

# Um modelo de análise regional para fins de planejamento econômico: integração de sistemas de regiões formais e funcionais\*

---

SPERIDIÃO FAISSOL  
Geógrafo da Fundação IBGE

## 1. Introdução

O objetivo do presente documento é o de esboçar um ângulo de análise dos problemas da organização espacial de um país — com vistas aos problemas de um país como o Brasil — que permitam de um lado uma compreensão do processo espacial propriamente dito, nas suas múltiplas componentes estruturais, que em síntese têm levado a uma dicotomia diferente na Geografia: a da concepção de sistemas de regiões formais e funcionais; de outro lado e a partir desta dicotomia, tentar estabelecer as bases de um modelo que, ao mesmo tempo que analise estas estruturas, diferenciadas pela própria natureza dos fenômenos que associa e interrelaciona, estabeleça parâmetros das relações de interdependência entre uma e outra estrutura, que contenham, pelo menos em sua essência conceitual, um caráter explicativo e uma implicação preditiva.

O ponto de partida teórico é o de que geógrafos têm, tradicionalmente, utilizado uma variada gama de técnicas e níveis de generalização, capazes de produzir dois tipos de agregados espaciais:

1. Regiões homogêneas, constituídas de agregados de unidades espaciais (municípios), formando unidades maiores (microrregiões), à base de critérios que definem atributos dos lugares considerados. Uma matriz destes atributos descreveria o conjunto de lugares e métodos

---

\* Documento apresentado à CONFEGE, Comissão Mista "J".

apropriados de álgebra matricial identificariam as menores distâncias conceituais entre estes lugares e eventualmente os agruparia em unidades maiores, com parcimônia, embora sem perda substancial de informação.

2. Regiões funcionais, constituídas de agregados de unidades espaciais, (municípios), formando unidades maiores por via de associação funcional. Relações entre os lugares e complementaridade estrutural seriam conceitos essenciais, geradores destas relações e acessibilidade seria um ingrediente básico desta forma de organização. Uma diferença essencial existiria na construção de uma matriz contendo tais informações, pois embora sendo ainda uma matriz  $n \times n$ , conteria pares de lugares nas linhas, ao invés de lugares. Relação implica dois elementos, e matrizes de relações têm sido tradicionalmente construídas, com lugares nas linhas e colunas descrevendo relações elementares; o par de lugares na linha torna possível a mesma técnica analítica multidimensional.

A idéia de que a primeira matriz que define uma estrutura e a segunda que descreve um comportamento são estruturalmente interdependentes em um estado de complexo equilíbrio, encontra ampla validação, não só nas teorias econômicas correntes, como nos próprios modelos funcionais e gravitacionais. A planície isotrópica seria, a rigor, a eliminação da primeira matriz, com a premissa de uniformidade estrutural.

Até mesmo a experiência das circunstâncias em que muitas ações de intervenção no processo econômico se fazem, sugere estas relações: às vezes se constrói uma estrada, cujo objetivo essencial é o de transformar as estruturas locais, ou às vezes se é levado a construir uma estrada para atender à demanda gerada por uma transformação na estrutura local.

Discutimos, em seguida, os problemas teóricos fundamentais de relações entre regionalização, desenvolvimento regional e planejamento econômico, não com vistas a esgotar a análise do tópico, mas apenas para situá-los no contexto em que são colocados no documento, portanto deixando a maior parte dos conceitos pressupostos ou indicados na literatura própria.

O modelo, na sua forma conceitual e implicações metodológicas, é, a seguir, apresentado, seguido de um exemplo extremamente simplificado de natureza apenas exploratória, mesmo porque a sua aplicação ainda é um objetivo a ser alcançado, parte que é da programação normal das atividades do Departamento de Geografia nos próximos anos.

A guisa de conclusão e recomendações, apresentamos quase que a própria programação do Departamento de Geografia do IBG, naquilo que diz respeito à formulação de modelos de Divisão Regional do Brasil, para fins específicos, entre os quais se pretende que um deles seja o planejamento econômico.

É importante, ao fim, assinalar que: partimos de um objetivo significativo a atingir, mas procuramos começar com formas simples de trabalho, embora sem perder de vista o objetivo final.

## **2. Regionalização, Desenvolvimento Regional e Planejamento Econômico**

É claro que um tema desta amplitude não caberia ser discutido no âmbito de um documento do tipo do que estamos apresentando. Entretanto parece necessário especificar e conceituar palavras e temas usados no mesmo, para evitar ambigüidades e por via de consequência estabelecer consistência metodológica.

Neste sentido regionalização é entendida como uma forma de organização do espaço, multidimensional por definição analítica, mas bifocalizada por necessidade e por uma razoável generalização das estruturas de interrelações entre as variáveis definidoras dos dois processos básicos: estrutura e comportamento dos lugares. Assim região tem uma dupla conceituação, definida por atributos ou por relações, e possui contornos diferentes, em função de uma variada gama de associações por similaridade estrutural e por complementaridade funcional.

Em termos estritos de uma organização econômica do espaço, (1), temos os dois problemas fundamentais: produção e distribuição. O processo de produção pode ser definido, locacionalmente, em termos de uniformidade espacial de agregados econômicos; a distribuição da produção pode ser definida em termos de uma demanda (que tem uma importante componente espacial sob a forma de custo de movimento), e a principal característica do mecanismo de atendimento desta demanda é um sistema de relações entre lugares.

Em suma, os dois processos de regionalização constituem, na realidade, os dois processos fundamentais da atividade econômica: a produção e a distribuição, estruturas e comportamentos, uma estática e outra dinâmica. Estruturas diferentes são interdependentes por complementaridade e a própria essência da idéia de complementaridade é a Relação. É este o fundamento conceitual de dois modelos de Divisão Regional já elaborados e publicados pelo Departamento de Geografia do IBG, sob a forma de *Microrregiões Homogêneas* e *Regiões Funcionais Urbanas*.

