http://www.jaist.ac.jp/iskss/

International Journal of
Knowledge and Systems Sciences
Vol. 3, No. 1, March 2006

Spatial Recovering of Agricultural Values from Aggregate Information: Sequential Downscaling Methods

G. Fischer, T. Ermolieva, Y. Ermoliev, H.T. van Velthuizen

International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria

Abstract

In this paper, we propose a downscaling procedure that provides a basis for recovery and estimation of incomplete, aggregate, unknown or indirectly measurable variables. It makes maximum use of information and dependencies on various levels relying on the cross-entropy maximization principle. We show that the maximum entropy principle can be viewed as the extension of the maximum likelihood principle. In this sense, the convergence of the proposed downscaling methods to solutions maximizing an entropy function can be considered as an analog of the asymptotic consistency analysis in traditional statistical estimation theory.

The main motivation for the development of the procedure has been a practical example of spatial estimation of agricultural production values. We briefly discuss the main challenges related to the choice of priors (location specific information) and their inherent uncertainties that to large extent determine the success of the downscaled results.

Keywords: Cross-entropy, Maximum likelihood, Sequential downscaling

1. Introduction

The estimation of global processes consistent with local data and, conversely, local implications emerging from global tendencies challenge the traditional statistical estimation and data inference methods, which are based on the ability to obtain an infinite number of observations from an unknown true probability distribution. For the new estimation problems that can be termed “downscaling” problems (by generalizing the definitions in [2]), we often have only limited, partial, aggregate or incomplete statistics. For example, we can collect regional data on agricultural production while seeking to estimate the production on the level of fine spatial units (e.g., grid cells in a geographic information system). Alternatively, we may have aggregate income and consumption statistics, or occurrences of natural disasters on global and national levels. These aggregate statistics, however, do not provide any clue as to potential alarming diversity of conditions at specific locations, e.g., poverty, catastrophic losses, hazardous pollution, epidemics.

In this article we propose a recursive sequential downscaling method that can be used in a variety of practical situations. The main idea is to rely on an appropriate optimization principle and use all possible constraints connecting observable and unobservable dependent variables. We prove the convergence of this method to the solution maximizing a cross entropy function. The method was used for spatially explicit estimation of agricultural production outlined in Section 2. The problem is to attribute known aggregate national or sub-national crop production and land use to particular locations (grid cell) in accordance with features derived from geographical datasets and consistent with agronomic knowledge. This section summarizes also the well-known maximum entropy principle [8]. Section 3 shows that the maximum entropy principle can be viewed as an extension of the maximum likelihood principle. Therefore, the convergence of downscaling methods to solutions maximizing an entropy function may be interpreted as an analog of the traditional statistical asymptotic consistency [11] analysis. Section 4 develops the sequential downscaling method. It is shown that the convergence of this method to the solution maximizing a cross-entropy function follows from the duality theory [6], which significantly simplifies proofs and clarifies the convergence properties especially in case of rather general constraints. This opens up a way for various modifications and extensions, e.g., to situations with uncertainties when the available higher-level information is imprecise or involves stochastic elements. Section 5 describes the numerical calculations. Section 6 concludes.

2. Downscaling Problems

2.1. Spatial Estimation of Agricultural Production

Agricultural production data [7] are available at national scale from FAO and other sources. These data give no clue as to the spatial heterogeneities of agricultural production within country boundaries. A downscaling method in this case has to achieve a plausible allocation of aggregate national production values to individual spatial units, say pixels, by using available evidences. Satellite-based land cover images can provide detailed current information (up to pixels) on crop land. Besides these there exists other important unobservable or only partially observable information significantly determining the patterns and intensities of crop production. For example, biological and soil conditions, variations in radiation, temperature, humidity and rainfall, the occurrences of frosts, floods, and droughts. FAO developed crop-specific suitability maps using spatial data of climate and soil/terrain conditions. Crop price is among the most important market signals to crop allocation.

The available information can be summarized as the following. Total agricultural area a_i in a pixel i, $i = \overline{1,m}$, is estimated from satellite-based land cover interpretation, the suitable area for crops in pixels comes from geographically detailed suitability surface studies. There may also be derived information on the yield u_{ij} of crop j, $j = \overline{1,n}$ in pixel i, the price π_j of crop j, the production V_j of crop j in the country. Let x_{ij} be unknown estimates of area shares of crop j in pixel i. Assuming the total crop land area in pixel i is known, x_{ij} satisfy equations

$$\sum_{j=1}^{n} x_{ij} = 1 , \ i = \overline{1,m} \tag{1}$$

$x_{ij} \geq 0$, $i = \overline{1,m}$, $j = \overline{1,n}$.

