of wheat been subject to unusual weather, or do they have access to off-farm job opportunities? These are logical extensions that are beyond the scope of this paper. In the same vein, the DRCU analysis can be extended to see what the conditions of banks are in those areas of the country where farms are having debt repayment difficulties (a study which one of the authors in fact carried out in the past).

Clearly the wavelet transform, and more particularly the methodology employed here (the redundant and hence shift-invariant transform, the noise modelling and denoising approach), is of very considerable benefit for analyzing such data and for handling such problems.

References

Antoniadis, A., and Pham, D.T., 1995, Wavelet regression for random or irregular design. Technical Report, University of Grenoble.

Bruce, A. and Gao, H., 1996, *Applied Wavelet Analysis with S-Plus* (New York: Springer).

Chui, C. K., 1992, *An Introduction to Wavelets* (New York: Academic Press).

Daubechies, I., 1988, Orthonormal bases of compactly supported wavelets. *Communications in Pure and Applied Mathematics*, **27**, 1271–1283.

DeVore, R. A., and Lucier, B. J., 1992, Fast wavelet techniques for near optimal processing. In *Proceedings of the IEEE Military Communications Conference*, 48.3.1–48.3.7.

Donoho, D. L., 1993, Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In *Proceedings of Symposia in Applied Mathematics*, **47**, 173–205.

Donoho, D. L., and Johnstone, I. M., 1992, Nonlinear solution for linear inverse problems by wavelet-vaguelet decomposition. Technical Report 403, Department of Statistics, Stanford University, Stanford, California.

Donoho, D. L., Johnstone, I. M., 1994, Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81**, 425–455.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D., 1995, Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society, Series B*, **57**, 301–369.

Fadda, D., Slezak, E., and Bijaoui, A., 1998, Density estimation with non-parametric methods. *Astronomy and Astrophysics Supplement Series*, **127**, 335–352.

Fan, J., and Marron, S., 1994, Fast implementations of nonparametric curve estimators. *Journal of Computational and Graphical Statistics*, **3**, 35–56.

Hall, P., and Nason G., 1996, On choosing a non-integer resolution level when using wavelet methods. Preprint, Department of Mathematics, University of Bristol, UK.

Hall, P., and Patil, P., 1996, On the choice of smoothing parameter, threshold, and truncation in nonparametric regression by nonlinear wavelet methods. *Journal of the Royal Statistical Society, Series B*, **58**, 361–377.

Hall, P., and Wand, M., 1996, On the accuracy of binned kernel density estimators. *Journal of Multivariate Analysis*, **56**, 165–184.

Härdle, W., and Scott, D., 1992, Smoothing by weighted averaging of shifted points. *Computational Statistics*, **7**, 97–128.

Holschneider, M., Kronland-Martinet, R., Morlet, J., and Tchamitchian, Ph., 1989, A real-time algorithm for signal analysis with the help of the wavelet transform. In J.M. Combes, A. Grossman and Ph. Tchamitchian, Eds., *Wavelets: Time-Frequency Methods and Phase Space* (Berlin: Springer-Verlag), 286–297.

Kerkyacharian, G., and Piccard D., 1992, Density estimation in Besov spaces. *Statistics and Probability Letters*, **13**, 14–24.

Kerkyacharian, G., and Picard, D., 1993, Density estimation by kernel and wavelet methods: Optimality of Besov spaces. *Statistics and Probability Letters*, **18**, 327–336.

Mallat, S. G., 1989, A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674–693.

Marron, J.S., and Chung, S.S., 1997, Presentation of smoothers: the family approach. Preprint, Department of Statistics, University of North Carolina, Chapel Hill.

Multi Resolutions Ltd., 1998, *MR/1: The Multiresolution Analysis Software*, User Manual, Version 1.0, 100 pp. (http://www.visitweb.com/multires or email multires@compuserve.com)

Nason, G., 1995, Choice of the threshold parameter in wavelet function estimation. In *Wavelets and Statistics*, Eds Antoniadis A., and Oppenheim G. (New York: Springer-Verlag).