Entretanto, sempre existiram dificuldades fundamentais de natureza operacional em tentar associar estes dois tipos de regionalização, a par de outras de natureza conceitual propriamente ditas. Procuraremos aqui apenas mostrar a significação dos dois para o entendimento dos problemas de desenvolvimento regional e por via de consequência para o planejamento econômico; fica também implícito que desenvolvimento regional e nacional e planejamento regional e nacional são conceitos correlatos, a rigor estritamente simétricos.

Poderíamos dizer que desenvolvimento regional tem a ver com os dois tipos de região, qualquer que seja o nível de tratamento do desenvolvimento regional que procurarmos abordar — local, regional ou nacional (por via de ser interregional), apenas com a afirmação de que o processo de desenvolvimento constitui um sistema de elementos interligados, ou que nenhuma área dentro de um país seria auto-suficiente e que portanto relações inter-regionais de produção e consumo fatalmente se estabelecem; o simples jogo das forças de mercado gera diferenciações regionais.

O problema fundamental é o de estabelecer-se o grau de interdependência entre os dois processos de organização do espaço: entre a geografia da produção e a de consumo, entre produção e fluxos. Um dos mais explorados aspectos de aplicação de modelos análogos nas Ciências Sociais tem sido o relativo aos modelos gravitacionais; eles, em essência, querem dizer que há uma relação de causa e efeito (a rigor de interdependência), entre a massa de um lugar e a de outro; entre uma unidade espacial homogênea (pois que o pressuposto do modelo gravitacional é o de uma massa homogênea) e outra, relação esta na qual o efeito da distância atenua ou acentua tal relação. O fato fundamental, que restringe a utilização de tais modelos e limita a sua operacionalidade é que os problemas de definição de massa e de efeito da distância são tão sensíveis e complexos, ao mesmo tempo que os tornam válidos somente a níveis altamente agregados, ou à custa de algumas premissas muitas vezes insustentáveis no mundo real. Estas dificuldades são agravadas pela quase inexistência de informações — em nível

próprio — de fluxos de mercadorias, pessoas, capital, etc., de forma a permitir comparações em diferentes níveis de generalização que permitam invalidar possíveis falácias ecológicas, contidas em modelos altamente agregados.

Discutiremos no capítulo seguinte problemas metodológicos relativos ao modelo que o Departamento de Geografia do IBG se propõe utilizar para tal fim, portanto aqui passaremos ao item seguinte: sua importância para o planejamento.

É claro que ao considerarmos o processo de desenvolvimento econômico como um sistema de relações simultâneas entre produção e consumo, ambos localizados e afetados por esta localização, uma metodologia analítica que permita a compreensão e comparação das duas estruturas, entre produção e fluxo, passa a ter uma conotação óbvia de planejamento. Mesmo considerando a natureza transversal (cross-section) de tal modelo, a repetição da análise em dois intervalos discretos de tempo ofereceria os parâmetros necessários à construção de simulações igualmente cabíveis para os propósitos do planejamento, inclusive porque apoiadas em interrelações concretas no âmbito dos dois subsistemas: o de produção e o de fluxo.

### **3. O modelo em seus problemas metodológicos. Um exemplo sumário e exploratório.**

Os dois processos de regionalização — homogêneo e funcional — já discutidos de forma sumária, podem ser descritos, matematicamente, em duas matrizes. Cada uma delas pode ser analisada separadamente e tem características próprias — na realidade como se disse antes, sempre foram analisadas mesmo sob a forma cartográfica, de forma independente uma da outra. A rigor, uma das controvérsias conceituais deriva da dificuldade em unir as duas matrizes em um só método analítico: a concepção de que são dois tipos de análise geográfica diferentes e inconciliáveis tem sido um tema de prolongado debate no âmbito da Geografia.

No caso de regiões homogêneas existe uma longa tradição analítica, e sempre constituiu o cerne dos estudos regionais. No caso de regiões funcionais, nodais, áreas de influência, etc., embora reconhecido de longa data, pelo menos desde geógrafos como HETTNER, é um tema mais recente e sobretudo tratado mais recentemente de forma mais ampla. Na realidade a Geografia Econômica foi, sempre, muito mais uma geografia da produção do que de consumo, inclusive porque os problemas de distribuição são mais difíceis de serem tratados de forma não matemática que os de produção. Por isso o estudo de regiões funcionais foi sendo ampliado e intensificado, ao longo da evolução teórico-quantitativa recente da Geografia. Recorde-se que o livro de CHRISTALLER, pioneiro neste sentido, data dos primeiros anos da década de 1930.

Atualmente tem havido considerável esforço metodológico no sentido de se desenvolverem técnicas adequadas à definição de regiões funcionais, ao lado de outras tentativas teorizantes de juntar os dois conceitos de homogeneidade e funcionalidade. NYSTUEN e DACEY realizaram um dos mais significativos, (2) tentando utilizar os conceitos da teoria dos grafos para definir regiões nodais ou funcionais, e resultados bastante promissores têm sido obtidos em numerosas tentativas similares, inclusive no Departamento de Geografia da Fundação IBGE.

Outra tentativa significativa foi realizada por BROWN (3), utilizando estatísticas geradas no contexto de uma concepção markoviana do processo, utilizando a noção de que “o tempo médio de primeira pas-

sagem”, estatística gerada em uma cadeia de MARKOV regular, podia ser entendida como uma medida relativa de distância funcional percebida, entre dois pares de lugares. A implicação era a de que a matriz de probabilidades de transição, que reflete a quantidade de movimento entre um lugar e outro (um estado e outro da cadeia), já traz consigo a informação total das causas que geram o movimento; a distância funcional, assim determinada, é conceitual, portanto não euclidiana.