This leads to the following plausible estimate $v_{ij} = a_i u_{ij} x_{ij}$ of production crop j in pixel i. Since the production V_j of crop j in the country is known, $\sum_{i=1}^{m} v_{ij} = V_j$, we also have equation

$$\sum_{i=1}^{m} a_{ij} x_{ij} = V_j \ , \tag{2}$$

where $a_{ij} = a_i u_{ij}$, $j = \overline{1,n}$.

2.2. The Maximum Entropy Formulation

Problem in Section 2.1. belongs to the so-called ill-posed since there may be an infinite number of feasible solutions x_{ij}, $i = \overline{1,m}$, $j = \overline{1,n}$ satisfying equations (1), (2). A unique solution can be obtained relying on the cross entropy principle, which guarantees a “best” solution making maximum use of the data and dependencies at hand.

The maximum entropy principle [8] has deep roots in the theory of information developed by Shannon [10] in 1948. It derives estimates of x_{ij} from the minimization of function

$$\sum_{j=1}^{n} \sum_{i=1}^{m} x_{ij} \ln x_{ij} \ , \tag{3}$$

since the entropy is defined as $-\sum_{j=1}^{n} \sum_{i=1}^{m} x_{ij} \ln x_{ij}$.

Usually, there exist some prior information on crop-specific area distributions, i.e., a prior distribution q_{ij} for crop j in pixel i, $q_{ij} > 0$, $i = \overline{1,m}$, $j = \overline{1,n}$. The prior can be based upon expert knowledge, available crop distribution maps, other relevant information, e.g., upon biophysical, soil, socio-economic characteristics. In this case, the cross-entropy maximization principle derives the estimates x_{ij} from minimization of the function

$$\sum_{j=1}^{n} \sum_{i=1}^{m} x_{ij} \ln \frac{x_{ij}}{q_{ij}} \ , \tag{4}$$

Minimization of practical global-wide problems at resolution of 5 min grid-cells utilizes a number of additional constraints which essentially increases computational time and makes the problem dependent on the choice of a “solver”. The alternative approach is to derive estimates x_{ij} from a certain “behavioral” principle. For example, it is reasonable to allocate crop j to pixels with maximum probable production values $\pi_j y_{ij}$. But it may lead to an overestimation or an underestimation of total known production V_j, $j = \overline{1,n}$, i.e., situations when the left hand side of (2) is greater or less than its right hand side, which requires a rebalancing procedure. Let us consider an idea of the balancing proposed by G.V. Sheleikovskii [1] for estimating passenger flows.

2.3. Projection of Interzonal Flows

The estimation may regard trade or migration flows between different regions, flows of passengers or transport in transportation systems, or flows of messages in communication systems. The downscaling methods estimate flows among given locations in a consistent way with statistics (or experts opinions) for expected total number of "departures" a_i from i-th locations and "arrivals" b_j in j-th locations. For unknown flows x_{ij} clearly $\sum_{j=1}^{n} x_{ij} = a_i$, $i = \overline{1,m}$, $\sum_{i=1}^{m} x_{ij} = b_j$, $j = \overline{1,n}$, i.e., we have a special case of constraints (1), (2) and a_{ij} = 1. Assume also that there is a prior probability q_{ij} for a "passenger" from i to choose the destination j. If a passenger from location i chooses the destination j with a prior probability q_{ij}, then the *expected* initial flow from i to j is $x_{ij}^0 = a_i q_{ij}$ with $\sum_j x_{ij}^0 = a_i$, $i = \overline{1,m}$, but there may be over (under) estimation of the statistics b_j on total inflows in j, $j = \overline{1,n}$ (i.e., $\sum_i x_{ij}^0 > b_j$ or $\sum_i x_{ij}^0 < b_j$). Calculate relative imbalances $\beta_j^0 = b_j / \sum_i x_{ij}^0$ and update $y_{ij}^0 = x_{ij}^0 \beta_j^0$, $i = \overline{1,m}$, $j = \overline{1,n}$. Now, $\sum_i y_{ij}^0 = b_j$, $j = \overline{1,n}$, but the estimate y_{ij}^0 may cause imbalance for departures a_i from i . Calculate $\alpha_i^0 = a_i / \sum_j y_{ij}^0$ and $x_{ij}^1 = y_{ij}^0 \alpha_i^0$, and so on. In summary, we can represent x_{ij}^k as $x_{ij}^k = a_i q_{ij}^k$, $q_{ij}^k = \left(q_{ij} \beta_j^{k-1}\right) / \left(\sum_j q_{ij} \beta_j^{k-1}\right)$, $i = \overline{1,m}$, $j = \overline{1,n}$. Assume $x^k = \{x_{ij}^k\}$ has been calculated. Find $\beta_j^k = b_j / \sum_i x_{ij}^k$ and $x_{ij}^{k+1} = a_i \left(q_{ij} \beta_j^k / \sum_i q_{ij} \beta_j^k\right)$, $i = \overline{1,m}$, $j = \overline{1,n}$, and so on. In this form the procedure can be viewed as a sequential redistribution of demands a_i from locations $i = \overline{1,m}$ among locations $j = \overline{1,n}$ by using a Bayesian type of rule for updating the prior distribution: $q_{ij}^{k+1} = q_{ij} \beta_j^k / \sum_i q_{ij} \beta_j^k$, $q_{ij}^0 = q_{ij}$. The update is done on an observation of imbalances of basic constraints rather than observations of random variables.

