A Neural Network Approach to Modeling Inter- and Intra-State Conflict

Paul R. Williamson
Global Vision, Inc., P.O. Box 4394
Ann Arbor, MI 48106-4394
telephone: 734-769-4877
paulrw@globechange.org

Myron S. Karasik
Member, Institute of Electrical and Electronic Engineers
ColdFrame Inc.
530 Lawrence Expressway, Unit 354
Sunnyvale, CA 94085-4014
telephone: 408-705-6870
coldframe@eudoramail.com

Prepared for a meeting of the Peace Science Society, International
Urbana, IL, November 6, 1994; author information updated February 23, 2001

1. Introduction to the Problem

A seldom addressed but important unsolved problem in inter- and intra-state conflict research is whether it is possible to use numerical data processed by electronic computer to predict the onset of disputes and wars over time with any greater reliability or specificity, or any further into the future than what is already done by expert human practitioners (the latter using a combination of intuition and existing analytical methods, however poor or good that may be). Our purpose in this discussion is to outline one plausible, empirically based approach to finding an answer. We will dynamically model conflict using an approach drawn from the area of artificial neural network modeling. The actual use described here, of this approach, has not been carried out; however we hope such a study and its results will soon follow.

Since we have no need to discuss the original biological type of neural network, herein we drop the phrase “artificial”. In this discussion we focus on inter-state conflict; however the technique of neural network modeling is applicable to intra-state conflict as well and we indicate very briefly how that application might work. In the next section we begin with a brief discussion of dynamic modeling and of some problems which have attended its application to human societal conflict. Following that, we introduce the concept of neural networks and indicate its possible application to our modeling topic. Then we discuss what data are appropriate and, of those, which are presently available for use; after which, in the next two sections, we discuss our proposed modeling approach and confront some limits to achieving a fully dynamic model. Finally, in a concluding section, we give an illustrative example of a few specific variables treated using the approach that we propose.

2. Dynamic Modeling

In attempting to use numerical models to predict conflict or other types of societal behavior, many complications arise concerning what form of model shall be tested. It is important to see what three of these complications are since they often have appeared to raise significant barriers to modeling, and because the approach that we outline appears to ease all of them. To see what they are, we need to formulate more precisely what we mean by “prediction.” Suppose one knows the values of N many variables of interest $x_i(t)$, $i = 1, \ldots N$ at a referent time t (where t refers, say, to a given year). And suppose one knows how to compute, *entirely* as a function of one or more of the original values themselves, the changes $\Delta x_i(t)$ in each such value from that time to the later time (say the next year) $t+1$. Then one *also* can know the values of each of the variables at the later time: they are given by

$$x_i(t+1) = x_i(t) + \Delta x_i(t) \qquad (1).$$

In that event we would say that one can “predict” from the one time period to the next. It is in this sense that we use the term prediction.

By a “dynamic model” we mean a situation of the above type, where one has the recipe to get from the one time period to the next period, in the form of a function for the changes. (The model may give the differential form dx_i / dt from which the changes Δx_i are then estimated.) In inter-state conflict research, a well known example of such a dynamic model is the arms race model proposed by Richardson (1960), based on a set of differential equations in which the rate of change of each armament variable is a function of itself and all other such variables.

In a dynamic model, the variables x_i which go into the change recipe are called *state variables*. Since the changes use only the original values, the whole process is self-contained. Using equation (1), one “bootstraps” from period t values to period $t+1$ values; one then uses the $t+1$ values to compute the changes for that period, then to bootstrap to the values of the next period; and so on. In this manner, starting in period $t=1$ and using the information for that period only, it is possible to compute the values for subsequent periods 2, 3, 4, ••• into the future. *If* the initial values were completely accurate and the change functions were correct, then all the future computations would be completely accurate also; but realistically, the initial values will contain errors which affect the accuracy of all subsequent calculations. Probably too, no change function would be entirely accurate. However, for some finite number of initial periods, the estimated future values might turn out to be of imperfect but tolerable accuracy. For weather prediction, that period of tolerable accuracy is a few days; for the solar system, it may be on the order of 100 million years (Murray 1993, p. 106.). How long it might be for global social phenomena we think should be determined by empirical study.

Let us consider the first of the three complications to which we just alluded. Quite likely, we do not know exactly, or even approximately, what the change function ought to look like, or there may be a great many candidates among which it is difficult to choose; so we could spend a long time trying to find the “correct” one. One aspect of the variety of choices presented by the candidates deserves special mention. In the standard dynamic modeling approach referred to above, system change at time t depends only on the values of the variables at that moment (in which case it may be said that the system is assumed to have no memory). Alternatively, system change at time t might well depend on values of variables at present but also at previous times $t-1$, $t-2$, ••• $t-n$, •••, indefinitely far into the past. (Then one could say the system did have a memory.) So the question is, how shall one address this multiplicity of possible schemes?

The second complication is, the plausible change functions include those which are non-linear, for which usually there is no general solution algorithm (such as “least squares”, etc. that exist for linear functions), thus we have no assurance of being able to use observed data to infer the parameters that best define the function. (As we intimated two paragraphs above, non-linearity may also give rise to chaotic phenomena which defeat all subsequent predictive attempts after finitely many periods, but how soon this happens should be answered by empirical means.)

Finally is the added practical difficulty of accommodating large numbers of state variables, especially when they may involve many disparate functional relationships. Multiple indicators may also complicate standard methods of estimating models to the extent that the latter require questionable statistical assumptions such as sufficient degrees of freedom, non-correlation between independent variables and residuals, etc.