O problema fundamental, entretanto, é que em ambos os casos partimos de uma matriz quadrada, com lugares nas colunas e nas linhas, e os fluxos medidos são singulares. É claro que numerosos procedimentos matemáticos podem somar, com ou sem ponderação, fluxos singulares, produzindo resultados compósitos (DACEY usou telefones por considerá-los uma forma singular que mais reflete o complexo de interrelações entre lugares).

BRIAN BERRY foi o primeiro geógrafo que procurou uma solução diferente, utilizando métodos fatoriais comumente usados na definição de regiões homogêneas, para análise de regiões funcionais; a simples diferença é a de que ele introduziu pares de lugares nas linhas, permitindo-lhe preencher as colunas com os diferentes tipos de fluxos, relevantes à definição da funcionalidade do processo regional. A vantagem adicional é a de que os fluxos são interrelacionados no contexto de uma matriz (seguindo-se o procedimento clássico de análise fatorial), e temos regiões funcionais eventualmente produzidas por similaridades estruturais dos próprios fluxos. Entretanto a motivação básica de BERRY, neste trabalho, foi a de desenvolver a sua “Spatial Field Theory” (4), que postula a interdependência entre as duas matrizes mencionadas anteriormente, isto é, entre um sistema de regiões homogêneas e funcionais.

BERRY assinala que “os padrões fundamentais espaciais que resumizam as características das áreas, bem como os padrões de comportamento espacial que são a essência das interações que se processam entre as mesmas áreas, são interdependentes e basicamente isomórficos” (5).

Segundo BERRY análises transversais em diferentes momentos de tempo podem revelar a natureza das transformações que se processam nestas duas estruturas segundo as seguintes alternativas:

1 — a) O sistema de interação entre dois lugares é função da forma que toma a estrutura do espaço.

b) Mudanças no comportamento, isto é, nas interações, resultam de mudanças nas características estruturais dos lugares, ao longo do desenvolvimento do processo espacial.

2 — a) As características estruturais dos lugares são basicamente dependentes de suas relações com outros lugares, isto é, estreitamente associadas às mesmas.

b) Mudanças nas interações entre lugares dão como resultado mudanças nas características estruturais dos referidos lugares.

A reversibilidade entre os dois aspectos do processo espacial constitui o fundamento desta teoria de integração espacial.

A fórmula matemática encontrada por BERRY foi a da correlação canônica, uma técnica de natureza multivariada, como tem sido a tônica de numerosas outras aplicadas à análise quantitativa de problemas geográficos. Ela combina as vantagens de uma análise fatorial, no sentido da parcimônia, portanto, com propriedades de síntese muito importantes, com as vantagens da análise de regressão, e todas as suas potencialidades preditivas.

Correlação Canônica é um procedimento matemático mais ou menos antigo, pois HOTELLING a utilizou desde 1935 (6), mas no campo

da Geografia, BRIAN BERRY foi dos primeiros a dar-lhe um conteúdo, quase que revolucionário, (7), ao utilizar suas propriedades para estabelecer as relações entre um sistema de regionalização formal (homogêneo) e outro funcional, atribuindo aos seus resultados uma capacidade preditiva matematicamente avaliável, por via da Regressão, de enorme significação prática e teórica. Outra aplicação de substancial importância foi feita por GAUTHIER, já agora com exemplos brasileiros, (8), procurando estabelecer as interrelações entre a acessibilidade de um determinado ponto e o crescimento urbano, identificando o que ele denominou de efeito de liderança da acessibilidade sobre o processo de crescimento urbano. Neste caso também a natureza potencialmente preditiva dos resultados de tais investigações parece muito óbvia, desde que fique matematicamente estabelecido uma espécie de parâmetro entre acessibilidade a um centro urbano ou a partir dele, e o crescimento urbano do referido centro e de um sistema de centros urbanos ao mesmo associado.

Em estudo recente, apresentado à reunião da Comissão de Métodos Quantitativos da União Geográfica Internacional, realizada no Rio de Janeiro, em abril de 1971, Michael Ray apresentou uma nova aplicação de correlação canônica (9), à qual alguns índices muito expressivos foram acrescentados (10). Em primeiro lugar o trabalho de RAY introduz "scores" dos lugares, o que ainda não havia sido feito antes. Esta contribuição — a do "escore", tem uma importância especial, pois que até agora o que se produzia era uma análise em termos de uma estrutura de inter-relações, quer dizer, em termos do processo que comandaria as relações entre os dois tipos de estruturas que se procuraria comparar, equivalente, de certo modo, à matriz de "loadings" na análise fatorial. O "escore", à semelhança do que ocorre na análise fatorial, constitui uma métrica na qual se posiciona cada lugar, por referência à acima mencionada estrutura de inter-relações, vale dizer segundo os efeitos que uma estrutura produz sobre a outra. Portanto, duas matrizes de "loadings" e de "scores".

O método produz medidas de interrelações entre um conjunto de medidas em um vetor  $Z_1$  e outro conjunto de medidas em outro vetor  $Z_2$ , quando cada vetor das medidas foram tomados em uma dada população.