Proof of the convergence of the method to the solution maximizing the cross-entropy function $-\sum_{ij} x_{ij} \ln\left(x_{ij} / q_{ij}\right)$ was established in [1] by using lengthy arguments essentially relying on specific properties of the transportation constraints. In [6] we prove the convergence using duality theory, which allowed us to take into account general constraints (2) and to significantly simplify and clarify the analysis of the convergence to cross-entropy maximization solution. This opens up an opportunity for various modifications, in particular, to situations with uncertain parameters a_i, b_j, a_{ij}. Its general scheme will be briefly outlined in Section 4.

3. Minimax Likelihood and Maximum Entropy

In many applications inherent uncertainty can be characterized or interpreted in probabilistic terms, either as frequencies of underlying random variables or (subjectively) by degree of our believe. For example, in the estimation of crop production defined by equations (1) and (2), we can think of values $x_{ij} > 0$, $\sum_{j=1}^{n} x_{ij} = 1$, as the probability (the degree of our belief) that a unit area of pixel i is allocated to crop j. This interpretation forms the basis of the cross-entropy principle. It is also used in the sequential downscaling methods.

A key problem in the probabilistic models of uncertainty is the estimation of the true probability distribution. The standard statistical estimation theory derives information on this distribution from observations of underlying random variables. Most naturally, the estimate has to maximize the probability that a given sample is observed, the maximum likelihood principle [5]. In downscaling, the random variables are practically unobserved or not accessible to direct measurements. Let us show that the maximum entropy principle can be viewed as an extension of the maximum likelihood principle.

Consider a similar to Section 2 situation with available aggregate data and some unknown probability distribution $p = (p_1, \ldots, p_r)$. In other words, there is an underlying random variable ξ with $\Pr ob[\xi = \xi_j] = p_j$. The available information on a is given by a random sample $\xi(1), \ldots, \xi(N)$ of N independent observations of ξ. A maximum likelihood estimate of the unknown probabilities $(p_1, \ldots, p_r)$ is obtained by maximizing the probability (likelihood) of observing the sample $\xi(1), \ldots, \xi(N)$, $\prod_{k=1}^{N} \Pr ob[\xi = \xi(k)] = \prod_{j=1}^{r} p_j^{\nu_j}$, subject to constraints

$$\sum_{j=1}^{r} p_j = 1 , \; p_j > 0 , \; j = \overline{1,r} , \qquad (5)$$

where ν_j is the number of times the value x_j has been observed, $\sum_{j=1}^{r} \nu_j = N$. Or, as logarithm is a monotonically increasing function, by the maximization of the log likelihood function

$$\ln \prod_{j=1}^{r} p_i^{\nu_i} = \sum_{j=1}^{r} \nu_j \ln p_j \ . \qquad (6)$$

Normalized by the number of observations N, the sample mean function is $1/N \sum_{i=1}^{r} \nu_i \ln p_i$.

This is a continuous, strictly concave function on the set of R^n determined by linear constraints (5). By using the Lagrangian function (or the more general fact of Proposition 1 below) we can derive that the unique solution maximizing (6) subject to constraints (5) is the empirical probability function

$$p_j^N = \nu_j / N \ , \ j = \overline{1, r} \ .$$

The log likelihood function (6) is the sample mean approximation of the expectation

$$E \ln p_\xi = \sum_{j=1}^{r} p_j^* \ln p_j \qquad (7)$$

where the unknown probability distribution p_j^* is approximated by the frequencies ν_j / N derived from an available sample of observations $x^1, \ldots, x^N$.