While a good part of the choice thus is simply in the quite considerable number of plausible attribute and behavioral indicators available (see the Appendix), again let us call attention to a particular aspect: Suppose we have a collection of indicators we believe relate to system change, where one of them is denoted by x_1 . Now the question arising is, do we really mean x_1 is the state variable or do we mean its change Δx_1 instead? That is, perhaps the effective relationship is between the amount by which the original variable *changed* during the period as state variable contributing to *its* change, the second order change $\Delta^2 x_1$, as the dependent variable. Going in the other direction, perhaps the *cumulation* of x_1 over time is the appropriate state variable, with x_1 the dependent change variable. (Given the latter possibility, the state variable at referent time t would be expressed as $\int_a^t x_1(s)ds$, where a denotes a constant fiducial time. For a wide class of functions, higher and lower ordered levels of this hierarchy, such as the third order change being dependent on its second order change as state variable, may be disregarded because these levels already are implicitly contained in the three named possibilities; Lynch and Truxal 1962.) Corresponding state variable questions arise for many candidate indicators.

The profusion of dozens of variants on the original Richardson arms race equations (Etcheson 1989) illustrates the first complication; the plausibility that the fatigue, reaction, and grievance coefficients are, themselves, varying (and unknown) functions of the armaments levels illustrates the second complication. While Richardson's equations explicitly involved only armaments levels, a more plausible realistic model would involve many other variables describing the interacting parties, their relationships mutually and with other parties, and the history of all such factors. The difficulty of integrating very many such variables without there resulting an unmanageable model, illustrates the third complication. Furthermore, the dilemma, whether a given indicator, its change, or its cumulation is the appropriate state variable of a dynamic system is suggested by the distinction, in international conflict research literature, concerning the *magnitude* of differences in national material capabilities versus *rates of change* in them, as correlates of international war, and the question which of these choices is the appropriate indicator (Houweling and Siccama 1988, Organski and Kugler 1980, Wayman 1989).

Neural network modeling has characteristics which help circumvent these complications about choosing the form of the function, finding the correct parameters of nonlinear functions, and integrating a large number of variables. The circumvention is achieved through a process in which such a model uses empirical examples to converge through iteration toward suitable specifications of these factors. As in all empirical testing, many cases are better than few; and there are dangers of false generalization from too few cases. However this process works on any number of cases and allows all candidate state variables to be used at once, nor does it assume anything concerning mutual correlations among the latter. And the process allows one to introduce inputs from many past time periods at once and to calibrate the parameters of the model in multiple passes, in each one adding additional past years to search for a satisfactory tradeoff between parsimony and accuracy.

Our approach to predicting inter- and intra-state conflict will make use of such characteristics and process. Again, our motivation for such an approach is the belief that this method can help overcome the indicated difficulties of explicit dynamic modeling. In addition, we will relax the definition of a dynamic model given above by seeking to develop a variant, in which some but not all of the variables are updated within the model. We return to the latter feature after discussing the concept of neural network modeling and the data to which we intend to apply it.

3. Artificial Neural Networks

By neural network is meant any of several schemes like those described in Hammerstrom (1993), Haykin (1994), and Rich and Knight (1991, chapter 18). We intend to use a "back propagation" scheme, composed of "input" and "output layers", and one or more "hidden layers". We will describe this scheme in the case of one hidden layer.

In brief, the idea of a back propagation network is as follows. The input layer (see Figure 1) consists of a series of components, each of which is simply a counter holding the numerical value characterizing one of the pieces of input information. We will return to the identity of these inputs below, after discussing the mechanics of the modeling approach.

The hidden layer consists of a second set of components and a set of weights called synaptic weights. Let j refer to the jth such hidden layer component. Each such component first computes a value $s(j)$ which is a weighted sum of the values of the input layer, minus a constant reference value: if $x(i)$ is the value of the ith input component, $w_1(j,i)$ is the synaptic weight applied to $x(i)$, and $w_1(j,0)$ is the reference value (also considered a synaptic weight), then we define

$$s(j) = \sum_i w_1(j,i)x(i) - w_1(j,0) \ , \tag{2}$$

across all values of i . Second, the jth component computes a new number based on a sigmoid (inverse logistic) function the argument of which is $s(j)$. The value of this new function, defined as

$$h(j) = 1/[1+e^{-s(j)}] \, , \tag{3}$$

evidently approaches 0 as $s(j)$ becomes more strongly negative and 1 as $s(j)$ becomes more strongly positive, and $h(j)=1/2$ for $s(j)=0$. Additional hidden layers would draw on the final computations $h(j)$ of the previous hidden layer in the same way as the first one draws on the values of the input layer.

Finally, the output layer consists of a third (or final) set of components, each of which corresponds to one of the empirically observed change variables, in a manner we will discuss later. Everything described in the hidden layer has a corresponding element in the output layer: A weighted sum defined as

$$y(k) = \sum_j [w_2(k,j)h(j)] - w_2(k,0) \, , \tag{4}$$

is computed across all values of j . Then the sigmoid function

$$o(k) = 1/[1 + e^{-y(k)}] \ , \tag{5}$$

is computed.

