WORKING PAPER 249

# Aspects of catastrophe theory and bifurcation theory in regional science.

A.G. Wilson

(Presented to the European Congress, Regional Science Association, London, August, 1979)

School of Geography,
University of Leeds,
LEEDS LS2 9JT,
ENGLAND.

July 1979

## 1. Introduction

Catastrophe theory and bifurcation theory have started to play important roles in regional science. The objectives of this paper are, first, to review briefly the progress made to date; and, secondly, to articulate the nature of the ongoing research tasks. The examples used will be restricted mainly to models based on spatial interaction concepts, though it is usually easy to see how the general argument can be extended. This also has the advantage that it demonstrates how familiar models have hitherto unnoticed properties which come to light when new methods are brought to bear on them.

There are now a number of texts on catastrophe theory - for example, those by Thom (1975), Zeeman (1977) and Poston and Stewart (1978) - and also on bifurcation theory more generally - for example, Hirsch and Smale (1974). The relevance of such methods for regional science has been set out by this author and others elsewhere in some detail - for example, Amson (1975), Wilson (1976), Poston and Wilson (1977) and Wilson (1978). This broad background will be taken as given.

Catastrophe theory is concerned with systems described by state variables, $\underline{x}$, and parameters, $\underline{u}$ with equilibrium states resulting from

$$\max_{\{\underline{x}\}} \ f(\underline{x},\underline{u}) \tag{1}$$

for some function, f. Thom (1975) has shown that, when there are four or less parameters, the equilibrium surfaces in $(\underline{x},\underline{u})$ space can always be transformed into certain characteristic topological forms. The main new feature of the theory is that the function f can have singular points at which equilibrium states merge or disappear. The values of the parameters at these points are said to be critical, and new forms of system behaviour can be observed there.

The equilibrium surface generated by (1) is the solution of

$$\frac{\partial f}{\partial \underline{x}} = 0 \tag{2}$$

The system is assumed to move rapidly back to equilibrium, if disturbed, according to equations such as

$$\underline{\dot{x}} = - \frac{\partial f}{\partial \underline{x}} \qquad (3)$$

and for this reason such systems are known as gradient systems.

The main new types of behaviour generated can be characterised as concerned with (i) jumps, (ii) hysteresis and (iii) divergence. In the examples presented here, we will be mostly concerned with jumps: for a small and smooth change of parameters, $\underline{u}$, through a critical point, the state variables, $\underline{x}$, *jump* to a new state..

Clearly (3) is not a general dynamical system. More generally, we would have

$$\underline{\dot{x}} = \underline{g}(\underline{x},\underline{u}) \qquad (4)$$

for a vector of functions, $\underline{g}$. The equilibrium points are the solutions of

$$\underline{g}(\underline{x},\underline{u}) = 0 \qquad (5)$$

and these, when they exist, determine much of the nature of the solution of the differential equations. These solutions usually take one of a number of well-known forms. For example, there may be convergent (ie. stable) or divergent (non-stable) equilibrium points, or they may be periodic (closed orbit or limit cycle), or chaotic. Bifurcation theory is concerned with the transition from one *type* of solution to another, again at critical parameter values. There is also one other kind of bifurcation - separatrix crossing - which involves a jump between alternative stable equilibrium points.

To summarise: we are focussing on dynamical systems described by eitherequations (1)-(3) or equations (4) and (5) and on critical parameter values at which the mode of behaviour of the system changes in a fundamental way (ie. bifurcation). In the examples below, we will be concerned with locational variables and the appearance or disappearance of locational structures.

Which characteristics of systems generate bifurcation? Essentially, there are two: first, *interdependence* between variables, which generates simultaneous equations - eg. for equilibrium points of the form (2) or (5); and, secondly, *nonlinearities*. The latter generate multiple solutions, or different types of solutions, for different parameter values and the transitions between different equilibrium points on different types of solution take place at at critical points. Since we focus, in part, on the more unusual properties of equilibrium points, this can be seen as providing a new interest in comparative statics.

