

A Web-based Expert System for Gypsy Moth Risk Assessment

W.D. Potter, X. Deng, J. Li, M. Xu, Y. Wei, I. Lappas
Artificial Intelligence Center, GSRC 111, University of Georgia, Athens, GA, 30605
M.J. Twery, D.J. Bennett
USDA Forest Service, Northeastern Forest Research Station, Burlington, VT
(Contact W.D. Potter at potter@cs.uga.edu)

Abstract

The gypsy moth is one of North America's most devastating exotic forest pests because it can cause the loss of valuable oak species, degraded aesthetics, loss of wildlife habitat, and detrimental effects on watersheds. Due to the increasingly wide infestation of the gypsy moth, it is important to develop decision aids that help assess the risks of this pest to our forests. Expert systems are a type of decision aid that could be applied to the area of risk assessment. We have developed the Gypsy Moth Expert System to estimate the risk that a forest stand faces from the gypsy moth based on the composition, structure, and management objectives of a particular forest. Risk assessment in this context is developed from forest susceptibility to infestation, vulnerability to damage caused by an infestation, and the hazard that management objectives for a forest may be affected if damage occurs. The system uses a straightforward set of if-then rules to classify risk. The development of a web-based expert system presented significant challenges to maintaining remote user processing integrity.

Keywords: Gypsy Moth, Risk Assessment, Expert System

1. Introduction

Gypsy moth (*Lymantria dispar* L.) is one of North America's most devastating introduced forest pests. The species originally evolved in Europe and Asia, and has existed there for thousands of years. In the late 1860's, the gypsy moth was accidentally introduced near Boston by E. Leopold Trouvelot (Liebhold et al. 1989). About 10 years after this introduction, the first outbreaks began in Trouvelot's neighborhood. By 1890 state and federal governments began

taking steps to eradicate the gypsy moth (Gerardi and Grimm 1979). These first attempts ultimately failed and since that time, the range of the gypsy moth has continued to spread (Elkinton and Liebhold 1990). It seems inevitable that the gypsy moth will continue to expand its range in the future because the gypsy moth is known for feeding on the foliage of hundreds of plant species in North America. Its most common hosts are oaks (*Quercus spp.*) and aspen (*Populus spp.*). Gypsy moth hosts are located throughout most of the continental United States (Montgomery and Wallner 1988).

Gypsy moth populations are typically eruptive in North America. In a forest stand, population densities may fluctuate from 1 egg mass per hectare to over 10,000 per hectare. When densities reach very high levels, trees may become completely defoliated. Several successive years of defoliation, along with adverse contributions by other biotic and abiotic stress factors, may ultimately result in tree mortality. In some northeastern forests where trees are young and strong, few trees may die but occasionally tree mortality may be very heavy (Doane and McManus 1981).

Due to the wide infestation of the gypsy moth, it is important to develop decision aids that help assess the risks of this pest to our forests. In addition to helping determine the risk of an infestation, these aids can also help when determining the type, duration, and intensity of a pest control management strategy. Expert systems are a type of decision aid that could be applied to the area of risk assessment. Expert systems are very popular today in a number of different domains (Benders and Manders 1993, Turban 1993, Stefik 1995, Holsapple and Winston 1996). Some of the most well known systems are used for diagnostic purposes. In a physician's assistant system for example, the system receives as input information pertaining to a patient's specific problem, and then provides as output one or more alternative diagnoses for the patient. The system possesses the knowledge of human experts including the manner in which an expert would determine a solution, so that it reaches the same conclusion that the expert would reach when presented with the same input information (Bobrow et al. 1986, Stefik 1995).

We have developed a web-based expert system called GyMEs (Gypsy Moth Expert System, pronounced "jimmies") to estimate the risk a forest stand faces from gypsy moth. GyMEs is an enhanced version of the GypsES expert system component (Thomas et al. 1998). GyMEs contains a revised and extended knowledge base as well as a more up-to-date inference

mechanism. The expert system component of GypsES was originally developed for a MacIntosh computer (Twery and Elmes 1991). The output of these systems is a rough assessment of the risk to a forest stand due to gypsy moth, which is characterized either as "very low", "low", "moderate", "high" or "very high". The input is a series of environmental factors that have been observed in the stand for the last five years, as well as general information regarding the stand and management strategy. More specifically, the user of the system is prompted to answer a number of questions related to the stand itself, such as the age of the stand, its defoliation and drainage history etc. Then, the system processes the user's responses and determines the risk assessment.