In our use of this approach, each of the outputs $o(k)$ corresponds to a "desired" quantity $d(k)$ which is a suitable transformation of some empirically observed quantity. In our application, this observed quantity is the change Δx_k in one of the input variables x_k . For instance, one of the outputs corresponds to a transformation of the annual change in the hazard rate of the dyad in question (Tuma and Hannan 1984; Williamson, Warner and Hopkins 1988). The ideal of the model is to make each output equal its corresponding desired quantity; and any discrepancy between the two supplies a non-zero component to the error vector $e(k) = o(k) - d(k)$. Qualitatively, the criterion of all such models is: find synaptic weights $w_1(j,i)$, $w_1(j,0)$, $w_2(k,j)$, and $w_2(k,0)$ for all i , j and k such that a measure of the variance of $e(k)$ is acceptably small. (A quantitative statement of the criterion is given by Haykin 1994.)

The question then becomes, can one find such numbers? Repeated empirical investigation as well as mathematical analysis suggests that under a very wide set of circumstances the answer is likely to be Yes, provided we are able to subject the model to a process of calibration called "training". In training, the initial synaptic weights are set by some plausible guess (possibly at "random") and the corresponding outputs are compared with the empirically observed values (criterion values) of a known outcome; then the synaptic weights are changed as a function of the error vector of output minus criterion values. When this process is repeated using additional criterion values from other known outcomes, the resulting error vectors typically become smaller. That is, the more outcomes known and used in training the model, the more accurate it will become.

4. Selection of Data

The subject of violent societal conflict is by no means self-contained. Many political, economic, demographic, and other aspects of human society affect, and are affected by, collective or mass human conflict. Beyond the societal aspects, are those more grossly physical aspects such as the traits of individual humans and many components of the non-human environment. In fact, our interest goes to netting all these societal and other elements together in a predictive scheme, in the belief that such a grand scale of integration is possible and desirable (Karasik and Williamson 1994, Williamson and Karasik 1994). Here, however we are concerned with a more narrow context, partly because the topic is intrinsically important and the existing data may be sufficient to allow useful findings, and partly to gain useful experience in methods that later can be applied more broadly.

Moreover, even the topic of societal conflict has a multitude of particulars which one could choose to emphasize. Again, because appropriate data are available and it is interesting and important, in the present discussion we focus on the issue of predicting variations in the hazard of militarized interstate dispute (MID) onsets (Gochman and Maoz 1984, pp.587-589) between named pairs of states (dyads). By "hazard" we mean roughly the rate at which the dispute onset is expected to occur per unit time, given the conditions yielding the hazard. (See Appendix note 1.) To find the total expected rate of onsets, one would multiply the hazard times the number of dyads at risk to dispute or, for different rates, sum across dyads.

An example is shown in Figure 2. Here we see a year-by-year estimate of the hazard of MID onsets for the United States - Soviet Union dyad from 1918 to 1976. This estimate was obtained in a previous analysis of dispute onsets (Williamson, Warner and Hopkins 1988). The dispute data were taken from a Correlates of War project data set which summarized all instances identified at the time the data set was finished, of involvement by nations which were interstate system members (Singer and Small 1972, Small and Singer 1980) during the period 1816 to 1976. (A major revision of this data set, to a more recent year, has since been completed.) Knowing when each dispute involving system members began, as we do from the Correlates of War data set, allows one to generate a corresponding series of hazard rate estimates for every dyad in every year that both parties were system members, from 1816 to the present.

Following reasoning similar to that which Richardson followed about armaments, changes in the hazard of dispute onset might be a function of its own value. In addition, as we suggested a moment ago, various other kinds of information describing the interaction between the parties, their relationships, and the past history of such factors are potentially relevant as well. Fortunately, data which match the broad geographical-temporal domain of the dispute data have also been developed. These data include indicators of the material capabilities and cultural characteristics of the system members; their dates of entry to and exit from the system, and classification to major or minor power status; their mutual diplomatic, alliance, and organizational ties; their involvement in inter-state and civil wars; and the presence or absence of geographical contiguity for each dyad. Additional data have been developed, having the same domain, which describe the political characteristics of the governing regimes of these system members (Gurr 1994). In the Appendix, we indicate in greater detail how we propose to use some of these data in a neural network model of interstate conflict.

5. An Approach to Training a Dispute Onset Model

As we just indicated, there is no shortage of interesting, relevant inputs to a neural network modeling scheme; indeed, these data were originally chosen to be developed exactly for their plausible relevance to the dynamics of inter-state conflict. Given the large number of training cases in the form of many dyad years, we also will be able to test a fair assortment of such variables in alternative combinations.

Nevertheless the geographical-temporal domain of these data has a structure which presents a dilemma. The system membership criterion is sufficiently broad to admit most or all entities that one would likely consider a nation state today; and the data cover these current system members with great thoroughness. As one goes back to earlier years, however, the number of qualifying states gradually declines so that, upon reaching the initial year 1816, nearly all the population has dropped out. Thus a data matrix of entities-variables versus years has a right-triangular shape, with the base corresponding to the most recent year and the apex to 1816. The dilemma is, we would like to make use completely at will, as model inputs, of the data for all entities and for all years, however we cannot do both because of this triangular shape. For instance, to use information involving India as an input, would require that data for it be extended back prior to 1947 when that state became recognized as independent (thus qualifying for system membership). In some cases, such as population, these data are clearly defined even if not readily accessible; in others, such as exchange of diplomatic representatives, it is hardly clear what such a datum would mean in principle. The other alternatives are: confine the analysis to post-1947 years (thus considerably impoverishing its scope), exclude Indian variables as inputs, or do a “multi-stage analysis”.