It is clear that these methods may offer a substantial contribution to systems analysis. They also offer, potentially, a significant new focus for planning. If critical points in parameter space exist, then it is important that planners should know this and where they are, either to 'encourage' the transition or to avoid it.

Finally, in this introduction, we outline the modes of use of the new theory. First, it is possible to attempt to apply the mathematical theorems directly. In catastrophe theory, for a particular problem, this involves making assumptions about the form of the equilibrium surface. The problem is that the canonical forms are well-known, but in a particular case, the system is unlikely to be in such a form. Thus, complicated transformations need to be specified. It is the over-simplified application of catastrophe theory, without the transformation being made explicit (and possibly infringing the assumptions about the *local* nature of the theory) which has generated much of the criticism (for example, in Zahler and Sussman, 1977). In the more general bifurcation theory, the mathematical results are more likely to be applicable in a direct way. The mode of use to be adopted here, and which is recommended, is that the theory is used in the first instance to make the analyst aware of the new kinds of behaviour which may be implicit in his models, but to continue to build detailed models of mechanisms of change. The experience of pure mathematicians, or of workers in other fields, can then often be used as the basis of insights, but theorems such as Thom's classification theorem are not used directly.

In the rest of the paper, first a well-worked out example is presented briefly to fix ideas, though in a wider setting than hitherto, including residential location, and then two further extensions. One extension arises out of interacting submodels, the second out of disaggregation and the introduction of new non-linearities. Some concluding comments, mainly about research tasks, are made at the end of the paper.

## 2. The singly-constrained spatial-interaction model

### 2.1 Retailing systems

The use of the spatial interaction model given by

$$S_{ij} = A_i e_i P_i W_j^{\alpha} e^{-\beta c_{ij}} \tag{6}$$

with

$$A_i = 1/\sum_j W_j^{\alpha} e^{-\beta c_{ij}} \tag{7}$$

is very familiar. The variables are defined as follows:

$S_{ij}$ : the flow of retail sales between zones i and j;

$e_i$ : the per capita expenditure by residents of zone i;

$P_i$ : the population of zone i;

$W_j$ : the size of the shopping centre in j, used as a measure of attractiveness;

$c_{ij}$ : the cost or distance of travel between i and j;

$\alpha, \beta$ : parameters.

The usual mode of use of the model is to feed in exogenous values of the variables on the right hand side of (6), including the $W_j$, compute the $S_{ij}$, and then calculate such terms as

$$D_j = \sum_i S_{ij} \tag{8}$$

This is the revenue attracted to j by this particular disposition of centres, $W_j$. It is then possible, for example, to plan an improved set of $W_j$'s by manual adjustment.

However, suppose we now add further equations which represent hypotheses which determine $W_j$ endogenously. One way of doing this is to assume that $W_j$ grows or declines according to profitability:

$$\dot{W}_j = \varepsilon(D_j - kW_j) \tag{9}$$

for suitable constants $\varepsilon$ and k. The equilibrium conditions are

$$D_j = kW_j \tag{10}$$

Then, equations (6), (7), (8) and (10) constitute a set of equilibrium equations (analogous to (1) or (2) in fact, though the form is different and thus will not be shown in detail here) which determine $W_j$ as a function of parameters like $\alpha, \beta$ and k, say, with all the other 'variables' assumed given and fixed.

We can then carry out an analysis of the dynamics of these equilibrium points. This is done in detail elsewhere, together with a more detailed list of relevant references (Harris and Wilson, 1978; Wilson and Clarke, 1979) and only the bare essence of the method will be presented here. Equations (8) and (10) give different methods of computing $D_j$ and, substituting from appropriate equations, these can be written as

$$D_j^{(1)}(W_j) = \sum_i \frac{e_i P_i W_j^{\alpha} e^{-\beta c_{ij}}}{\sum_k W_k^{\alpha} e^{-\beta c_{ik}}} \tag{11}$$