Ogden, R, T., 1997, *Essential Wavelets for Statistical Applications and Data Analysis* (Boston: Birkhäuser).

Pinheiro, A., and Vidakovic, B., 1995, Estimating the square root of a density via compactly supported wavelets. DP 95-14 ISDS, Duke University.

Scott, D. W., 1992, *Multivariate Density Estimation: Theory, Practice, and Visualization* (New York: Wiley).

Scott, D. W., and Whittaker, G., 1996, Multivariate applications of the ASH in regression. *Communications in Statistics A: Theory and Methods*, **25**, 2521–2530.

Shensa, M. J., 1992, Discrete Wavelet Transforms: Wedding the à trous and Mallat algorithms. *IEEE Transactions on Signal Processing*, **40**, 2464–2482.

Starck, J.-L., Murtagh, F., and Bijaoui, F., 1998, *Image Processing and Data Analysis: The Multiscale Approach* (Cambridge: Cambridge University Press).

Strang, G. and Nguyen, T., 1996, *Wavelets and Filter Banks* (Wellesley: Wellesley-Cambridge Press).

U.S. Department of Agriculture, 1996, Crop Production, 1996 Summary, CrPr 2-1(97). National Agricultural Statistics Service, Washington, D.C.

Vannucci, M., and Vidakovic, B., 1996, Preventing the Dirac disaster: wavelet based density estimation. DP 96 ISDS, Duke University.

Walter, G. G., 1992, Approximation of the delta function by wavelets. *Journal of Approximation Theory*, **71**, 329–343.

Walter, G. G., 1994, *Wavelets and Other Orthogonal Systems With Applications* (Boca Raton, Florida: CRC Press).

Wand, M. P., and Jones, M. C., 1994, *Kernel Smoothing* (London: Chapman & Hall).

Werthenbach, C., and Herrmann, E., 1998, A fast and stable updating algorithm for bivariate nonparametric curve estimation. *Journal of Computational and Graphical Statistics*, **7**, (1), 61–76.

Table 1. Abbreviated income statement for wheat farms, 1995

Item	Less than 20 percent	20 to 59 percent	60 percent or higher	All farms
Number of farms	124,974	73,573	52,021	250,568
Total value of production	214,176	116,887	77,076	157,146
Value of wheat production	14,471	40,414	63,247	32,215
Wheat production (bu.)	3,772	10,055	15,313	8,013
Gross cash income	203,634	123,789	81,206	154,772
Livestock sales	67,718	19,245	5,763	40,623
Crop sales (incl. net CCC loans)	109,399	79,053	56,558	89,518
Government payments	7,886	7,716	7,587	7,774
Other farm-related income	18,631	17,775	11,297	16,857
Less: Cash expenses	160,730	99,464	62,784	122,406
Variable	124,624	71,482	43,347	92,146
Fixed	36,106	27,982	19,437	30,260
Equals: Net cash farm income	42,904	24,325	18,422	32,366
Less:				
Depreciation	19,485	13,535	9,971	15,763
Labor, non-cash benefits	666	341	284	491
Plus:				
Value of inventory change	42	2,559	6,979	2,221
Nonmoney income	4,838	4,115	2,602	4,162
Equals: Net farm income	27,633	17,124	17,748	22,495

Source: Agricultural Resource Management Study, USDA.

Table 2. Farm business debt capacity utilization, 1995

	Debt capacity utilization ratio			
Item	Below1.0	1.0 to 1.5	1.5 to 2.5	Above 2.5
Number of farms	389,663	37,278	36,095	73,204
Percent of farms	72.7	7.0	6.7	13.7
Net farm income	68,250	-1,363	-16,077	-70,490
Plus: Depreciation	21,054	23,601	18,130	23,883
Plus: Capital interest	5,533	17,294	15,126	16,785
Plus: Estimated capital lease	2,059	1,405	1,296	3,597
Equals: Income for debt coverage	96,896	40,937	18,475	-26,225
Maximum loan payment	77,517	32,750	14,779	0
Total reported debt	97,504	271,556	239,787	309,103
Maximum feasible debt	438,822	222,071	131,564	48,858
Debt capacity utilization ratio	0.22	1.22	1.82	6.33

Source: Agricultural Resource Management Study, USDA.