Hotteling (2) usou o método, conforme assinala RAY (9), para determinar a possibilidade de substituição do vetor  $Z_1$  pelo vetor  $Z_2$  na medida de uma observação. Daí advém a sua evidente conotação preditiva, e também uma clara conotação de inter-relações, desde que cada uma das medidas tenha um significado relevante na pesquisa geográfica; o seguimento de tal proposição é o de que dados dois conjuntos de mensurações em uma determinada área, as relações estruturais entre ambas possam ser exploradas entre as duas por via de correlação canônica e a posição de cada lugar, na área considerada, possa ser definida em uma métrica semelhante à da análise dimensional ("scores" dos lugares nos algoritmos correntes de análise fatorial), já levado em conta o efeito de um conjunto de características sobre o outro. A importância do método advém precisamente desta possibilidade, isto é, dadas as duas matrizes, no contexto das quais existiriam interrelações, mas entre as quais existiria uma certa interdependência (do tipo da que existiria entre um sistema de regiões formais e funcionais), de estabelecer-se o grau de interdependência com a conseqüente capacidade de calcular-se uma, dada a outra. Repetida uma análise deste tipo em dois momentos de tempo, ficaria evidenciada a trajetória do processo comum e interdependente, e possível a interferência dos efeitos de transformações em uma delas, desde que a outra fosse alterada por via de ações de inter-

venção normalmente de origem governamental. Esta a sua potencialidade preditiva, de evidente significação para objetivos de planejamento do desenvolvimento econômico.

A vantagem deste método, assinalada por RAY (9), é a de que, da mesma maneira que a análise fatorial, tal técnica produz uma matriz de estrutura fatorial canônica, que revela não só as interrelações no interior de cada conjunto de dados, mas também as interrelações entre um conjunto e outro; portanto intravetorial e intervetorial.

COOLEY e LOHNES descrevem, em detalhe, (10) o procedimento matemático que difere do procedimento fatorial simples no sentido de que este "seleciona funções de testes que tenham máxima variância, sujeito à restrição da ortogonalidade. O modelo canônico seleciona funções lineares que tenham máxima covariância entre domínios, sujeito ainda à restrição de ortogonalidade. Geometricamente, o modelo canônico pode ser considerado uma exploração da extensão em que indivíduos ocupam a mesma posição relativa em uma medida do espaço, assim como na outra." pp. 169. Em termos estritamente espaciais, queremos saber em que medida o vetor  $Z_1$  mapeia um lugar na mesma medida que o vetor  $Z_2$  o faz; isto significa que, nas correlações no interior de um domínio (ou de uma das matrizes), o fator canônico, com o qual cada variável se correlacionaria, é, em primeiro lugar, aquele fator que se correlaciona, de forma maximizada com outro semelhante da outra matriz; a seguir este fator correlaciona-se com cada variável que o compõe, indicando, assim, e simultaneamente, covariância nos dois domínios e estrutura da variação em cada um deles. Esta última composição é importante, porque ao mesmo tempo indica quais os elementos que mais fortemente contribuem para tal fator maximamente correlacionado e permite o cálculo da redundância, como se verá mais adiante, através, ao mesmo tempo, do peso do fator (o seu eigenvalue) e sua correlação como outro fator da outra matriz.

O algoritmo apresenta ainda a possibilidade de calcular testes estatísticos, tanto do quiquadrado, como do lambda, de maneira a se verificar a hipótese nula de  $Z_1$  não estar relacionado com  $Z_2$ . Este teste é realizado para o primeiro fator canônico e, afastada a hipótese nula, os cálculos para os fatores seguintes prosseguem interativamente, cada um com uma correlação menor entre os dois domínios. COOLEY e LOHNES adotam, como regra apenas convencional, considerar trivial uma correlação inferior a 0,30, quando se costuma utilizar 0,40 na análise fatorial comum; é que a restrição de uma correlação, no interior de cada matriz e covariância máxima com a outra matriz, torna prudente diminuir o nível de definição do trivial na correlação.

Em seqüência, o algoritmo calcula e imprime diversas matrizes:

1. Uma matriz de dispersão para o conjunto de variáveis utilizadas nas duas matrizes, na qual os valores aparecem com sua média tornada igual a zero e dispersão em torno desta média. A significação da matriz, além de permitir outros cálculos subseqüentes, é a de oferecer desde logo uma visão não só da variação dos valores, de forma comparada, mas também sua dispersão em torno da média.

2. Uma matriz de correlações, ainda com as variáveis de todas as duas matrizes, que permitirá uma análise das correlações que existem entre todo o conjunto de variáveis nos dois domínios, uma com cada outra.

3. Uma matriz de pesos dos vetores, estandardizados, na Regressão de cada elemento de  $Z_1$  em todos os elementos de  $Z_2$ , já agora uma matriz para cada domínio, isto é, eventualmente um domínio A e outro domínio B. Esta matriz permitirá a definição posterior de uma outra de estrutura e outro padrão.

4. Uma matriz estrutural, derivada da anterior, contendo as correlações de  $p_1$  elementos de  $Z_1$  com a função de regressão definidos nos mesmos para estimar os elementos  $p_2$  de  $Z_2$ , função esta que é nada mais nada menos que o fator canônico. O fator canônico é a regressão de um vetor de  $Z_1$  sobre outro de  $Z_2$ .

Além dos testes que são calculados para se verificar a validade da análise, o algoritmo computa algumas outras estatísticas. Vários índices são calculados a partir da matriz 4, de estrutura fatorial canônica.

1. A soma dos quadrados dos valores, ao longo das linhas de uma ou de outra matriz, produz a comunalidade da variável, que é a proporção da variância explicada ou extraída pelos vários fatores canônicos; a soma das colunas (também o quadrado para eliminar os sinais negativos), dividida pelo número de linhas, numa ou noutra matriz dá a proporção da variância generalizada em todo o domínio, por cada fator canônico. Portanto, a primeira mede a distribuição da variável ao longo de todos os fatores, a outra o peso de cada variável num determinado fator, seguindo, ainda aí, caminhos semelhantes ao da análise fatorial, com a diferença de que nesta última o eigenvalue é a soma simples do quadrado das correlações, sem a divisão pelo número de linhas.

2. A redundância de cada fator em um conjunto de medidas, quando o fator correspondente no outro conjunto é conhecido, é dada pelo produto da proporção da variação extraída pelo fator com o quadrado da correlação canônica, quer dizer coloca a correlação canônica entre um fator de uma matriz e outro fator de outra matriz como função modificadora ou modificada da significação do fator propriamente dito. Isto equivale a dizer que uma correlação elevada entre um fator e outro pode ser “reduzida à insignificância se sua correlação canônica com o fator correspondente do outro domínio for baixa”, pp. 2 do manuscrito.