One of the important characteristics of our expert system, as well as almost all expert systems, is that it does not need to have the answers to every input question in order to reach a conclusion. Sometimes the conclusion can be drawn based on the user's answer to a single question, and at other times many or all of the questions must be answered to determine an assessment. This is a short-circuit feature characteristic of consulting expert systems. Another characteristic of GyMEs is that the system will not ask the same question twice. For example, if one line of reasoning is being followed by the system and the system needs to change to another line of reasoning, then responses given earlier will automatically be applied to this new approach. The system retains the information given to it by the user for future use. Thus, eliminating unnecessary user interaction. The most significant characteristic of GyMEs is that it is web-based, that is, it runs over the Internet. A forest manager can "surf" into the GyMEs web site and interact with the system to assess a particular forest stand. This is significant because each interaction over the web is disjoint or physically separate from every other interaction. In order to provide a (logically) cohesive interactive session, the system must bridge these gaps.

In the following sections, we present our approach to the risk assessment problem, the knowledge that our system contains, and the conclusions that it may draw. A more precise description of the system's implementation is also presented.

2. Risk Assessment

The risk assessment evaluation in GyMEs is based on a number of parameters, using a structure developed by Gottschalk and Twery (1989) and described in detail by Bennett (1995).

Each parameter may have numerous acceptable values. But, what is the risk of a gypsy moth infestation? Risk is the chance we face of losing something of value combined with the expected loss based on our actions within the time frame of an appropriate planning horizon. The more valuable something is, the larger the loss we face if a loss-inducing event occurs. When managing a forest stand we take certain actions to achieve certain specified goals for the stand. Based on the conditions of the stand (i.e., how vulnerable the stand is to damage) and the estimated impact of an expected gypsy moth population on stand defoliation, we determine the level of damage (loss) that can be expected. The risk measures the extent of disruption the gypsy moth is expected to have on our management goals during the time of concern. This disruption corresponds to the loss we face if, in this case, we do not take any action to lessen the gypsy moth impact on the stand (Twery 1991, Bennett 1995). If the stand conditions and defoliation prediction indicate a very high risk, then there is a significant chance that the future condition of our stand will be different from what we want (our established goal).

In GyMEs, risk is derived from a hazard rating combined with a defoliation prediction. To determine the hazard rating, we use the susceptibility and vulnerability of a stand. A defoliation prediction is an estimate of the amount of defoliation damage a stand will incur from some level of pest infestation. Associated with an infestation is the likelihood that the stand will in fact incur defoliation. This is called the stand susceptibility. The stand's vulnerability is an assessment of the probable degree of mortality that would result from defoliation. Finally, combining susceptibility, vulnerability, and management objectives leads to a hazard rating. This, combined with a defoliation prediction, then brings us back to the risk assessment. Additional details of these concepts are discussed more fully in (Bennett 1995).

Assessing the risk of a gypsy moth infestation requires knowledge and expertise due to the large number of parameters that must be considered. If we take all the combinations of parameter set values, we have the total solution space for this problem. If we add a single additional parameter value, the total number of combinations increases exponentially. In addition, the problem is aggravated by the fact that the gypsy moth not only causes damage to a forest's health, but also interferes with the overall quality of the ecosystem. This is further complicated by the fact that a forest manager's goals for the forest may have little to do with the actual health of the trees, but more, for example, to do with the happiness of hikers passing

through the forest. Accurately assessing and predicting the risks of infestation by the pest is important for prescribing the proper short-term and long-term forest management practices. With a web-based expert system, the knowledge and expertise of the domain expert can be easily adopted and utilized by many users and at many different locations without the presence of the expert.

3. Expert Systems Development

The following are the steps used in developing the GyMEs prototype. These include problem identification, knowledge base design, and knowledge base validation (Stefik 1995).

An expert system requires a precise domain. The domain must be well organized and well understood. The selected application within the domain will need to require expert knowledge in order to solve specific problems within that domain. Otherwise, there would be no need for an expert system; a standard algorithmic search scheme would be more suitable in that case. This means that the types of problems encountered within the domain should exhibit combinatorially explosive solution spaces when using an exhaustive search scheme to find a solution. For example, in the diagnosis domain, as the number of disorders (diseases) increases linearly, the number of possible diagnoses increases exponentially (i.e., there are exponentially more combinations of diagnoses to consider). This type of growth in the total number of solutions is called combinatorial explosion. Clearly, gypsy moth risk assessment is an appropriate domain for expert systems development.