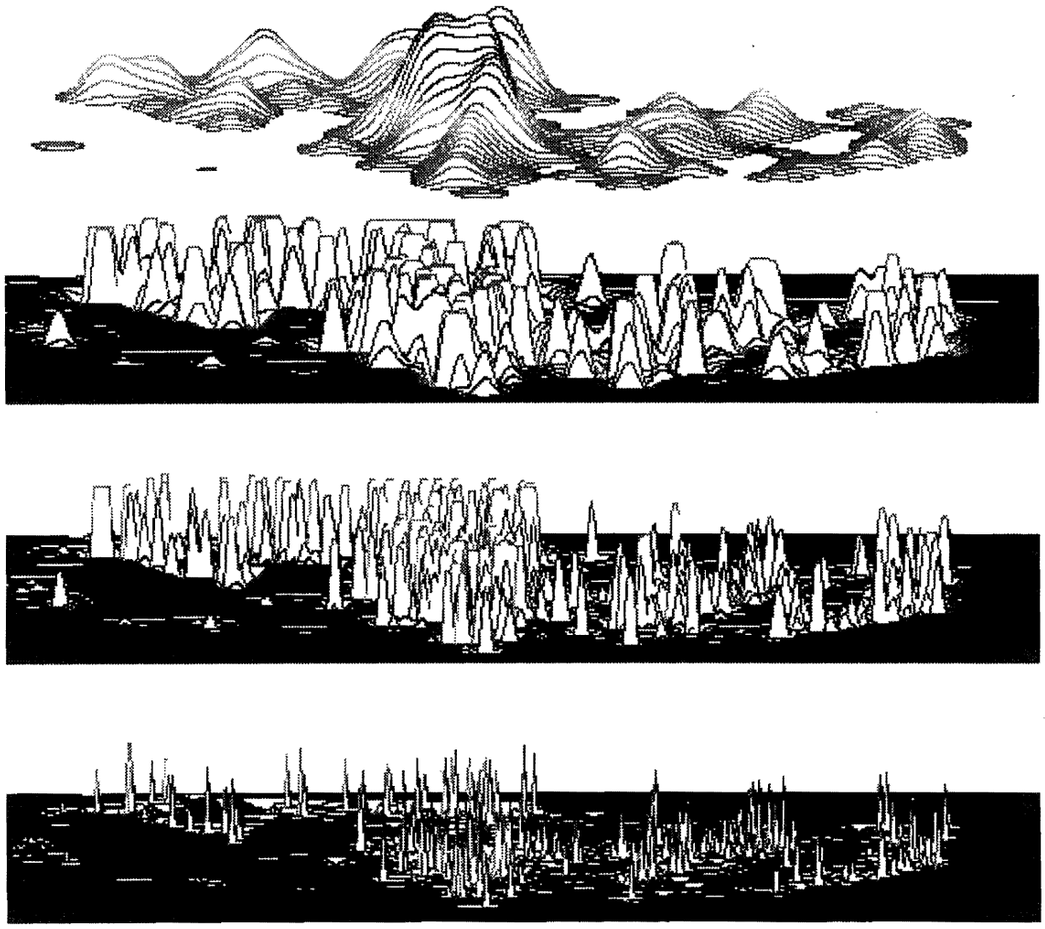


Figure 1: Perspective plot (from south) of 4-level à trous transform, ordered from bottom to top, of the 1995 wheat dependence data. The thresholded "capping" is at the 3σ level and is used for display purposes only.