It is the latter alternative which we favor. By this approach, we mean that a preliminary study first will be done using as inputs information about 1) the characteristics and interactions of the dyad in question, 2) its constituent states, 3) some aggregate characteristics and behavior of the global system, and 4) the same about a handful of states the identities of which go back to the year 1816. The latter group of states will include those parties-- Britain, France, Russia, United States --which have retained their system identities throughout the period and who are judged by expert historical opinion to have held major power status during all or part of it. In addition, this information will include either the same variables for the remaining parties-- Germany, Japan, China, Austria-Hungary, and Italy --which are judged to have held major power as well as sovereign political status during parts but not all of the study period; or perhaps a composite of them. (The latter would be because of the ambiguity of evaluating variables during periods of non-membership. Some have argued that China should be counted as a system member from 1816 rather than from 1860, as at present. We are inclined to agree, but appropriately changed admission criteria have not been applied systematically to the whole data set, nor are the data readily available; so we shall keep it as is.)

Note that the output of a given iteration of the model will be change variables for a *single* dyad over one year; each dyad-year thus is to be modeled using a separate iteration of model inputs and calculations. The training iterations for each dyad will extend as far back as the year following the first MID onset involving it. All such iterations will be used to obtain a single global model the synaptic weights of which are inferred from all dyad years. Finally, at the end of this first stage, inputs will be deleted which appear to make less important contributions to the accuracy of the outputs.

After obtaining a model based on the first stage training, we will go to the second. This second analysis will begin with those inputs surviving the winnowing process of the first stage; and it will start with the synaptic weights previously found. To these elements will be added inputs for which data are unavailable (or undefined) for one or more years 1816 to the present, but are available over the history of the particular dyad in question. For instance, characteristics and behavior of India and Pakistan may each be added to the inputs of the analysis for the other. As before, inputs making the less important contributions will be deleted at the end of the training session. At this juncture, rather than a single global model we now will have trained individual models that are particular to each dyad.

In the third and final stage, we will group dyads together and re-run the training sessions for those groups. Choosing them will reflect that all members of a given group either have been major powers (one group) or (for the other groups) the members will be directly or indirectly connected by geographical contiguity of their borders and / or by a history of mutual dispute or war behavior. (So if states A and B have mutually disputed, and states B and C have a common border, we may put the dyads composed of A, B, and C into one group.) Again, the training will begin with those inputs surviving the winnowing process and with the synaptic weights of the previous stage. To those elements may be added inputs that are particular to the group in question concerning the entire group rather than its members. At this juncture, we would need an altered criterion for the initial year of the group series; an appropriate one might be the first year after which at least one of its dyads has experienced a dispute.

The distinct feature of this final stage is, the outputs of a single iteration will concern *all* group members rather than dyads taken one by one. This will give an opportunity for the model to reflect structural peculiarities in which the various parties have specialized roles within the region of which they are a part.

The question with which we began is whether it is possible to create a numerical simulation that predicts with acceptable accuracy the ups and downs of the hazard rates and other variables? While the approach outlined above cannot guarantee that all possible relationships are detected, certainly it can detect many, given sufficiently many cases on which to operate. How many such cases are needed, depends on how complex a relationship is actually appropriate to the data; this would need to be found by actual trial. However there are a great many cases to draw on. The previous study of the timing of disputes (Williamson, Warner and Hopkins 1988) could draw on an excess of 20 thousand dyad-years of experience, consisting of each pair of nations in each of the years after they first were on opposite sides of a dispute. The recently completed revision of the MID data set resulted in more than double the number of cases, which will correspondingly enlarge the available pool of dyad years. These numbers would be applicable to the first stage study. Later stages, in which distinct dyads and groups are studied separately, would have greatly fewer numbers, in most cases considerably less than the

185 or so years of the whole data series; thus the opportunity of altering the model at these later stages would be very limited. These stages would have the function of correcting the earlier results, where warranted by historical experience, but not of replacing those results.

6. Non-Dynamic Variables in the Model

Earlier, at the end of Section 1, we alluded to an easing of the criteria compared to those required of a fully dynamic model, to which idea we now return. Let us call a variable x_i "endogenous" if its change function (or differential) is computed within the model. To fit our previous definition of a dynamic model, every variable would need to be endogenous. The key to having such a self-contained model, of course, is in being able to compute all the changes using *only* endogenous variables in the input layer. Suppose the changes Δx_i also depended on the values of some other variables y_k, for which one was *not* able to accurately compute the changes from one period to the next? Let us call these latter variables "exogenous". Then the initial period 1 input values would not be sufficient; one would need to wait until the future values of the exogenous variables were observed, before using them to help compute the subsequent round of Δx_i values.

While our ultimate goal is a fully dynamic model, we assume the realistic situation for inter- and intra-state conflict is probably many variables will need to be treated as exogenous, at least at this juncture, thus preventing that goal from being realized. An important source of exogenous variables is the attributes and behaviors that nominally are internal to the individual states. Among such "internal" variables are those which describe intra-state conflicts such as civil wars. To be able to represent the dynamics of such variables requires an entirely separate program of modeling. Such a program may be similar in its use of neural network approaches; however we leave that topic to a future occasion. (Note, however, that if the nation state were itself to be composed of parts, then a similar logic to what we have suggested here for dyads might inform such a modeling effort. In particular, perhaps a civil war can be viewed as an interaction among elements that have formed themselves into two groups or geographic regions-- that is, civil war is evidence that the nominally unitary state has formed itself into a dyad. Compare this with the complimentary suggestion by Wilkinson (1987 and 1989), that interactions including war between the elements of a dyad is evidence that its parts have joined into a single entity.) A related barrier to achieving a fully dynamic model is the omission from consideration of the environmental and other non-societal elements to which we previously alluded.