and

$$D_j^{(2)}(W_j) = kW_j \tag{12}$$

where the two $D_j$'s are shown as functions of $W_j$. The equilibrium condition is then

$$D_j^{(1)}(W_j) = D_j^{(2)}(W_j) \tag{13}$$

and equation (11) and (12) show (13) to be a set of non-linear simultaneous equations in all the $W_k$'s. Because both non-linearities and interdependencies are present, we can expect to find bifurcation properties. The form of equations (11), (12) and (13) enable us to gain some geometrical insights into this: $D_j^{(1)}$ can be plotted against $W_j$; $D_j^{(2)}$ against $W_j$; and the equilibrium points are the intersections of the two curves. Examples, for $\alpha>1$ and for three different k values, are shown in Figure 1*. In case (a), there is only one equilibrium point; in case (c), there are two stable equilibrium points and one unstable. Case (b) is clearly a critical case. The varying parameter here is the slope of the $D_j^{(2)}$ line, k. If the stable equilibrium points are plotted against k, we obtain

*The proof that the curves take these forms are presented elsewhere (Harris and Wilson, 1978). The $D_j^{(1)}$ curve could have more than one point of inflection, but this is unlikely and we restrict ourselves to the curves shown in this analysis.

Figure 2. This is immediately recognised as a form of the fold catastrophe. The critical value of k is zone dependent and may be called $k_j^{crit}$. Thus, the analysis shows that, for each zone, either retail development is not possible (case (a)), or it is (case (c)). This demonstrates how a catastrophe theory phenomenon can be generated directly by an extension of a familiar model.

The insight gained so far is valuable - and indeed this mechanism occurs in a number of other contexts and fields (Boddington, Gray and Wake, 1977; Poston and Stewart, 1978; Papageorgiou, 1979; and Varian, 1979, for example). However, much of the research interest lies in further complexities. For example, it can be shown that when the equilibrium state $\{W_j\}$ is critical with respect to k, it is also critical with respect to all other parameters. In the first instance, this means with respect to $\alpha$ and $\beta$ in addition to k. Also, it can be shown that this form of jump behaviour (for changing k) only exists when $\alpha>1$. Thus $\alpha=1$ is a critical point of a cusp kind and k and $\alpha$ generate a cusp surface, though not in canonical form. When $\beta$ is added, Thom (1975), in effect warns us that the complexity of the singularities may be up to the swallowtail type.

However, the argument can and must be extended further. A critical equilibrium point is also critical with respect to variables we considered fixed: $e_i, P_i, c_{ik}, W_k$, for all i and k for example. Changes in these can also cause jumps. The dependence on $W_k$, $k\neq j$, is particularly difficult to handle because all these other zones obviously have jump mechanisms analogous to that set out for zone j above. The nature of the complexities arising are discussed by Wilson and Clarke (1979).

It can be shown that there is a surface in parameter space - say $(\alpha,\beta,k)$-space - on which the equilibrium state is critical. The elucidation of the shape of this surface remains an important ongoing research task.

## 2.2 Residential location

The simplest form of residential location model has people being located around workplaces, say as

$$T_{ij} = B_j W_i^{res} E_j e^{-\mu c_{ij}} \tag{14}$$

where $W_i^{res}$ is the residential attractiveness factor, $T_{ij}$ is the number of people who work in j and live in i, $E_j$ is the number of jobs in j, $c_{ij}$ is the usual generalised cost of travel from i to j and $\mu$ is a suitable parameter. The balancing factor, $B_j$, is given in the usual way as

$$B_j = 1/\sum_k W_k^{res} e^{-\mu c_{kj}} \tag{15}$$

There are a number of different features, clearly, between residential location and retail location, but we can make some interesting points if we pursue an exactly analogous argument in the first instance. First, we need a measure of residential structure and we taken number of houses in i, $H_i$ and we taken $H_i^{\alpha}$ to be the measure of attractiveness, $W_i^{res}$. This topic should obviously be dealt with in a much more complicated way, but we reserve that for later and for elsewhere. Then, an analogous differential equation for $H_i$ would be