The knowledge base is the core component of any expert system since it contains the knowledge acquired from an expert in the field. Typically, a knowledge engineer is responsible for working with an expert to build the knowledge base for the system. The knowledge engineer must perform a detailed analysis of the inference process and develop the prototype knowledge base. The tasks involved in developing the GyMEs knowledge base include knowledge acquisition, knowledge representation, knowledge programming, and knowledge refinement.

The objective of knowledge acquisition is to obtain facts and rules from the expert that will allow the system to draw expert level conclusions. The process of knowledge acquisition is very time-consuming and difficult, especially if the knowledge engineer is unfamiliar with the domain. In our case, there were several knowledge engineers working on the project and among

them were very knowledgeable foresters. For GyMEs, the expert knowledge was originally derived from co-author Twery of the USDA Forest Service (Bennett 1995). The resulting knowledge base was later expanded, analyzed and refined. As a result, additional rules were added and many were updated.

After the acquisition of the knowledge begins, a prototype system implementation is begun to test the early stages of the system. This process involves encoding the expert knowledge into the proper format for the computer. Representing and encoding the facts and relationships that constitute knowledge is the next step in the system implementation. There are many established approaches of representing knowledge, for example, semantic networks, rules and logic expressions. "If - Then" style rules are, however, widely used because they are easy to understand and enhance. They facilitate the addition of an explanation facility early in the development process or as an add-on later in the development. GyMEs employs a typical rule-based knowledge representation.

During the knowledge programming stage, we first design an overall framework and systematic representation scheme based on the rules derived from the expert. The knowledge is assembled into an organized rule base for the inference engine to interpret and use. The process involves coding facts, rules, objects and relationships found in the domain in the programming language for the system. The following are examples of rules taken directly from the GyMEs knowledge base.

```
risk(very_low) :-  
    Defolp == 1.           % if the defoliation prediction is very low (i.e., 1), then the  
                           % risk is very low
```

```
risk(very_low) :-  
    Defolp == 2,           % if the defoliation prediction is low (2) and  
    hazard(very_low).      % the hazard rating is very low, then the risk is very low
```

```
risk(very_low) :-  
    Defolp == 2,           % if the defoliation prediction is low (2) and  
    hazard(low).           % the hazard rating is low, then the risk is very low
```

```
risk(very_low) :-  
    Defolp == 2,           % if the defoliation prediction is low (2) and  
    hazard(moderate).      % the hazard rating is moderate then the risk is very low
```

```
risk(very_low) :-
```

Defolp ::= 3, % if the defoliation prediction is moderate (3) and
hazard(very_low). % the hazard rating is very low, then the risk is very low

In the above rules, the risk assessment is determined to be very low, if any one of the following conditions is satisfied:

1. the defoliation prediction equals to 1, OR
2. the defoliation prediction equals to 2 AND the hazard rating is very low, OR
3. the defoliation prediction equals to 2 AND the hazard rating is low, OR
4. the defoliation prediction equals to 2 AND the hazard rating is moderate, OR
5. the defoliation prediction equals to 3 AND the hazard rating is very low.

Notice there are five different sets of conditions that could lead us to conclude the risk assessment is very low. Each condition set is different and is based on some input received from the user and some previous conclusion determined by the system. In this case, the first rule regarding the defoliation prediction requires an input value entered by the user. If the user rates the potential for defoliation as very low, then the system concludes that the estimated risk is likewise very low. On the other hand, all of the other rules are dependent on this user input and the additional condition related to the hazard rating. According to the expert, these value combinations for hazard rating and defoliation prediction lead us to conclude a very low risk assessment. Here, the hazard rating relates to the health of the trees in the stand, the stand management objectives, and the potential for gypsy moth habitation. These parameters are derived from other rules and user input data.

4. GyMEs – the Gypsy Moth Expert System

The GyMEs knowledge base includes over 300 facts and rules for deciding the level of gypsy moth risk. The knowledge base is divided into six categories, including stand susceptibility, disturbance history, stand condition, site factors, intervention strategy and defoliation predication. Each category in the knowledge base is further subdivided into 3 to 5 levels and each level is determined by detailed information provided by the user. For example, stand susceptibility is determined by the stand age and species composition of the stand. An organization chart of the knowledge base is presented in Figure 1.