Esta redundância ainda é particularmente importante, porque ela não é igual e em sentido contrário, precisamente porque embora a correlação canônica seja simetricamente igual, a proporção da variância que cada fator explica em cada matriz não é necessariamente igual. Assim dado o domínio descrito pela matriz de *status* econômico por exemplo, um fator pode correlacionar-se com outro da matriz de *status* social de forma elevada e a proporção da variação extraída pelo fator da primeira matriz sendo elevada, a redundância assume um valor elevado. Entretanto, dado o fator correspondente na matriz de *status* social, eventualmente com uma baixa proporção da variância explicada, para a mesma correlação (que é no caso simétrica obviamente) a redundância assumirá um valor baixo, eventualmente “reduzido à insignificância”. Isto equivale a dizer, na realidade, o significado de um fator sobre o outro, em termos da descrição de uma relação de dependência, principalmente se considerarmos que o procedimento analítico constitui, a rigor, de uma regressão de um fator sobre o outro.

Da mesma forma que a soma do quadrado dos valores em cada variável, ao longo dos diversos fatores, produz a comunalidade da variável, a soma das redundâncias de cada fator em um domínio, dado o outro domínio, produz um índice que representa a redundância total de um conjunto de fatores em um domínio sobre o conjunto de fatores em outro domínio; isto significa a redundância de uma estrutura (a dos atributos) dada a outra (a dos fluxos), que no fundo constitui uma regressão de uma variável sintética sobre outra sintética, cada uma representando uma estrutura complexa.

A extensão desta concepção e deste procedimento analítico, como o fez BRIAN BERRY, a uma matriz que nas linhas conteria pares de lugares, e não lugares tomados isoladamente, tornaria possível a análise de duas estruturas complexas, uma representando uma regionalização formal, a outra representando uma regionalização funcional.



No caso da regionalização funcional pode-se partir de pares de lugares nas linhas, desde que nas colunas os valores já representariam, por definição, a interação entre aquele par de lugares.

No caso da regionalização formal o procedimento é mais laborioso, pois ou parte-se dos pares de lugares e cada variável atributo seria representada pela diferença entre o lugar A e o lugar B (o par de lugares como indicado na regionalização funcional) ou os dados iniciais já seriam "scores" dos lugares derivados de análise fatorial previamente preparada e de novo teríamos a diferença entre os valores de A e B, representando o valor do par AB, aí então comparável ao valor na matriz que representasse a regionalização funcional. Análise fatorial nas duas matrizes produziria resultados, cujos valores já teriam passado por idêntico processo de agrupamento em fatores (portanto com variáveis já correlacionadas e máxima parcimônia), ganhando-se uma etapa no algoritmo da correlação canônica.

A principal dificuldade para análise de detalhe das estruturas de interação entre pares de lugares, e sua comparação com a estrutura dos lugares propriamente dita, é, ao mesmo tempo, a falta de dados de fluxos entre pares de lugares de um lado e de outro o fato de os fluxos serem fortemente assimétricos, sendo muitas vezes difícil, quase impossível montar uma matriz de fluxos entre pares de lugares, sem um elevado número de zeros nas colunas da matriz respectiva. Uma vez adotado um nível de generalização, em termos de unidades taxonômicas observacionais, no que diz respeito aos lugares e de agregados de fluxos que reflitam o processo de interação entre tais unidades taxonômicas, então pode-se proceder à análise subsequente; é preciso ressaltar, entretanto, que isto não é fácil de se obter, nem operacionalmente nem conceitualmente. Operacionalmente a dificuldade é a da obtenção de dados de fluxos suficientemente desagregados em termos de unidades espaciais e suficientemente agregados em termos de tipos de fluxos, que a matriz seja representativa de áreas funcionais e os fluxos constituam unidades de interação efetivamente significativas.

A dificuldade conceitual é que as unidades taxonômicas observacionais, para a medida dos fluxos, serão, necessariamente, as mesmas unidades utilizadas para se mapear os atributos dos lugares e eventualmente as diferenças estruturais entre um lugar A e outro B, para ser utilizado na matriz de estrutura. Como conceber uma unidade taxonômica observacional que não contenha, intrinsecamente, autocorrelação espacial, seja no contexto da regionalização formal se a unidade for derivada de análise de fluxos, seja no contexto da regionalização funcional se a unidade for derivada da análise das características estruturais dos lugares? Uma solução de compromisso teria que se apoiar na premissa de que, dada uma unidade qualquer, o propósito subsequente da análise seria o de identificar a estrutura de interrelações entre as duas matrizes, ignoradas eventuais autocorrelações espaciais. Por esta via do compromisso o caminho seria o do teste de várias hipóteses sobre agregação de dados em unidades espaciais diferentes, a fim de se verificar o melhor nível da resolução do problema, enquanto se desenvolve a pesquisa na direção de uma maximização das correlações canônicas por via de reagrupamento das próprias unidades observacionais, em um processo iterativo que abrangeria simultaneamente linhas e colunas das duas matrizes; a dificuldade seria, possivelmente, a extrema laboriosidade do processo iterativo, que consumiria um tempo muito grande de computação, o que o tornaria muito dispendioso, especialmente se as unidades fossem numerosas. Em todo o caso o caminho está aberto para pesquisas, em um setor de análise que promete ser um dos mais proveitosos, dada a sua potência analítica e seu conteúdo preditivo.