$$\dot{H}_i = \rho(P_i - qH_i) \tag{16}$$

where $\rho$ and q are suitable constants. $P_i$ functions here as the demand for housing at i and q something like the cost per house. $P_i$ is given by

$$P_i = \sum_j T_{ij} \tag{17}$$

and so this incorporates the effect of competition of other places in exactly the same way as happens for the shopping centres. The equilibrium conditions are

$$P_i = qH_i \tag{18}$$

An exactly analogous analysis to the shoping centre case would then show some zones where housing development would take place, and some where not.

However, residential location is clearly not exactly analogous to retail facility location. A number of complicating factors can be identified which we discuss briefly. Consider again the differential equations (16). The parameter q, for example, represents the cost of housing and will be a function of location, i - and a very complicated function at that. It will itself depend on things like density and hence on $H_i$, and this will introduce new non-linearities into the equation.

In equation (16), we took $H_i$ as a measure of $W_i^{res}$. This is one field where there is scope for a great deal of research. It is certainly the case that residential attractiveness is made up of a large number of factors, say of the form:

$$W_i^{res} = X_{1i}^{\gamma_1} X_{2i}^{\gamma_2} X_{3i}^{\gamma_3} \ldots \qquad (19)$$

where the $\gamma$'s are suitable parameters. Some of the $X_{ki}$'s will then be the variables which appear directly in differential equations. For example, one might be $H_i$, another might be $L_i^{res}$ - the amount of land in i taken up by residential development. Some of the $X_{ki}$'s may be 'memory' variables which keep a record of earlier developments of a kind which influence future attractiveness. Examples of these might be density and amount of open space. Memory variables of this type are constructed from other variables which do appear in differential equations or from intrinsic characteristics of zones. Then, an outcome of this kind of research would be to replace equations (16) by equations in a wider range of variables which took their place in a composite attractiveness factor.

There is one very special problem with residential location: there is an obvious bound on the consumption of land for residential development in a zone. This is not recognised in equation (16) at present. There are various ways of coping with this. Firstly, we could do more work on the term q and make it, in effect, a complicated production function of other variables. For example, when $H_i$ rises to a point where densities are so high that further building becomes prohibitively expensive, then this fact would be registered by an increasing value of q and $\dot{H}_i$ would become negative in equation (16). An alternative procedure (which is in effect an approximate form of the previous method) would be to take a constant or simple value of q up to a certain value of $H_i$ which represented a maximum density and then let it be zero after that. This sort of method might work well in a simulation model.

There are also other complexities peculiar to residential dynamics. First, we know that there must be a lot of inertia in residential development: once a particular house or group of houses is built, then unless something very unusual happens, it has a life of perhaps sixty years and often very much longer. Further, the development is often lumpy: houses are not typically built singly in such a way that a continuous variable differential equation might be good representation. This suggests that we have to look much more to simulation models in which this process can be represented directly. There is also an accounting problem which differentiates residential modelling from retailing: the structures, the houses, last much longer than the assignment of people to them. This means that the dynamic assignment of people to the stock should also be represented by difference equations rather than the allocation equations (14) and (15). This yet again will lead to the introduction of new non-linearities and new bifurcation properties. The kinds of accounting frameworks involved have been presented elsewhere (Wilson, 1974, Chapter 11).

Finally, it may be interesting to try new tacks altogether for residential modelling and not to work directly by analogy with the retailing case. A good example of this is the work of the Brussels school (Allen, et al., 1978). The essence of this is to use the population variables directly in the differential equations rather than to model structure variables. Perhaps the main difficulty with this method is that the structural variables, as we noted in another context, change much more slowly than some of the population variables and this could lead to problems if they are the internal variables of a 'fast' equation and are themselves changing. In effect, some assumption is needed rather like that of the original Lowry model: that residential structural development takes place in response to, and always balances with, population change by location. However, there are clearly alternative possibilities to explore.