Figure 1 illustrates that the user is prompted to input information on stand condition, crown condition, site factors, disturbances and species composition. The knowledge base uses these inputs to determine the severity of each of the six category values and eventually provides a gypsy moth risk assessment for the stand. The rules are compiled to represent the knowledge base for GyMEs. The following are additional examples of the rules (in an English format for easy reading) maintained by the system:

```
IF
    the stocking percentage is greater than or equal to one percent
    AND
    the stocking percentage is less than or equal to eighty percent
    AND EITHER
        the crown condition is good
    OR
        the crown condition is fair
THEN
    the stand condition is good.
```

In the above rule, the stocking percentage is an index used to estimate the amount of growing space being utilized by trees within the stand. The crown condition is a rating determined using the amount of damaged or dead limbs in the trees' crowns. These factors give us an indication of the stand's condition or pre-existing stress level. The user provides the values for these factors as a result of a query that the system prompts the user with. As it turns out, this is a special rule where the English version shown above does not correspond exactly with the actual rule in the knowledge base. If the user satisfies the stocking percentage conditions yet does not know the crown condition (e.g., no data available) then the system concludes that the stand condition is good by default. Of course, the user may backtrack to this particular point in the prompt sequence once information is available on crown condition. Assuming a good or fair rating, the stand condition would be determined to be good.

```
IF
    the defoliation prediction is greater than or equal to thirty percent
    AND
    the defoliation prediction is less than sixty percent
    AND
    the hazard rating is moderate
THEN
```


the risk is moderate.

In this rule, the defoliation prediction value is determined using field data indicating gypsy moth egg mass counts. Based on egg mass counts, stands are categorized according to the level of defoliation expected in the coming year. The hazard rating is based on additional factors including stand vulnerability and management strategy (Bennett 1995). A hazard exists when conditions in the forest indicate damage that is likely to be incurred due to a gypsy moth infestation would impair the forest management objectives.

As a web-based application, the GyMEs user interface was originally designed using the hypertext markup language (HTML) and the common gateway interface (CGI) to allow the user to access the system with a web browser such as Microsoft Internet Explorer or Netscape Communicator (Brown and Honeycutt 1998). Generally, the user is also referred to as the web surfer or the client while the computer system being accessed may be referred to as the web site or the server. The interface contains three main parts: a generic form that allows information to be obtained from the user, a mechanism that allows interaction with the inference engine in order to perform various computations on that information, and a generic form that allows results to be presented on the user's screen. The main reason for using HTML as a front end was that by doing so we could take advantage of the fact that there exist web browsers for virtually every platform you might be interested in using for surfing the web.

The screen layout consists of two horizontal frames or windows. The top horizontal window shows the current question being asked by the system. This window also contains a list of possible answers with radio buttons or an input box for the user's response. The user can move the cursor to the appropriate answer and mouse click the "Submit" button located in the lower part of the input box. After the user selects or types in an answer and submits it, the system stores the information in a file and displays it on the second window. The risk assessment will be presented whenever the system receives enough information to reach an assessment conclusion.

As mentioned earlier, GyMEs employs a backward chaining inference engine. Therefore, the system is given a specific goal or conclusion to determine based on the order of the rules in the knowledge base and then proceeds to search the knowledge base for rules that have that goal as its conclusion. Once the inference engine finds an appropriate rule, it recursively searches

through the knowledge base to determine if it can satisfy the premise (conditions) of that rule. If this is not possible, GyMEs looks for another rule with that original goal as its conclusion (backtracks), and so on until a conclusion is found.

GyMEs uses a prolog inference engine located on the server (i.e., the machine at our web site) to manipulate the assessment knowledge base. Since each interaction with the server from a client (the web surfer) is an independent event, the server also maintains a client specific file of known facts and derived conclusions for the current decision making session. After each interaction, the server consults this file before invoking the inference engine. Armed with the facts and derived results in the file, and the user's latest input, the inference engine processes the rules in the knowledge base to determine the next conclusion. If this is a risk assessment, then the user is notified. If not, the server results file is updated and the prompt for the next piece of required information is determined and sent to the client's browser. Prior to implementing this client specific file scheme, only one client at a time was allowed to access GyMEs. However, we devised a scheme where any number of clients could surf into the GyMEs web site and run the system. The scheme is based on the server keeping track of each client's session file, and being able to relate these session files to the proper client.

Before fully implementing our own session file scheme for tracking individual client interactions, we converted our prolog system to LPA Prolog with the Pro-Web feature (Shalfield 1998). Pro-Web handles all the client/server interactions automatically, making the GyMEs web site truly multi-user in an almost transparent fashion. That is, using Pro-Web gives us full multi-user capability so that any number of surfers may surf into the system. However, in order for Pro-Web to handle the problem of relating clients to sessions, there are some restrictions on how the web interface is developed. To the client, this is totally invisible, which is exactly what we wanted.