*Um exemplo sumário e exploratório.* Do que ficou dito na apresentação sintética do modelo, a sua essência matemática, que é a correlação canônica, constitui um caso especial de Regressão, no qual a variável dependente é uma matriz e a independente é outra matriz. O exemplo que vamos apresentar é o de um caso particular de regressão, no qual a variável dependente é a distância funcional entre lugares, e variáveis independentes hipotetizadas como ligadas àquela distância são introduzidas, compondo um modelo causal clássico no uso de regressões. A distância funcional utilizada foi derivada da aplicação de um modelo Markov I (cadeia regular), representada pelo “tempo médio de primeira passagem”, utilizando-se dados de migrações de origem e destino urbano, tabuladas ao nível das 10 regiões utilizadas nas Tabulações Avançadas do Censo Demográfico. A tabela I indica estas distâncias, para cada par de região, em números inteiros e arredondados, para simplificação, e contém, por definição, a informação que se pressupõe ter determinado a decisão de migrar. Como o sistema urbano de São Paulo é o que supõe-se o maior emissor de informação ao migrante potencial, de origem e destino urbano, a distância para ele é menor de todas. A variância e o desvio padrão no sistema, não sendo grandes, tornam o sistema relativamente integrado, daí termos usado apenas a média das distâncias de cada lugar, (para São Paulo, por exemplo), obtendo assim um valor único para cada lugar, a ser usado como variável dependente. Por analogia esta seria a matriz de relações ou de comportamento, como BERRY a denominou. As variáveis independentes, em número de oito, definiriam os atributos dos lugares, portanto, a matriz de estrutura como BERRY a denominou, caracterizando a primeira a regionalização funcional, a segunda regionalização formal. Dada a natureza preliminar e exploratória do exemplo, estamos partindo do pressuposto de que estas oito variáveis (veja-se tabela 2), reflitam os principais aspectos da estrutura formal, relevantes no contexto da decisão de migrar de uma origem urbana ou rural para um destino urbano.

TABELA I

MATRIZ TEMPO MÉDIO DE PRIMEIRA PASSAGEM										
MIGRAÇÕES RURAL-URBANO					PARA DEZ REGIÕES BRASILEIRAS					
1	99,3	845,2	365,5	70,10	78,1	30,3	21,1	39,6	269,1	67,10
2	802,2	317,9	267,7	88,4	72,2	44,8	14,8	33,4	261,8	45,7
3	933,0	829,5	164,5	82,4	72,1	37,1	10,7	29,1	258,9	56,7
4	971,6	879,6	416,8	55,5	67,2	48,3	6,4	25,2	255,3	52,8
5	969,3	881,4	420,2	83,9	33,7	35,3	9,10	28,1	257,9	54,8
6	976,3	883,7	417,8	54,10	66,0	6,5	14,6	32,8	262,3	62,2
7	975,3	884,1	420,8	104,3	75,1	58,2	2,1	20,5	251,1	54,9
8	977,7	886,6	422,4	106,3	78,0	60,1	7,4	5,10	233,9	57,7
9	985,1	895,5	432,0	111,9	86,7	65,8	17,4	16,8	23,7	67,2
10	931,8	340,8	415,1	97,5	57,6	50,2	13,3	30,10	257,9	9,7

Dados os elementos essenciais do modelo — na sua forma tentativa e simplificada como foi esclarecido — foram feitas dez regressões, uma para cada região, utilizando-se as mesmas oito variáveis independentes, e a distância média como variável dependente. É importante assinalar que mesmo considerando o fato de termos dez observações apenas, estamos usando oito variáveis, o que torna o resultado da regressão estatísti-

camente discutível, o objetivo é o de apenas demonstrar o método; como a Regressão foi feita "stepwise", cada variável independente foi controlada.

Não é o objetivo do presente documento apresentar resultados empíricos sob a forma de uma análise e interpretação de um conjunto de dados. Apresentamos de um lado os valores das distâncias funcionais obtidas na análise Markoviana e agora alguns exemplos isolados das Regressões feitas, apenas como ilustração metodológica e como indicativas da validade do modelo, nesta sua fase experimental.

TABELA II

*Variáveis Independentes*

	% Pop. urbana	% Pes- soal curso Sec.	% Pes- soal curso Sup.	% Pes- soal ocup. ind.	% Pes- soal ocup. serv.	Renda + 500	Renda (200,500)	Renda (200)
I	45,1	12,7	0,7	11,2	19,0	5,4	18,1	61,7
II	27,6	7,1	0,4	6,4	11,4	1,5	5,3	78,2
III	46,3	10,2	0,8	12,0	20,6	3,6	8,1	75,9
IV	41,6	9,5	0,6	11,3	18,0	4,0	10,3	70,2
V	51,7	17,2	0,8	14,5	23,5	5,3	14,5	70,6
VI	88,0	28,7	1,9	25,6	43,5	19,6	33,2	42,5
VII	80,4	32,4	1,8	31,2	30,8	17,6	30,2	45,7
VIII	36,1	14,4	0,7	10,3	16,9	5,4	17,0	56,8
IX	50,1	19,6	1,1	17,1	22,9	8,0	20,8	51,5
X	48,0	14,7	0,8	11,5	23,8	6,3	16,5	68,8

Por exemplo, os principais resultados da regressão das distâncias médias de São Paulo, com as variáveis indicadas, mostram em primeiro lugar que a variável mais importante em explicar a distância funcional, que, por sua vez, é a resultante do processo de migrações internas, é a relativa a pessoal ocupado em Serviços. Não é nem a população urbana nem o pessoal ocupado na indústria, embora deva-se considerar aí que estamos utilizando números relativos e não absolutos, pois estamos procurando uma relação estrutural e não o poder de atração de massas padrão. O valor de B para São Paulo é de 0,85, enquanto seu erro padrão é de 0,35. O valor de F para este passo da Regressão é de 2,89 para 3/6 de grau de confiança, o que o coloca entre 0,25 e 0,10 de probabilidade de ser aleatório, portanto, com razoável grau de significação estatística. Ao mesmo tempo se considerarmos o racional do processo migratório, podemos facilmente compreender que o setor serviços sempre tem constituído o natural desaguadouro inicial do migrante, (no caso de São Paulo mais do migrante inter-regional do que o intra-estadual), mas de qualquer maneira parece ser o mais significativo no agregado.