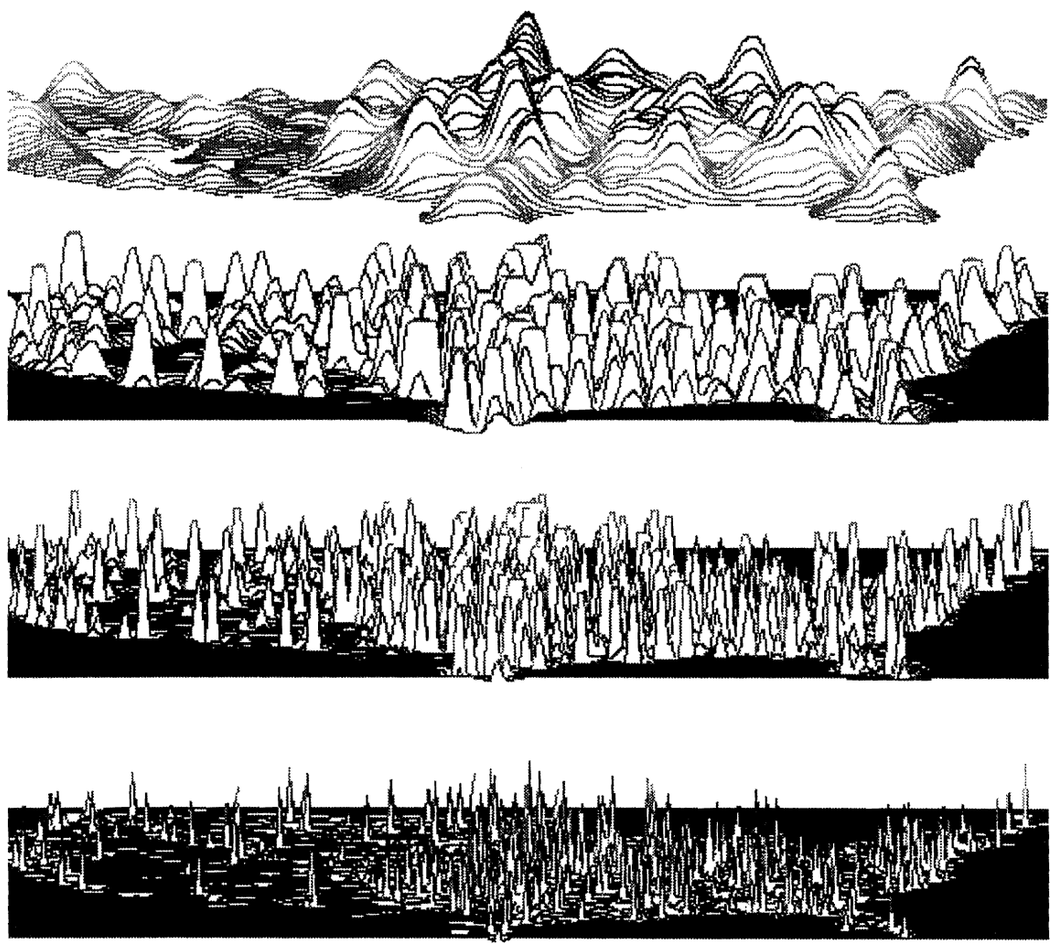


Figure 2: Perspective plot (from south) of 4-level à trous transform, ordered from bottom to top, of the 1995 debt capacity utilization ratio, DCUR, data. The thresholded "capping" is at the 3σ level and is used for display purposes only.