5. Future directions

After compilation of the rules and making any necessary additions/changes, we partially verified the knowledge base to help ensure its accuracy. Verification is a necessary step to ensure that there are, for example, no dead-end lines of reasoning embedded within the knowledge base that would result in a "risk assessment unknown" conclusion derived via the

inference process. There must also be a cross check of the results against the conclusions provided by the domain expert, if possible. For GyMEs, this step is currently in progress.

One of the first issues for future research is the continued validation and expansion of the GyMEs knowledge base. One type of validation needed is criterion-related validity to determine the relationship between the expert system outcome and the advice from the domain expert. Another issue is the reasoning process used by the GyMEs inference engine. After a user enters information, the system should provide a way to track the logical reasoning process, and then output the information if the user requests it. Another issue to be addressed in the near future is the inclusion of an explanation facility. When a user is prompted for some information about the stand, it would be beneficial for them to have the ability to ask the system why it needs that information. Also, when an intermediate conclusion or risk assessment is derived, users may want to know how that conclusion was derived. This type of facility increases the utility of an expert system.

A further goal is to develop GyMEs as a module that can be incorporated within the GypsES software system. This phase of development will soon be completed. However, our secondary goal is to enhance the web-based version. Our proposed approach is based on using java servlets (Moss 1998) to handle the server interaction to the inference engine. Servlets are applications running on the server to manage client requests. This approach will allow us more flexibility than currently provided by Pro-Web and will improve the server's ability to handle client requests.

References

- Bennett, D.B. (1995) Modeling a decision process: hazard and risk assessment for areas threatened by gypsy moth infestation. Master's Thesis. West Virginia University. Morgantown, WV.
- Benders, J. and F. Manders (1993) Expert systems and organizational decision-making. *Information and Management*. 25: 207-231.
- Bobrow, D.G., S. Mittal and M. Stefik (1986) Expert systems: perils and promises. *Communication of the ACM*. 29: 880-894.
- Brown, M.R. and J. Honeycutt (1998) *Using HTML 4*, Fourth Edition. Que Corporation, Indianapolis, IN.

- Doane, C.C. and M.L. McManus, eds. (1981) The gypsy moth: research toward integrated pest management. USDA Forest Service Technical Bulletin 1584.
- Elkinton, J.S. and A.M Liebhold (1990) Population dynamics of gypsy moth in North America. *Ann. Rev. Entomol.* 35: 571-596.
- Gerardi, M.H. and J.K. Grimm (1979) The history, biology, damage, and control of the gypsy moth. Associated University Press, Inc. Cranbury, New Jersey.
- Gottschalk, K.W. and M.J. Twery (1989) Gypsy moth impacts in pine-hardwood mixtures. pp. 50-58, *Proc. Pine-Hardwood Mixtures: A Symposium on Management and Ecology of the Type.* 1989 April 18-19; General Technical Report. SE-58. USDA Forest Service, Southeastern Forest Experiment Station.
- Holsapple, C.W. and A.B. Winston (1996) *Decision Support Systems – A Knowledge-Based Approach.* West Publishing Company. New York.
- Liebhold, A., V. Mastro, and P.W. Schaefer (1989) Learning from the legacy of Leopold Trouvelot. *Bulletin of the Entomological Society of America.* pp. 20-22.
- Montgomery, M.E. and W.E. Wallner (1988) The gypsy moth – a westward migrant. p. 353-375. In: Berryman, A.A. (ed.) *Dynamics of Forest Insect Populations.* Plenum Publishing Corporation.
- Moss, K. (1998) *JAVA Servlets,* McGraw Hill Publishing, New York.
- Shalfield, R. (1998) *LPA Pro-Web 1.0 User Guide.* Logic Programming Associates, Ltd. London. England.
- Stefik, M. (1995) *Introduction to Knowledge Systems.* Morgan Kaufmann, California.
- Thomas, S.J., S.L.C. Fosbroke, and A.B. Cumming (1998) *GypsES: Decision Support and Project Management – User’s Guide, Version 1.0.* USDA Forest Service. Northeastern Forest Experiment Station. NA-TP-01-98.
- Turban, E. (1993) *Decision Support and Expert Systems: Management Support Systems.* 3rd ed., Macmillan, New York.
- Twery, M.J. (1991) Effects of defoliation by gypsy moth. In: Gottschalk, K.W.; Twery, M.J.; Smith, S.I., eds.; *Proceedings USDA InterAgency Gypsy Moth Research Review;* January 22-25, 1990; East Windsor, CT. Northeastern Forest Experiment Station, General Technical Report NE-146.