In downscaling problems the available information about the unknown probability distribution p_j^*, $j = \overline{1, r}$ is given not by a sample of observations, but by a number of constraints of type (1) and (2) connecting this distribution with characteristics of observable variables. Let us denote by P the set of all distributions satisfying these constraints. If $x = (x_1, ..., x_r) \in P$, then we can consider

$$\sum_{j=1}^{r} x_j \ln p_j \qquad (8)$$

as an approximation of the expectation function (7) similar to the log likelihood function (6). We can now derive a conclusion analogous to maximization of the sample mean approximation $(1/N) \sum_{j=1}^{r} \nu_j \ln p_j$.

Proposition 1.

If x is an approximate probability distribution $x = (x_1, ..., x_r) \in P$, then

$$\max_{p \in P} \sum_{j=1}^{r} x_j \ln p_j = \sum_{j=1}^{r} x_j \ln x_j \qquad (9)$$

The log likelihood function (8) is defined for any feasible probability distribution $x \in P$. The worst-case principle leads to minimization of the maximum log likelihood function defined by (9):

$$\min_{x \in P} \max_{p \in P} \sum_{j=1}^{r} x_j \ln p_j = \min_{x \in P} \sum_{j=1}^{r} x_j \ln x_j \ ,$$

i.e., to the principle of maximizing entropy $-\sum_{j=1}^{r} x_j \ln x_j$. In the case of a given prior distribution q_{ij}, we may require the minimization of the difference between the log likelihood function (8) for $p \in P$ and the log likelihood function $\sum_{j=1}^{r} q_j \ln p_j$ for the given prior q_j, $j = \overline{1, r}$ from P:

$$\min_{x \in P} \left[\max_{p \in P} \sum_{j=1}^{r} x_j \ln p_j - \sum_{j=1}^{r} x_j \ln q_j \right]$$

$$= \min_{x \in P} \sum x_j \ln \frac{x_j}{q_j}.$$

In [6] we prove the assertion for distributions that belong to a certain parametric class, i.e., a parametric maximum entropy principle, which is important for situations when the balance equations of type (1) and (2) are given in some probabilistic sense. This is a key issue in dealing with uncertain parameters.

4. Sequential Downscaling

Consider the cross entropy maximization problem for spatial allocation of agricultural production:

$$\sum_{j=1}^{n} \sum_{i=1}^{m} x_{ij} \ln(x_{ij} / q_{ij}) \ . \qquad (10)$$

subject to constraints (1)-(2), where $q_{ij} > 0$, $a_{ij} > 0$ are given, $a_i > 0$, $b_j > 0$, $i = \overline{1, m}$, $j = \overline{1, n}$. For simplicity (and without loss of generality) we assume $x_{ij} > 0$. The prior distribution q_{ij} is normalized, i.e., $\sum_{j=1}^{n} q_{ij} = 1$, $i = \overline{1, m}$.

For a continuous, strictly convex function on a non-empty compact set of an Euclidian space there is a unique optimal solution to the minimization problem. Consider the Lagrangian function:

$$L(x, \lambda, \mu) = \sum_{i,j} x_{ij} \ln(x_{ij} / q_{ij})$$

$$+ \sum \lambda_i (a_i - \sum_{j=1}^{n} x_{ij}) + \sum_{j=1}^{n} \mu_j (b_j - \sum_{i=1}^{m} a_{ij} x_{ij}).$$

Since the optimal solution is positive, the optimality conditions lead to

$$\frac{\partial L}{\partial x_{ij}} = \ln \frac{x_{ij}}{q_{ij}} + 1 - \lambda_i - a_{ij}\mu_j = 0 \quad , \quad i = \overline{1,m} \quad ,$$

$j = \overline{1,n}$,

i.e., the optimal solution can be represented analytically as $x_{ij}(\lambda,\mu) = q_{ij} e^{\lambda_i - 1} e^{a_{ij}\mu_j}$, $i = \overline{1,m}$, $j = \overline{1,n}$ for some λ_i, μ_j.