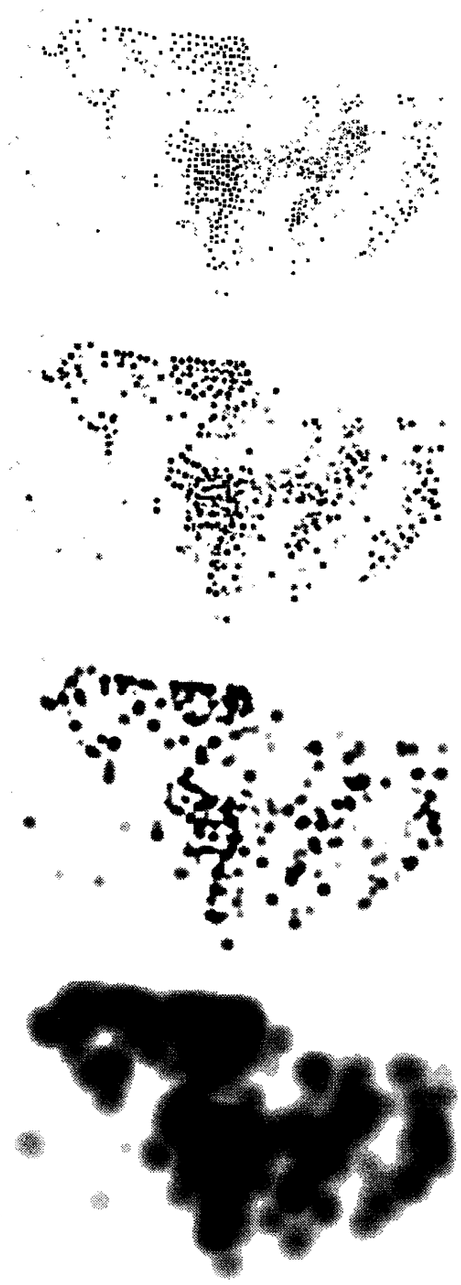


Figure 3: From top to bottom: Wheat dependency data; detail or wavelet coefficients (non-negative values shown) at successive resolution levels, d_1, d_2, d_3, c_3.

Figure 4: From top to bottom: Debt capacity utilization ratio, DCUR, data; detail or wavelet coefficients (non-negative values shown) at successive resolution levels, d_1, d_2, d_3, c_3.

Figure 5: From top to bottom: Wheat dependency data; approximations formed by $d_2 + d_3 + c_3$, $d_3 + c_3$ and c_3, where the original data has been additivitely decomposed into detail datasets d_1, d_2, d_3, together with a smooth background, c_3.

Figure 6: From top to bottom: Debt capacity utilization ratio, DCUR, data; approximations formed by $d_2 + d_3 + c_3$, $d_3 + c_3$ and c_3, where the original data has been additivitely decomposed into detail datasets d_1, d_2, d_3, together with a smooth background, c_3.

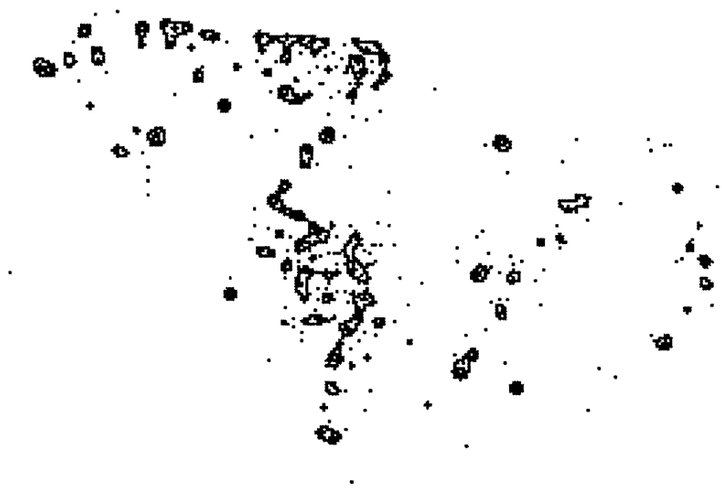


Figure 7: Contour plot of 4-level à trous transform of the wheat dependence data. Contours correspond to 3σ levels (stationary Gaussian distribution at each level).