Já no caso da região VI, constituída da Guanabara e Estado do Rio, a variável mais significativa é renda superior a Cr\$ 500,00, seguida da de pessoal ocupado nos serviços, ambas com erro padrão igual à aproximadamente metade do valor de B, a primeira com  $R = 0,62$  e a segunda com 0,71, mas a primeira com valor F proporcionalmente maior, embora sem grande diferenciação.

Para o Estado do Paraná a variável mais importante é renda superior a Cr\$ 500,00, pois além do valor R ser o mais alto (0,92), o erro de B é metade de seu valor (0,62 para 1,12), o valor F é de 4,75, signifi-

cativo entre 0,10 e 0,005 de probabilidade de ser aleatório, o que é também o mais elevado. A esta altura parece importante chamar a atenção para o fato de estarmos tratando de migrações de origem e destino urbano, diferente do resultado que se obteria se fosse destino rural, no caso particular do Paraná; o mesmo padrão foi observado para a região 9.

Como o objetivo não é, a rigor, analisar o comportamento regional, mas apenas ilustrar uma metodologia analítica aplicada às relações entre estruturas e comportamentos, bastaria apenas mencionar que a única região em que o percentual de população urbana foi a variável mais importante foi a região II (Maranhão e Piauí), embora a nível de significação bastante baixo; na região IV (Sergipe e Bahia) nenhuma das variáveis chegou a ser significativa, o mesmo ocorrendo com a região X, muito heterogênea, e de migrações muito recentes e afetadas pelo Distrito Federal, para apresentar qualquer padrão discernível.

Ainda pareceria relevante salientar, ao fim desta rápida descrição de alguns aspectos do modelo, e esclarecer que ele constitui um exemplo extremamente simplificado, desde que de um lado definimos um vetor (distância funcional), a partir de um setor do processo de migrações internas (o de origem e destino urbano), e de outro lado definimos como relevantes, para definir a estrutura, oito variáveis bastante correlacionadas entre si, e caracterizadoras quase que estritamente do nível de desenvolvimento. Subjacente a esta escolha está, por assim dizer, a hipótese de que as migrações estão estritamente relacionadas com as oportunidades econômicas, ora descritas pelo nível de renda, pelo setor serviços ou pelo industrial. Implícita fica a idéia de que na própria raiz de uma aplicação bem sucedida de modelos deste tipo, estará a adequada escolha e fundamentação teórica das variáveis relevantes. O algoritmo usado (Biblioteca do Instituto Brasileiro de Informática, e componente do Standard Package for Social Sciences, SPSS), dado as múltiplas estatísticas que calculam e imprimem e ao fato de ser a regressão do tipo "stepwise", permite a detecção de numerosas incongruências teóricas, facilitando a sua correção posterior. No presente esboço algumas estariam em ordem, e estarão sendo feitas ao longo do processo da pesquisa, mas não necessariamente neste documento que, como se disse, é exploratório, indicativo, essencialmente preliminar.

#### **4. Conclusões e Recomendações**

A concepção do processo de regionalização como capaz de ser descrito em duas matrizes diferentes, que contenham: de um lado os atributos dos lugares posteriormente transformados de forma a se constituírem em distâncias escalares entre os lugares em um espaço multidimensional e, de outro lado as relações entre estes lugares, sob a forma de fluxos de vários tipos que, por sua vez, definam o sistema de interação entre cada par de lugares, parece corresponder, de forma coerente, aos requisitos da teoria geográfica sobre regiões.

Ao mesmo tempo tal concepção torna possível a aplicação de métodos adequados à comparação de duas matrizes, por via da correlação canônica, um caso particular de Regressão combinado com análise fatorial, que vem sendo desenvolvida pela via de algoritmos apropriados, visando atender às necessidades da pesquisa geográfica.

O significativo da combinação dos dois métodos é a possibilidade de usar uma matriz como um todo, como variável independente, gerando fatores que primeiro se correlacionem, maximamente, com outro fator gerado na outra matriz, seguindo assim a linha da regressão de um sobre o outro. Assim fica definida uma relação causal de enorme signi-

ficado prático para fins de planejamento. A seguir variáveis de cada domínio são correlacionadas com cada fator, definindo assim uma estrutura interna do fator, já previamente definido como associado a outro, do outro domínio. À medida que se desenvolve a tecnologia adequada, testes especiais e medidas de significância vão sendo adicionadas, facilitando e tornando menos sofisticado seu uso por parte de técnicos não familiarizados com o rigor matemático de procedimentos assim complexos.

Um modelo deste tipo está sendo trabalhado no Departamento de Geografia da FIBGE, visando essencialmente à análise dos dois sistemas de regiões, formais e funcionais, com vistas à sua utilidade para planejamento. É preciso salientar que não se trata, necessariamente, de um modelo de divisão regional para fins de planejamento, pois sendo o planejamento um processo específico, os seus objetivos podem ser melhor atendidos tanto a partir de áreas homogêneas como de áreas funcionalmente associadas.

A perspectiva que tal procedimento analítico oferece, dadas as suas características básicas — a combinação de um modelo causal como o de regressão, com outro de inter-relações do tipo análise fatorial — abre amplas oportunidades de utilização para fins de planejamento, desde que aplicado com as devidas cautelas, e apoiado em concepções teoricamente consistentes. Abre também perspectivas, quase que neste mesmo contexto, à utilização de parâmetros derivados, em modelos de simulação, estes naturalmente especificamente voltados para formulação de políticas alternativas de desenvolvimento.