Twery, M.J. and G.A. Elmes (1991) Hazard rating for gypsy moth on a macintosh computer: a component of the GypsES system. In: Gottschalk, K.W.; Twery, M.J.; Smith, S.I., eds.; Proceedings USDA InterAgency Gypsy Moth Research Review; January 22-25, 1990; East Windsor, CT. Northeastern Forest Experiment Station, General Technical Report NE-146.

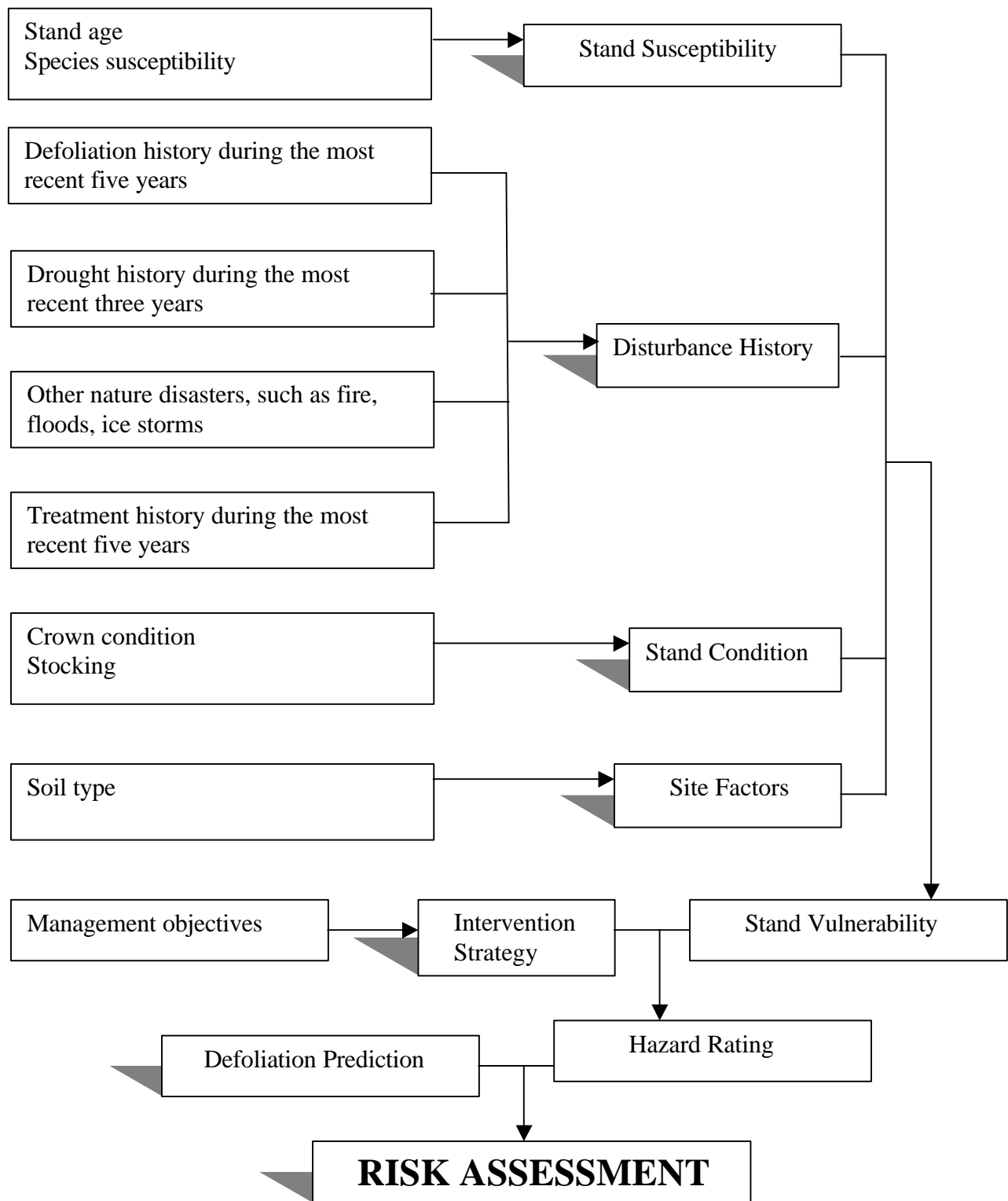


Figure 1. Flow chart of knowledge base