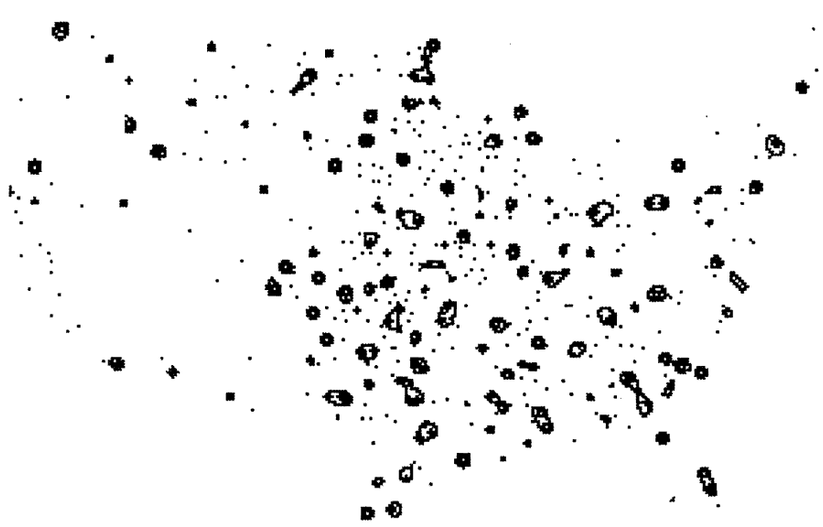


Figure 8: Contour plot of 4-level à trous transform of the debt capacity utilization ratio, DCUR, data. Contours correspond to 3σ levels (stationary Gaussian distribution at each level).

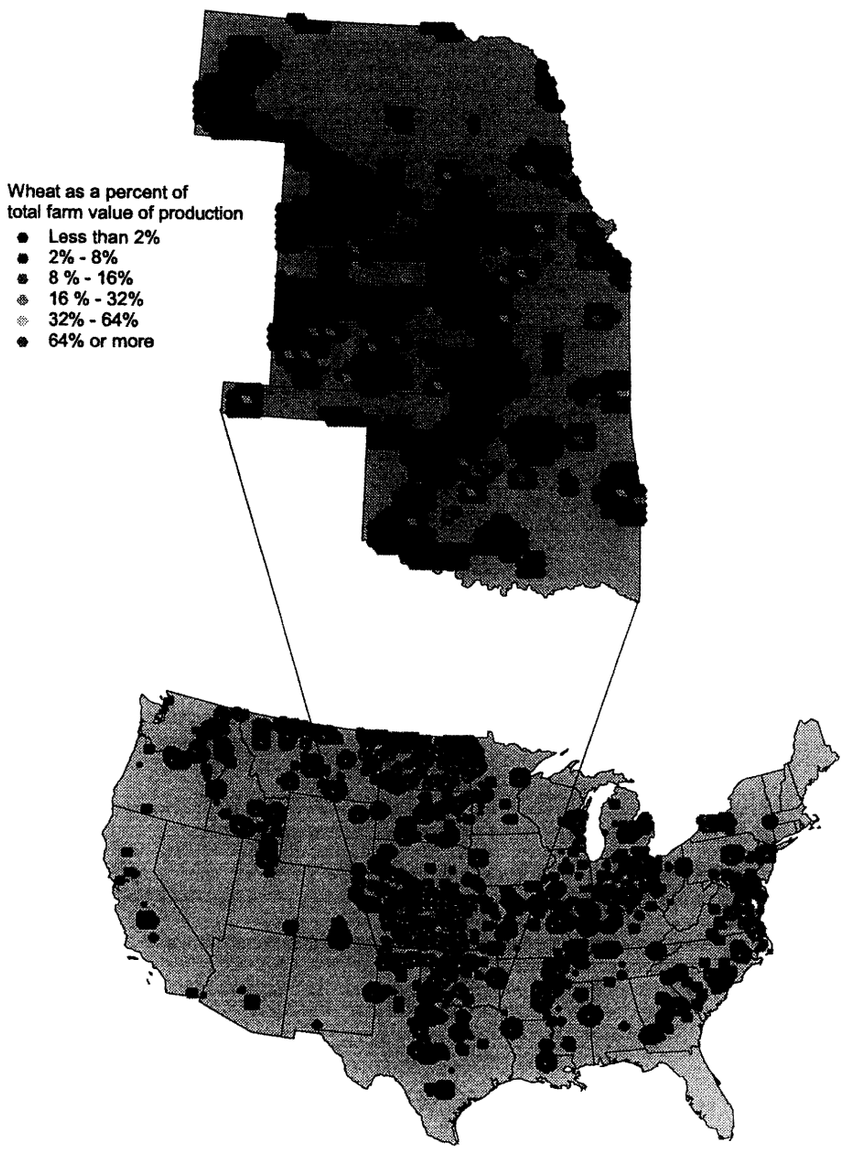


Figure 9: Result of wavelet regression (heteroscedastic noise model) of wheat dependence data.