Mas não é só neste campo que a correlação canônica pode contribuir para a melhor compreensão de processos espaciais; ela é igualmente promissora no terreno da combinação de aspectos físicos do meio-ambiente, com os relativos à estrutura agrária, onde as características físicas do território podem ser conjugadas aos aspectos socioeconômicos estruturadores da atividade agrária. Este é um campo novo, inteiramente inexplorado, mas obviamente fértil em resultados analíticos.

(1) Um economista definiria o mesmo problema como sendo de organização espacial da economia, reduzindo a diferença entre um ângulo de análise e outro à consideração do espaço como variável endógena ou exógena no modelo.

## BIBLIOGRAFIA

- 1) NYSTUEN e Dacey — A graph-theoretic Interpretation of Nodal Regions, in *Spatial Analysis*, Edit. B. J. L. Berry e D. F. Marble, Prentice Haall Inc., Englewood Cliffs, N. Jersey, 1968, 407 a 418.
- 2) BROWN, Lawrence — Functional distance: an operational approach — *Geographical Analysis* — n.º 1, volume II. The Ohio State University Press — Jan. 1970. 76-83.
- 3) BERRY, J. L. Brian — Essay on Commodity Flows in India, *Research Paper da Universidade de Chicago*, n.
- 4) BERRY, J. L. Brian — “A synthesis of formal and functional regions using a Field Theory of Spatial Behaviour”, in *Spatial Analysis — A reader in Statistical Geography*, Ed. por Brian Berry e Duane Marble, 1968.
- 5) HOTELLING, H. — The most predicable criterion, *Journal of Educational Psychology*, XXVI (1935), 139-42.
- 6) BERRY, J. L. Brian — Essays on Commodity Flows and the Spatial Structure of the Indian Economy, Chicago: University of Chicago, *Research Paper* n. 11, 1966.
- 7) GAUTHIER, Howard — Transportation and the Growth of the São Paulo Economy, *Journal of the Regional Science*, VIII (Summer, 1968) 77-94.
- 8) RAY, D. Michael and LOHNES, Paul R. — “Canonical Correlation in Geographical Analysis” — Apresentado na Reunião de Métodos Quantitativos da União Geográfica Internacional, Rio de Janeiro, abril de 1971 e Ray, D. Michael — The spatial Inter-relationships of Economic and Cultural Differences: A canonical Ecology of Canadá, *Economic Geography*, Special Issue on Factorial Ecology.
- 9) COOLEX, William W. e LOHNES, Paul R. *Multivariate Data Analysis*, John Wiley & Sons, Inc., 1971.

## SUMMARY

This paper is proposing to present an analysis model of spatial organization of a country, aiming the economic planning which, starting from a dichotomic conception of spatial process (formal region and functional region), could be of explicative and predictive character.

In its first part the A. has formed his opinion about the matter used in this article, trying to give it better methodological consistence. The regionalization is defined by attributes and relations, as production and distribution, that are fundamental process of economy. These two concepts, establishing inter-regional relations among structures and fluxes, are important to understand the regional development. And finally, being possible to observe the concrete interrelations in two discrete points between the two subsystems, would be useful for planning.

Then the A. follows discussing the model in its theoretic basis, that is the essence of the work. He presents several contributions, as from Nystuen and Dacey to Bryan Berry, passing by Brown, Gauthier and Michael Ray, in order to place the example which he presents as being of exploratory character and simplified, seeking rather to indicate the validity of the model than to present empiric results. He uses the functional distance as dependent variable (behaviour matrix) and eight independent variables to define the attributes of places (structure matrix). It is implicit in the model that the choice of relevant variables is essential for the success of its application.

As conclusion it becomes clear that the conception of the regionalization process described by two different matrices is important to the regional geographical theory, since it is possible the application of methods as of *canonic correlation*, which combines a causal model (regression) with other one of interrelations (factor analysis), having, for this reason, great importance for planning.

Versão de Joaquim Quadros Franca.

## RESUMÉ

L'article se propose de présenter un modèle d'analyse de l'organisation spatiale d'un pays pour être utilisé comme plan économique, qui partant d'une conception dicotomique du processus spatial (Région formelle et Région fonctionnelle) soit de caractère explicatif et prédictif.

Dans la première partie il définit, pour une meilleure consistence méthodologique, les thèmes employés dans l'article. La régionalisation est définie par attributs et par relations, processus essentiels de l'économie: la production et la distribution. Ces deux concepts établissant des relations inter-régionales de structure et flux, sont importants pour comprendre le développement régional. Et finalement pouvant observer en deux points discrets les interrelations concrètes entre les deux sous-systèmes, c'est utile pour dresser des plans.

Ensuite l'auteur passe à discuter le modèle dans sa base théorique que est le cerne du travail. Il présente les diverses contributions, de Nystuen et Dacey à Bryan Berry en passant par Brown, Gauthier et Michael Ray, de façon à situer l'exemple, qu'il présente comme étant de caractère exploratoire et simplifié, cherchant plus à indiquer la validité du modèle que de présenter des résultats empiriques. Il utilise la distance fonctionnelle comme variable dépendante (matrice de contenant) et 8 variables indépendantes pour définir les attributs des endroits (structure de base). C'est compris dans le modèle que le choix des variables pertinentes est primordial pour que son application ait une bonne issue.

Comme conclusion, c'est clair que la conception du processus de régionalisation décrit par deux matrices différentes est importante pour la théorie géographiques régionale, vu qu'elle rend possible l'application de méthodes comme celle de la *correlation canonique* qui combine un modèle de cause (régression) avec un autre d'interrelations (analyse factoriale) pouvant pour cela même valoir une grande utilité pour la planification.

Versão de Maria Cecília Bandeira de Mello.