Still another limitation is in those inputs that are dichotomous, denoting the presence or absence of some characteristic, such as "major power status", or the occurrence or non-occurrence of some event, such as the formation or termination of an alliance. In such cases, to make the characteristic endogenous, we need some means of quantifying, as itself a variable, the underlying propensity or likelihood that the dichotomy will alter its value. The change in such an underlying, continuous variable would then be the corresponding model output. This raises two kinds of problems. First, is representing each such underlying propensity and assigning empirically determined values to it. Second, is the need for a mechanism which translates the propensity into outcomes of change events which actually do or do not occur at each iteration of the model. It may be useful to address the first problem using the same techniques on which the dispute hazard rate has already been based. The second problem can be addressed with a pseudo random number generator; alternatively, it may be possible to use the neural network techniques for generating such dichotomies. In any event, the chief point to be made is that all the possibilities mentioned in this and the previous paragraph involve considerable work which we think it wiser to defer beyond this first modeling attempt.

While the resulting limits are real, it may turn out that many exogenous variables change comparatively regularly or slowly, or with benign impact. For instance, variables reflecting the material capabilities of states, diplomatic relations between them, mutual geographic contiguity or isolation, or other attributes or structural features, may (on most occasions) change much less rapidly than variables describing conflict within and between states. Although likely examples are as yet unclear, it may also turn out that the impact of some exogenous variables, on the rate of change of the endogenous, is comparatively small. In either case, it might then be acceptable to approximate the influence of the exogenous variable by treating its values as constant or as changing according to some simple externally imposed approximation such as a constant rate of change. Identifying such possibilities would be a goal of the model development.

7. Illustrative Example and Conclusions

Finally, let us consider how a modeling program would look, using some of the variables drawn from the appendix and applying the stage 1 through 3 methods described above. As a specific group of system members, consider an “Indian subcontinent group” composed of India, Pakistan, Bangladesh, Sri Lanka, and the Himalayan mountain kingdoms. (Why not also include Afghanistan and Burma, etc.? At this juncture, the choice of the group members is purely judgmental; and certainly alternative combinations could be tried. But for now let us stick with our first choice.) It should also be noted that the variables presented here and in the appendix are, themselves, in some instances rather elaborate transformations of simpler data. In principle, no such “pre-cooking” of the inputs is needed: given sufficiently many cases, a neural network is capable of emulating these functions itself-- provided they in fact are optimal representations of the input-output relationships. Nevertheless, we use such transformations in the hope of speeding up the convergence of the model to accurate representations; and also, to help the result make sense to human intuition.

As indicated for the first step, we train a model consisting of *all* dyads existing in the global system during our data period 1816 to the present. In due course, a number of endogenous variables can be developed as hazard functions of each of the dyadic events, such as diplomatic recognition, cataloged in Group 3 of the appendix; all of which are of the first type of input, characterizing the dyad. At this juncture the one available such output is the estimated hazard of mutual MID onset (henceforth, HAZ). An exogenous dyadic variable would, for instance, be the highest level of hostile behavior exhibited by any participant in the prior dispute involving the present dyad members as antagonists (HLH). An exogenous state characteristic is values of military personnel of the two parties as a fraction of the world total (MP1 and MP2). Because stage 1 development will pass over all dyads, we would seek to assign symmetric roles to the two values; that is, we want their synaptic weights to be equal. (For suppose, in the US-UK dyad, the US value is in “slot A” of the model inputs and the UK value, in “slot B.” Then, in considering all the dyad years, when we get to, say, France-Germany, which value shall be put in which slot?) This desired symmetry can be assured by using their product $\mathrm{MP1} \times \mathrm{MP2}$ at the input layer (among several possible methods).

Note, however , at stage 1 of the present model there is no *endogenous* state characteristic; for to what one of the several dyads to which the given state belongs, should we assign the mechanism for incrementing the value? Here we see the place where a distinct modeling scheme is needed for each state actor, taken individually. While the details of that topic exceed our present purpose, we note in passing that the hazard of civil war and other internal conflict events would be contained in such a scheme. These and other attributes and behaviors might affect, and be affected by, the dyadic variables; that is, each would supply inputs to the other. Thus, in a higher-level global model, the two types of scheme would be linked together.

Turning to global systemic and major power characteristics, an example of the former is material capabilities concentration based on information-theoretic entropy across 8 major powers. (See Appendix note 14. This measure requires the numbers from all the parties to be put into it, so it actually is at the systemic rather than individual level.) This characteristic could be evaluated separately for any of five capability indicators; for our example let us choose urban population (MC-URB). For indicators of the nine separate major powers, let us choose (again) numbers of military personnel relative to the world total (MAJ1, ... MAJ9). We note in this context there is no need for the earlier symmetry, since the values of the nine majors are to be presented to the input layer in every one of the dyad-year examples. It is perfectly reasonable that, say, the United Kingdom, may have a role systematically distinct from the other eight parties. (Indeed, one might test the impact of characteristics of additional system members; however, as we mentioned in section 5, the triangular shape of the data set puts a limit on this.)

Finally, we note that the modifications concerning inputs from multiple past years, and annual cumulation and incremental versions of variables, which we discussed in section 2, would be tried out. The considerable number of dyad-years available as examples (cases) in stage 1 gives plenty of room to do so. For instance, one might try annual increments in relative military personnel in place of, or in addition to, absolute values; and one could input these alternatives from multiple past years.