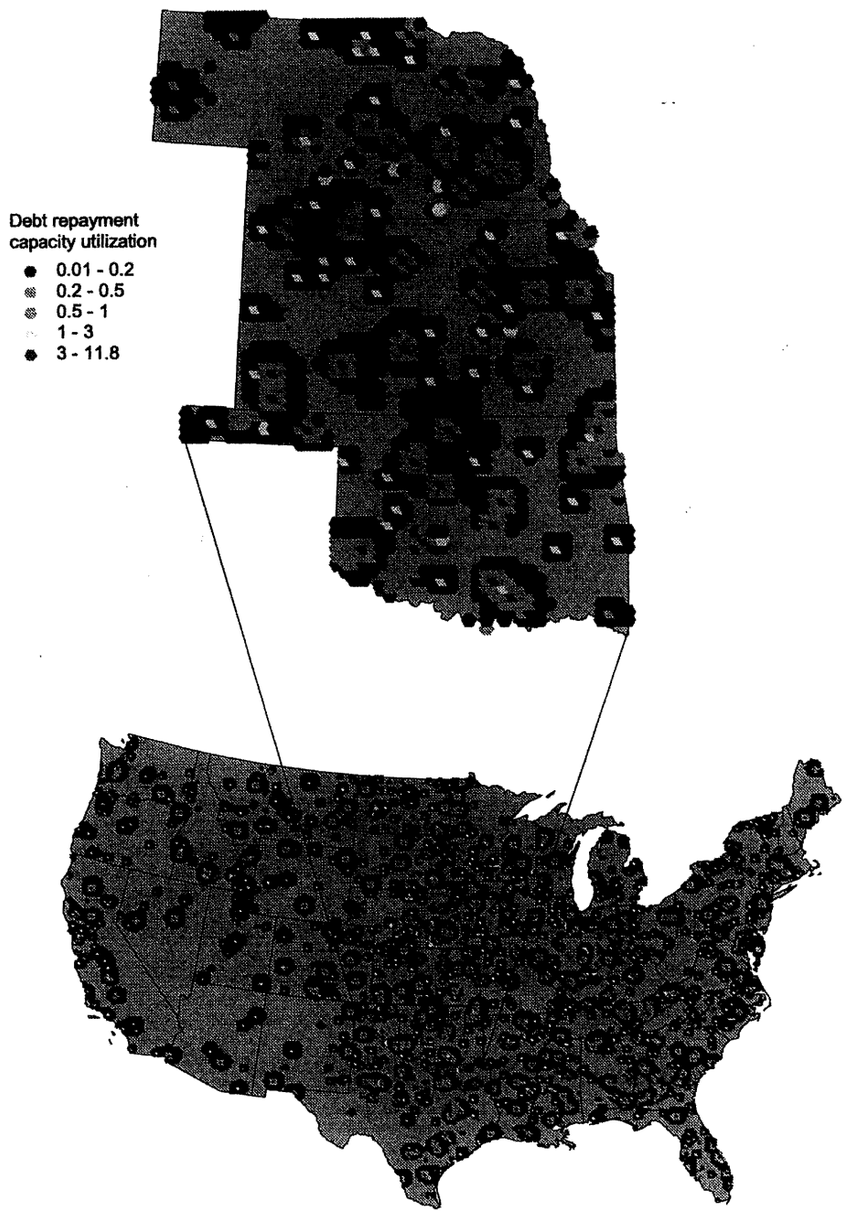


Figure 10: Result of wavelet regression (heteroscedastic noise model) of debt capacity utilization ratio, DCUR, data.