The dual problem reads: find Lagrange multipliers (λ_i, μ_j), $i = \overline{1,m}$, $j = \overline{1,n}$, maximizing function

$$\varphi(\lambda,\mu) = \min_x L(x,\lambda,\mu) = L(x(\lambda,\mu),\lambda,\mu) \qquad (11)$$

From general results of convex analysis (see, for example, [9]) it follows that $\varphi(\lambda, \mu)$ is a strictly concave continuously differentiable function and the optimality condition can be written as

$$\frac{\partial \varphi}{\partial \lambda_i} = a_i - \sum_{j=1}^{n} x_{ij}(\lambda,\mu) = 0 \, , \, i = \overline{1,m} \, , \qquad (12)$$

$$\frac{\partial \varphi}{\partial \mu_j} = b_j - \sum_{i=1}^{m} a_{ij} x_{ij}(\lambda,\mu) = 0 \, , \quad j = \overline{1,n} \, . \qquad (13)$$

To solve the dual problem (11) let $\alpha_i = e^{\lambda_i - 1}$, and the optimal solution $x_{ij}(\alpha,\beta) = q_{ij}\alpha_i e^{a_{ij}\mu_j}$, $a_i > 0$, $i = \overline{1,m}$. If $a_{ij} = 1$, $i = \overline{1,m}$, $j = \overline{1,n}$, i.e., for the transportation constraints, and using notation $\beta_j = e^{\mu_j}$, the optimal solution can be represented as $x_{ij}(\alpha,\beta) = q_{ij}\alpha_i\beta_j$, $i = \overline{1,m}$, $j = \overline{1,n}$. This formulation is typical for the so-called gravity models [2]. Consider the following sequential method for updating variables $\alpha = (\alpha_1,...,\alpha_m)$, $\mu = (\mu_1,...,\mu_n)$ and $x = (x_1,...,x_m)$ to satisfy the optimality conditions (12) and (13). From aggregate data and prior distribution, compute $x_{ij}^0 = a_i q_{ij}$. Clearly, x_{ij}^0 satisfies (1), but constraints (2) may be violated. At step k, for given $x^k = \{x_{ij}^k\}$, find μ_j^k satisfying equations $\sum_{i=1}^m a_{ij} x_{ij}^k e^{a_{ij}\mu_j^{k-1}} = b_j$, $j = \overline{1,n}$. The left hand side of this equality is a monotonic function and the scalar value μ_j^k can be easily calculated. Calculate $y_{ij}^k = x_{ij}^k e^{a_{ij}\mu_j^k}$, and derive $\alpha_i^{k+1} = a_i / \sum_j y_{ij}^k$, $i = \overline{1,m}$, $j = \overline{1,n}$. Update x_{ij}^k to $x_{ij}^{k+1} = y_{ij}^k \alpha_i^{k+1}$, or $x_{ij}^{k+1} = x_{ij}^k \alpha_i^{k+1} e^{a_{ij}\mu_j^k}$, $i = \overline{1,m}$, $j = \overline{1,n}$, repeat until convergence is reached. In summary, the procedure involves a sequential updating of an a priori probability distribution q_{ij} by using a Bayesian type of rule $q_{ij}^k = q_{ij}\gamma_j^k / \sum_j q_{ij}\gamma_j^k$, $\gamma_j^k = e^{a_{ij}\mu_j^k}$, where γ_j^k is calculated using "observations" of imbalances instead of using observations of real random variables from unknown true probability distributions. The sequence $x^k = \{x_{ij}^k, i = \overline{1,m}, j = \overline{1,n}\}$, $k = 0,1,...$, converges to the solution x^*, $x^* = \lim x^k$, $k \to \infty$, of the cross-entropy problem (10) under constraints (1) and (2). For the transportation constraints $a_{ij} = 1$, $i = \overline{1,m}$, $j = \overline{1,n}$, the proposed method is reduced to Sheleikovskii's method, Section 2.

5. Practical Applications

The proposed method can easily be modified to reflect problem-specific peculiarities of constraints (1) and (2). A simplifying situation occurs when function $e^{a_{ij}\mu_j}$ is approximated by a function $A_{ij} f_j^{\mu_j}$, $i = \overline{1,m}$, $j = \overline{1,n}$, for some parameters $A_{ij} > 0$, $f_i > 0$, $i = \overline{1,m}$, $j = \overline{1,n}$, and μ_j varying in accordance with the range of plausible imbalances in (22). The solution of (22) in this case amounts to computing $\beta_j^{k+1} = b_j / \sum_i a_{ij} A_{ij} x_{ij}^k$, where β_j is defined as $\beta_j = f_j^{\mu_j}$. This proposed method has been applied globally for downscaling aggregate national and sub-national data on crop production and land use for all major countries [7]. The data on country-specific crop production (rain-fed and irrigated) was obtained from FAO. The list of crops included all major crops such as wheat, rice, maize, potato, soybean, pulses, oil crops, coffee, tea, tobacco, and cotton. Each allocation unit, usually countries was subdivided into regular grid cells with a spatial resolution of 5 by 5 minutes latitude/longitude (i.e., approximately 10 square kilometers at the equator). For each spatial unit the algorithm used information and constraints on the total cultivated area in this unit, percent of irrigated land, multi-cropping index (i.e., how many harvests may be obtained per year from a piece of land). The prior information guiding the downscaling algorithm also included suitability and attainable crop yields in pixels, crop prices, population density, and farming systems characteristics.