## 3. Interacting fields

The retail model can be seen as a set of fields around expenditure points, $e_iP_i$, which determine retail structures, $W_j$. These centres will be the sites of retail (and all service) employment and thus will, partly, determine $E_j$. The residential model is a set of fields around employment points, $E_j$, and they determine $P_i$. Thus $E_j$ is partly determined by the $P_i$'s and each $P_i$ determined by the $E_j$'s. When the two submodels of the preceding subsection are combined into something like a Lowry model framework – with the addition of any minimum size or density constraints as required, together with the usual equation

$$E_j = E_j^B + E_j^R \tag{20}$$

(where $E_j^B$ and $E_j^R$ are basic and retail employment respectively) – then it is clear that there is a strong coupling between the two subsystems. This can be seen as the addition of interdependencies rather than non-linearities but will affect the positions of equilibrium points. It will also affect bifurcation behaviour, at least in the following way: any jumps in a $W_j$ will directly cause jumps in each $P_i$, and vice versa. Such a secondary jump may well take the variable concerned through a critical point and cause further jumps.(which is not unlike the 'domino effect' of Isard and Liossatos, 1978). These will be difficult phenomena to interpret, but this is an important research task.

## 4. Disaggregation

We now show how new non-linearities can be introduced. Consider a two sector disaggregation of the retail model, using an obvious notation:

$$S_{ij}^g = A_i^g e_i^g P_i \hat{W}_j^g e^{-\beta^g c_{ij}} \qquad g=1,2 \tag{21}$$

with

$$A_i^g = 1/\sum_j \hat{W}_j^g e^{-\beta^g c_{ij}} \qquad g=1,2 \tag{22}$$

The attractiveness of a place for a particular good, g, might depend on the size of retail facilities in total – say

$$W^*_j = W^1_j + W^2_j \tag{23}$$

so that

$$\hat{W}^1_j = W^{*\alpha^1_1}_j W^{1\alpha^1_2}_j \tag{24}$$

$$\hat{W}^2_j = W^{*\alpha^2_1}_j W^{2\alpha^2_2}_j \tag{25}$$

for suitable parameters $\alpha^1_1, \alpha^1_2, \alpha^2_1, \alpha^2_2$. The balancing conditions would apply in each sector:

$$D^1_j = k_1 W^1_j \tag{26}$$

$$D^2_j = k_2 W^2_j \tag{27}$$

and the dynamical analysis can then be carried through as before.

It can be shown that the results are similar to the aggregate ones but with $\alpha^1_2$ and $\alpha^2_2$ playing the role of $\alpha$ for sectors 1 and 2 respectively. That is, if either is greater than 1, jumps are possible; otherwise, not.

Let

$$\alpha^1_1 + \alpha^1_2 = \alpha^{(1)} \tag{28}$$

$$\alpha^2_1 + \alpha^2_2 = \alpha^{(2)} \tag{29}$$

Then if sector 1 is taken as consumer goods and sector 2 as durable goods, we might expect

$$\alpha^{(2)} >> \alpha^{(1)} \tag{30}$$

and for $\alpha^{(1)}$ to be of the order of 1. This would suggest that $\alpha^1_2$ is likely to be less than 1 and $\alpha^2_2$ *possibly* greater than 1. Generally, we would be more likely to have jumps for higher order goods. The argument presented suggests that, the more complicated the system becomes, the less likely is jump behaviour to occur, at least for lower order sectors.

## 5. Concluding comments

The main general conclusion to be drawn from this essay is that the techniques offered by catastrophe theory and bifurcation theory are likely to be very important for regional science and for planning whenever non-linearities and interdependencies are present in a dynamic or equilibrium framework. Indeed, the main purpose of this paper has been to illustrate how the case studies which have already been partially worked out can be extended. It is important to recognise that the models presented here are relatively simple and the world correspondingly more complex. The next obvious step in relation to this paper, for example, is to incorporate the principles of disaggregation, probably using a hierarchical framework, within an interacting fields' model. This can then be seen as a contribution to central place theory (cf. Wilson, 1978). From the methods' side of the argument, only the simplest of the mathematical techniques which are available have yet been used. On both counts, therefore, the ongoing analytical task is formidable.