Moving, in our example to the second and third stage analyses, attention now would shift from the individual dyad level to the “Subcontinent” group level consisting of variables describing India, Pakistan, Bangladesh, Sri Lanka, and the Himalayan mountain kingdoms, and the dyads formed among them. In conducting this stage, we would begin with the results of the first; that is, we would start with the annual change in HAZ for output, as a function of the magnitude of HAZ, and of HLH, MP1 × MP2, MC-URB, and MAJ1 through MAJ9 as inputs. This function would be defined by synaptic weights connecting these inputs to each of several neurons in the hidden layer and, from the latter, to the change in HAZ, as determined in the global first stage analysis.

In the second stage, to these data would be added characteristics specific to the group members that we have chosen. For instance, we might include the relative military personnel of each group member (MP1, MP2, etc.) Since, at this juncture, the state or dyadic identity of each “slot” is retained over all the cases, just as for the major power characteristics, these group variables may be asymmetric, i.e. have differentiated roles (for instance India distinct from others, and so on). We could also add regional exogenous variables such as concentration of urban population within the region (MC-URB). To the previous dyadic level synaptic weights, we would add small random or other suitable starting values for all other connections, the latter involving inputs from these new variables to the old hidden layer neurons. A further series of trials would allow whatever alterations might occur in this augmented scheme of weights.

Finally, in the third stage, inputs from stage one variables to *other* dyads in the group would be allowed, also. At this juncture, we also would include addition *endogenous* variables, in the form of attributes and behaviors of the group members that plausibly are linked as both inputs and outputs relative to the dispute onset hazard. An example is the “residual” of military personnel, that is, the degree to which actual personnel level of each group member deviates above or below its long-run trend (MP-R1, MP-R2, ...). The suggestion is that, while the trend may be relatively autonomous, the residual components are strongly coupled to the changes in hostility reflected in the dispute hazard. Again, the old synaptic weights would be augmented by weights connecting each input to the hidden layer neurons in each other dyadic neural net, and a series of trials would be allowed to produce a final set of synaptic weights.

The latter two stages of the above process would be carried out wherever it seemed reasonable to group actors and their corresponding dyads together. The final result would be a set of groups or individual dyads covering all members of the interstate system, and a series of components, each estimating the changes in HAZ and, where grouping was carried out, in MP-R1, etc., as a function of themselves and the other inputs that we have just mentioned. To reiterate the principal advantage of this approach, one not need guess functional form of these relationships in advance; through repeated introduction of training examples, the models often converge toward error minimizing configurations.

Success with this process, even to the point of accurate prediction of cases based on *new* data, would hardly conclude the inquiry, however. Even less than with more conventional modeling approaches, predictive accuracy would only be the beginning. In more conventional approaches, one begins by positing some kind of explicit functional relationship among data--presumably reflecting some conception of the substantive relationships at work --*then* one tests the model. If it shows acceptable accuracy, then presumably one's previous substantive conception provides an interpretation of that happy result. Neural network models, however, work quite well without our knowing the underlying function; so discovering the latter remains our task. In that sense, the usual modeling process is stood on its head. Still another way to make this point is that the effective synaptic weights, once determined, would themselves be new data telling us about the structure of relationships among the old. And we would have, in the model, a way of generating ("simulating") the observable results of the functional relationships. Thus the neural network is also a measuring instrument, revealing previously unseen data. Explaining that data then becomes part of our scientific agenda.

Appendix. Proposed Inputs, Neural Network Conflict Study

Group 1. dyadic indicator, endogenous

Estimated hazard of mutual MID onset [1]

Group 2. member state indicators, endogenous

Magnitude of residual of over-time estimate of [2, 3] ...
  Military personnel
  Military expenditures
  Iron or steel production
  Total population
  Urban population

Group 3. dyadic indicators, exogenous

Number of parties which are majors in this dyad (0,1,2)

Characteristics of prior mutual MID
  There was a prior dispute?
  A clear initiator of the dispute can be identified?
  Dispute consisted of a war not preceded by a codeable non-war event?
  Dispute is linked to previous interstate war involving one of the adversaries? [4]
  Dispute is linked to previous civil war involving one of the adversaries? [4]
  Number of sides on which there was a major power (0,1,2)
  Number of pairs of opposed parties in the dispute.
  Number of pairs of opposed major powers in the dispute.
  Highest level of hostile behavior exhibited by any participant [5]
  Duration of dispute.

Characteristics of prior mutual war
  There was a prior war?
  Number of sides on which there was a major power (0,1,2)
  Number of pairs of opposed parties in the war.
  Number of pairs of opposed major powers in the war.
  Number of fatalities.
  Distribution of fatalities [6]
  Duration of war.

Diplomatic recognition

  Level / rank
  Presence or absence

Common alliance membership
  Number of common memberships
  Presence or absence

Common IGO membership
  Number of common memberships
  Presence or absence

Geographical contiguity
  Presence or absence, most proximate criterion
  Presence or absence, least proximate criterion

Cross national majority-minority demographic linkage [7], alternate groups based on ...
  Religion
  Ethnicity
  Language

Elapsed time since [8]...
  onset of oppositely sided / same sided ...
    prior war [9]
    2nd prior MID / war
    3rd prior MID / war
  termination of oppositely sided / same sided ...
    prior MID / war
    2nd prior MID / war
    3rd prior MID / war
  most recent establishment / termination of mutual...
    diplomatic recognition
    military alliance
  most recent mutual exchange of territory.

Pair-wise entropy, 5 relative material capability indicators [3, 10]
  Military personnel
  Military expenditures
  Iron or steel production
  Total population
  Urban population.