6. Concluding Remarks

This paper reviews some features of a new class of estimation problems, which have been termed “downscaling” problems, with unobservable and uncertain variables. We present a method that relies on appropriate optimization principles and uses all possible constraints connecting available ‘prior’ information at locations with other observable and unobservable dependent variables. For practical applications, the choice of appropriate ‘priors’, their inherent uncertainties and imprecision are among the major challenges of the downscaling methodology, ultimately determining the success of these procedures.

Extensive testing of the proposed procedure for downscaling of agricultural production, consistent with national statistics and compatible with various geographical and technical ancillary sources of information, has demonstrated that the iterative downscaling procedures are converging fast, allow for great geographical detail and are very flexible in model specification and detail.

In this paper we analyze numerical downscaling procedures only for situations when aggregate observed information is available and used as constraints on average values. For many practical situations this assumption may be insufficient and the procedures may need to be extended into more rigorous probabilistic treatment. For example, a prior probability q_{ij} for a “passenger” from location i to chose destination j generates a random flow ξ_{ij}, $\sum_j \xi_{ij} = a_i$, (a_i denoting aggregate departures from location i), leading to total random inflows $\sum_i \xi_{ij}$ in destination j. The analysis of only average flows $\overline{\xi}_{ij} = a_i q_{ij}$ $\overline{x}_{ij} = a_i q_{ij}$ may not be sufficient, e.g., for many facility location problems. Parameters a_i of constraints (2) for some problems with flows can be associated with a risk to loose a part of flow on the way (i, j) from i to j. Again, more rigorous risk based analysis requires probabilistic treatment of these constraints.

References

[1] L. M. Bregman, “Proof of the convergence of sheleikhovskii’s method for a problem with transportation constraints”. *Journal of Computational Mathematics and Mathematical Physics* 7/1, 191-204, Zhournal Vychislitel’noi Matematiki, USSR, Leningrad, 1967.

[2] M. F. P. Bierkens, P. A. Finke, P. de Willigen, *Upscaling and Downscaling Methods for Environmental Research*. Kluwer, Dordrecht, The Netherlands, 2000.

[3] Y. Ermoliev, *Stochastic Programming Methods*, Nauka, Moscow (In Russian), 1976.

[4] Y. Ermoliev, R. Wets (eds.), *Numerical Techniques of Stochastic Optimization: Computational Mathematics*, Springer-Verlag, Berlin, Germany, 1988.

[5] R. A. Fisher, “On the mathematical foundations of theoretical statistics”, *Philosophical Transactions of the Royal Society of London*, Series A222, 309-368, 1922.

[6] G. Fischer, T. Ermolieva, Y. Ermoliev, H. T. van Velthuizen, “Sequential downscaling methods for estimation from aggregate data”, In: K. Marti, Y. Ermoliev, G. Pflug, M. Makovskii (Eds.), *Proceedings of the IFIP/IIASA/GAMM Workshop On Coping with Uncertainty*, Springer-Verlag, Heidelberg (forthcoming).

[7] G. Fischer, H. T. van Velthuizen, M. M. Shah, F. O. Nachtergaele, “Global agro-ecological assessment for agriculture in the 21st century: methodology and results”, in *International Institute for Applied Systems Analysis*, Laxenburg, Austria, 2002.

[8] A. Golan, G. Judge, D. Miller, “Maximum entropy econometrics: robust estimation with limited data”, Series in *Financial Economics and Quantitative Analysis*, John Wiley & Sons Ltd, Baffins Lane, Chichester, West Sussex PO19 1UD, England, 1996.

[9] R. T.Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[10] C. Shannon. “A mathematical theory of communication”, *Bell System Technology*, Journal 27, 379-423, 1948.

[11] A. Wald, “Note on the consistency of the maximum likelihood estimate”, *Annals of Mathematical Statistics* 20, 595-601, 1949.