It is also important, though difficult, to seek to develop some associated empirical work, partly to test the models and ideas, partly because the use of these methods in planning will not be convincing otherwise. For once, a new idea, while immediately applicable in principle with its focus on criticality and stability, does not seem to be immediately applicable in practice. But the promise is substantial.

## References

P.M. Allen, J.L. Deneubourg, M. Sanglier, F. Boon and A. de Palma (1978) The dynamics of urban evolution, Volume 1: interurban evolution, Volume 2: intraurban evolution. Final Report to the U.S. Department of Transportation, Washington D.C.

J.C. Amson (1975) Catastrophe theory: a contribution to the study of urban problems? *Environment and Planning*, *B*, 2, 177-221.

T. Boddington, P. Gray and G.C. Wake (1977) Criteria for thermal explosions with and without reactant consumption. *Proceedings of the Royal Society of London*, *A*, *357*, 403-22.

B. Harris and A.G. Wilson (1978) Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models. *Environment and Planning, A, 10*, 371-88.

M.W. Hirsch and S. Smale (1974) *Differential equations, dynamical systems and linear algebra.* Academic Press, New York.

W. Isard and P. Liossatos (1978) A simplistic multiple growth model. *Papers, Regional Science Association, 41*, 7-13.

G. Papageorgiou (1979) Private communication.

T. Poston and I.N. Stewart (1978) *Catastrophe theory and its applications.* Pitman, London.

T. Poston and A.G. Wilson (1977) Facility size vs. distance travelled: urban services and the fold catastrophe. *Environment and Planning, A, 9*, 681-6.

R. Thom (1975) *Structural stability and morphogenesis.* Benjamin-Addison Wesley, New York.

H.L. Varian (1979) Catastrophe theory and the business cycle. *Economic Inquiry, 17*, 14-28.

A.G. Wilson (1974) *Urban and regional models in geography and planning.* John Wiley, Chichester and New York.

A.G. Wilson (1976) Catastrophe theory and urban modelling: an application to modal choice. *Environment and Planning, A, 8*, 351-6.

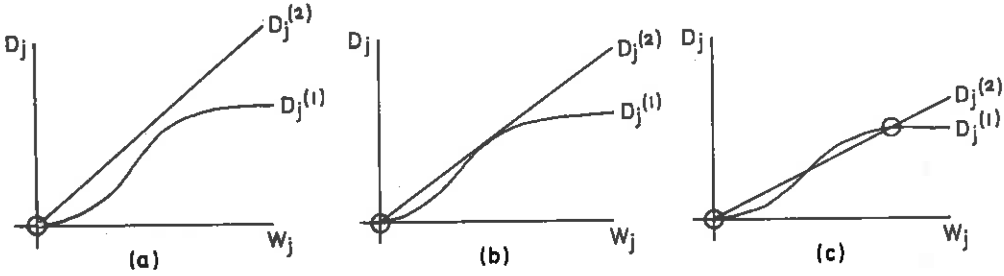
A.G. Wilson (1978) Spatial interaction and settlement structure: towards an explicit central place theory. In A. Karlqvist, L. Lundqvist, F. Snickars and J.W. Weibull (Eds.) *Spatial interaction theory and planning models.* North Holland, Amsterdam, 137-56.

A.G. Wilson and M. Clarke (1979) Some illustrations of catastrophe theory applied to urban retailing structure. In M. Breheny (Ed.) *London Papers in Regional Science.* Pion, London.

R.S. Zahler and H.J. Sussman (1977) Claims and accomplishments of applied catastrophe theory. *Nature, 269*, 759-63.

E.C. Zeeman (1977) *Catastrophe theory.* Addison-Wesley, Reading, Mass.

Figure 1.




○ = stable equilibrium points

Figure 2.