Group 4. member state indicators, exogenous

Magnitude / rate of change, relative capability shares for [11] ...
  Military personnel
  Military expenditures
  Iron or steel production
  Total population
  Urban population.

Number of currently underway events or interactions in which party is a participant
  MIDs (excluding wars)
  Wars
  Diplomatic exchanges
  IGO memberships
  Military alliances.

Sum of MID onset hazard rates, named party to all others. [12]

Homogeneity of domestic demographic groups [13] , based on ...
- Religion
- Ethnicity
- Language.

Elapsed time since [8]...
- Acquisition of system membership
- Most recent acquisition of major power status
- Most recent loss of major power status
- Most recent entry to / exit from participation in MID / war involving ...
  - Major powers on both sides
  - Major powers on one side only
  - Minor powers only.

Group 5. systemic and regional indicators (all exogenous)

Material capabilities concentration based on information-theoretic entropy of 5 national indicators [3, 14]
- Across 8 major powers, equally weighted
- Across all system members, weighted by sum of MID onset hazards [15]

Total number of MID onsets in system / region, observed versus estimated [16]
- arithmetic difference method
- log ratio method

Current calendar time.

Number of ...
- MIDs / wars underway in system / region in which ...
  - majors are on both sides.
  - majors are on one side only.
  - majors are on neither side.
- dyads currently on opposite sides of a MID / war in system / region, composed of ...
  - both majors
  - one major- one minor
  - both minors
- system / regional members
  - classed as majors
  - in total
- pairs of ...
  - majors in system / region
  - nations classed as system / regional members.

"Effective" number of pairs of system members [17]

Notation and terminology:
1) When two terms are separated by / , then both denote categories, for each of which the stated indicator is to be separately evaluated.
2) “Endogenous” variables are those the annual increments of which will be modeled; all other variables are “endogenous”.
3) MID denotes militarized interstate dispute.

4) ? denotes indicator in the form of a question answered yes or no.

Notes to specific entries:

[1] The hazard rate of an event can be defined as follows: Consider a group of entities, each of which is at "risk" to experiencing a given event. We call this group the "risk set"; and whenever one of the entities in question experiences the given event, we consider this entity to be removed from the risk set. Let s denote the elapsed time since the entities in question were first members of the risk set and let $f(s)$ be the proportion of the original membership still in the set at time s; so, by definition $f(0)=1$. Then the hazard rate $h(s)$ is defined by $h(s) = -d \ln f(s) / ds$ (Tuma and Hannan 1984; Williamson, Warner and Hopkins 1988). This definition meshes with other common statistical concepts; for instance $h(s) =$ constant defines a Poisson distribution for the event.

[2] Logarithm of material capability indicator is estimated by bivariate linear regression against time, of which the residual is shown. Our conjecture is that this deviation from the trend is more likely to be coupled to the conflict behaviors than the trend itself; for instance the rise in military personnel and expenditure, or in steel production above the "normal" may be coupled to foreign disputes and wars, whereas the long run trend is unchanged (Organski and Kugler 1980).

[3] Energy consumption is omitted due to absence of data prior to 1860 (majors) and 1920 (minors).

[4] Linkages are coded on the possibility that they are aggravating factors.

[5] Threat to use force = 1, display of force = 2, use of violence = 3, war = 4 (Gochman and Maoz 1984).

[6] Based on formula for information theoretic entropy. Let p be the proportion of total fatalities incurred by one of the two sides. Then the distribution number is given by $-[p \ln p + (1-p)\ln(1-p)] / \ln 2$.

[7] Adapted from information theoretic entropy. Let p_{1i} and p_{2i} be the proportion of group i relative to total population in state 1 and 2, respectively and N be the number of such groups. Then the linkage measure is given by $-\sum_{i=1}^{N} p_{1i} p_{2i} [\ln p_{1i} + \ln p_{2i}] / \ln N$. This measure will have a large value when one group is all the population of one party and half the population of the other, which combination is conjectured to aggravate the potential for conflict.

[8] Elapsed time may contribute in two ways. First, elapsed time since a most recent prior event is a surrogate for the hazard rate of recurrence of the event. Plausible values of the latter vary in a regular way with elapsed times after previous such events (but, given sufficient resources, we would prefer to estimate the hazard rates of these events and use them directly as model inputs). Second, times since 2nd and previous occurrences may be surrogates for unobserved factors that contribute to hazards and other variables. (For instance, the hazard of an MID onset may reflect the combined effect of how long it has been since the most recent *several* such onsets.

[9] Time since prior MID is omitted since the hazard rate is directly estimated.

[10] Same as the formula for fatalities distribution in note 6, except that p is now the proportion of capabilities held by one party relative to their joint total.

[11] Relative capability share is the value of a capability indicator divided by the world total. For each year, the rate of change of the logarithm of this capability share might be estimated by a modified bivariate linear regression against time, in which prior years receive progressively declining weight and future years receive no weight, in contributing to the estimate.

[12] For each state j, its hazard rates λ_{jk} in relation to other states k for the given year are summed across all k to form an estimate of its net propensity to engage in disputes with other parties (Williamson 1994).

[13] Defined as information-theoretic entropy based on group population as fraction of state total. Let p_i be the proportion of population of group i over total state population and let N be the number of such groups. Then the measure is given by $-(1/\ln N)\sum_{i=1}^{N} p_i \ln p_i$. This measure is evaluated separately using each of the three indicated types of group division.

[14] The relative entropy H is defined the same as the measure of note 13, except that p_i is now the proportion of the indicated capability held by state i relative to the group total, where the group is 8 major powers (Britain, France, Russia, Japan, Germany, Austria-Hungary, Italy, United States) or all system members. (China is omitted from the list of major powers due to missing data prior to 1860.) The concentration measure is given by $K = 1 - H$. There is some evidence that the logarithm of the risk of dispute escalation to global war is proportional to dK/dt (Williamson 1994).

[15] In this weighting scheme, each capability value is multiplied by the total dispute hazard estimate of note 12 to yield a new value $c_j = x_j \sum_{k=1}^{p} \lambda_{jk}$, where x_j is the capability of the given nation j. This weighted value is re-normed through division by the system total of such normed capabilities to yield a new value p_i' in place of the p_i of note 14 (Williamson 1994).

[16] The estimated number of dispute onsets, year-by-year, is obtained as the sum of hazards across all dyads in the system or region, given by

$$\Lambda = \sum_{k<j} \lambda_{jk} ,$$

where j and k index all system or regional states. The discrepancy between this and the empirically observed number O of actual dispute onsets is the basis of the measure in question. In the arithmetic difference method the discrepancy is measured by $\mathrm{O} - \Lambda$; in the log ratio method, it is given by $\ln(\mathrm{O}/\Lambda)$. Elsewhere one of us has argued that this discrepancy is a plausible measure both of a "system breakdown" at the time of major war outbreaks, and of an associated non-linearity in scaling from dyadic to systemic conflict risk which is manifested as dispute onset contagion among the dyads (Williamson, Warner and Hopkins 1988, Williamson 1994).

[17] Let M, m, and g denote the number of major powers, minor powers, and number of pairs of minor nations each sharing a common geographical border, respectively. Then one reasonable estimate of the effective number of pairs is $[M(M-1)/2] + Mm + g$.

Bibliography

Etcheson, Craig (1989). *Arms Race Theory: Strategy and Structure of Behavior*. Westport, CT: Greenwood Press.

Gochman, Charles S. and Zeev Maoz (1984). "Militarized Interstate Disputes, 1816-1976: Procedures, Patterns, and Insights." *J. Conflict Resolution* 28 (Dec.): 585-616.

Gurr, Ted (1994). “Peoples Against States: Ethnopolitical Conflict and the Changing World System.” *International Studies Quarterly* 38, No. 3 (Sept.): 347-377.

Houweling, Henk W. and Jan G. Siccama (1988). *Studies of War*. Dordrecht, The Netherlands: Martinus Nijhoff Publishers.

Hammerstrom, Dan (1993). “Neural Networks at Work”. *IEEE Spectrum* (June) , pp. 26-32.

Haykin, Simon (1994). *Neural Networks. A Comprehensive Foundation*. New York: Macmillan.

Karasik, Myron S. and Paul R. Williamson (1994). "Modeling Complex Human Social Dynamics Using Neural Networks of Fuzzy Controllers," *Proceedings*, World Congress on Neural Networks, San Diego.

Lynch, William A. and John G. Truxal (1962). *Signals and Systems in Electrical Engineering*. New York: McGraw-Hill.

Murray, Carl (1993). “Is the Solar System Stable?” in Hall, Nina (ed.) *Exploring Chaos: a Guide to the New Science of Disorder*. New York: W. W. Norton (pp. 96-107).

Organski, A. F. K. and J. Kugler (1980). *The War Ledger*. Chicago: Univ. of Chicago Press.

Rich, E. and K. Knight (1991). *Artificial Intelligence*. New York: McGraw Hill.

Richardson, Lewis F. (1960). *Arms and Insecurity*. Pittsburgh: The Boxwood Press and Chicago: Quadrangle Books.

Singer, J. David and Melvin Small (1972). *The Wages of War, 1816-1965: A Statistical Handbook*. New York: Wiley.

Small, Melvin and J. David Singer (1980). *Resort to Arms: International and Civil Wars, 1816-1980*. Beverly Hills, CA: Sage.

Tuma, N. B. and M. T. Hannan (1984). Social Dynamics: Models and Methods. Orlando, FL: Academic Press.

Wayman, Frank W. (1989). "Power Shifts and War." Presented at Annual Meeting, International Studies Assn., London (29 Mar. - 1 Apr.). Dearborn, MI: Dept. Pol. Sci., Univ. of Mich.

Wilkinson, David (1987). "Central Civilization." Comparative Civilizations Review 17 (Fall): 31-59.

Wilkinson, David (1989). "Civilizational Convergence and War." Presented at Annual Meeting, International Society for the Comparative Study of Civilizations, Berkeley, CA (1-4 June). Los Angeles: Dept. Pol. Sci., Univ. of Calif., Los Angeles.

Williamson, Paul R. and Myron S. Karasik (1994). "A Proposal for Developing and Implementing a Global Information and Forecasting Service." Presented at Chicago Area Sigma Xi Clubs Forum on Sustainable Development, Loyola University, Chicago. Ann Arbor, MI 48106-4394, P. O. Box 4394: Global Vision, Inc.

Williamson, Paul R. (1994). “Capabilities Concentration, Dispute Contagion, and Global War: Findings, Synthesis, and Speculations.” Ann Arbor: Correlates of War Project, Dept. Pol. Sci., University of Mich. Unpublished.

Williamson, Paul R., John Warner, and Stephen A. Hopkins (1988). "A Model of International Dispute Onsets With Preliminary Application to the Impact of Nuclear Weapons." Ann Arbor, MI: Correlates of War project, Dept. Pol. Sci., Univ. of Mich. Unpublished.