A MIXED PARADIGM REASONING APPROACH TO PROBLEM SOLVING IN INCOMPLETE CAUSAL-THEORY DOMAINS

by
John Douglas Hastings

A dissertation submitted to the Department of Computer Science
and The Graduate School of The University of Wyoming
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in
COMPUTER SCIENCE

Laramie, Wyoming
December 1996

To The Graduate School:

The members of the Committee approve the dissertation of John Douglas Hastings presented on September 25, 1996.

L. Karl Branting, Chairman

Jeffrey Van Baalen

Stanley R. Petrick

Jeffrey A. Lockwood, Outside Member

John R. Cowles

APPROVED:

Henry R. Bauer, III, Head, The Department of Computer Science

Donald S. Warder, Dean, The Graduate School

Hastings, John Douglas, A Mixed Paradigm Reasoning Approach to Problem-Solving in Incomplete Causal-Theory Domains, PhD, Department of Computer Science, December, 1996.

Many complex physical systems such as biological, ecological, and other natural systems are characterized both by incomplete models and limited empirical data. Accurate prediction of the behavior of such systems requires exploitation of multiple, individually incomplete, knowledge sources.

This dissertation describes model-based adaptation, a technique for integrating case-based reasoning with model-based reasoning to predict the behavior of biological systems characterized both by incomplete causal models and insufficient emprical data for accurate induction. This approach is implemented in CARMA, a system for rangeland grasshopper management advising. CARMA implements a process model derived from protocol analysis of human expert problem-solving episodes. CARMA's design attempts to emulate the speed, graceful degradation, opportunism, and explanatory ability of human experts.

CARMA's ability to predict the forage consumption judgements of expert entomologists was empirically compared to that of case-based and model-based reasoning techniques in isolation. This evaluation confirmed the hypothesis that integrating model-based integrating model-based and case-based reasoning can lead to more accurate predictions than the use of either technique individually.

Acknowledgements

I would like to thank Karl Branting and Jeff Lockwood for all their help on this research. Your advice, criticisms, and suggestions made everything go as smoothly as possible. Above all, your doors were always open whenever I had a question.

I would like to thank my other committee members, John Cowles, Stan Petrick, and Jeffrey Van Baalen for helping with the final version of this paper at a busy time of the semester.

Thanks also go out to the entire Computer Science Department for everything you have taught me. Thank you, Janet and Carol for all the "little" things.

Thanks to my family and friends for all the support given to this "career" student and my wife Naomi for all the motivation in the final phases of the project.

Table of Contents

Chapter 1

Introduction

1.1 Predicting the Behavior of Physical Systems

One of the most striking characteristics of human problem-solving behavior is the ability to exploit multiple knowledge sources and reasoning techniques. This ability is important because problem solving often occurs in an environment of incomplete knowledge. Automating this ability requires techniques for integrating multiple problem-solving paradigms in a flexible manner.

For example, many types of diagnostic, monitoring, and planning tasks require prediction of the behavior of physical systems[1]. Precise models exist for the behavior of many simple physical systems. However, models of biological, ecological, and other natural systems are often incomplete, either because a complete state description for such systems cannot be determined or because the number and type of interactions between system elements are poorly understood. Empirical methods, such as case-based reasoning, decision-tree induction, or statistical techniques, can be used for prediction if sufficient data are available. In practice, however, many biological systems are characterized both

[1]In the life sciences, physical systems are contrasted with biological systems, but in this paper biological systems are classified as a subset of physical systems.

by incomplete models and insufficient empirical data for accurate induction. Accurate prediction of the behavior of such systems requires exploitation of multiple, individually incomplete, knowledge sources.

1.2 Research goals

The research described in this dissertation attempts to satisfy both theoretical and practical goals. Theoretically, this research demonstrates how multiple knowledge sources may be integrated for the purpose of providing accurate predictions about the behavior of physical systems whose causal theory is incomplete. Underlying this approach is the view that computer systems should emulate the human capacity to employ whatever reasoning technique is most appropriate for a given task or knowledge type. The goal of this approach is to enable computer systems to optimize the use of the diverse and incomplete knowledge sources available to decision-makers and to produce patterns of reasoning that resemble those of human decision-makers. This integration will be shown in the context of rangeland grasshopper management advising, a specific task arising within rangeland management that requires predictions in a biological system characterized both by an incomplete model and insufficient empirical data for accurate use of empirical techniques.

The main focus will be a specific integration technique, model-based adaptation, for combining an incomplete causal model with case-based reasoning to improve predictions about the forage loss caused by grasshoppers in a rangeland ecosystem. The technique of integration to predict forage loss is compared to other problem-solving approaches. Ablation tests are used to evaluate the relative contributions of the incomplete causal model and case-based reasoning to predict forage loss.

On the practical side, the rangeland grasshopper management advising task has been implemented in a system termed CARMA (CAse-based Range Management Adviser). CARMA demonstrates that integrating various reasoning paradigms can lead to a useful advising system. CARMA's advice is evaluated by comparing it to the advice given by entomologists.

1.3 Reader's guide to the dissertation

This chapter motivates the dissertation in the context of predicting the behavior of physical systems with an incomplete causal theory domain. Chapter 2 introduces rangeland grasshopper management advising as a specific task that requires making such predictions. Although many of the components of a rangeland grasshopper ecosystem are known, the causal model is incomplete in that the interactions among the components are only partially understood. The incompleteness of the various knowledge sources requires integrating them to produce the most accurate advice.

Chapter 3 describes CARMA, a system that advises ranchers about the best response to rangeland grasshopper infestations by integrating the multiple problem solving paradigms (specifically model-based and case-based reasoning) used by human experts.

Chapter 4 describes how CARMA's case-based reasoning component learns match and featural adaptation weights in order to maximize its predictive accuracy.

Chapter 5 details the evolution of CARMA through various configurations based on tests of the forage consumption prediction component. An evaluation of CARMA's final configuration is provided in terms of forage consumption predictive accuracy and treatment recommendation quality.

Chapter 6 compares this dissertation to other research efforts.

Chapter 7 discusses the contributions of this research and proposes future work.

Chapter 2

An Information Processing View of
Rangeland Grasshopper Management

This chapter introduces the task of rangeland grasshopper management advising. Rangeland grasshopper management advising is a specific task arising within rangeland management that requires making accurate predictions about the behavior of a physical system with an incomplete causal theory. The absence of a complete and accurate model necessitates integrating a variety of individually incomplete knowledge sources. The applicability of various knowledge sources to this task is described in a process description of expert problem solving. In performing this process, human experts exhibit several characteristics that are desirable to emulate in a computer system.

2.1 Rangeland grasshopper management

In most years and locations, the majority of grasshopper species are innocuous or even beneficial to grassland ecosystems. Of over 300 species of grasshoppers in the western United States, perhaps only 15 can be considered serious pests; many of the other species are beneficial in terms of controlling weeds, nutrient cycling, and food for wildlife (Lockwood 1993a; Lockwood 1993b). However, large-scale grasshopper outbreaks are capable of inflicting serious economic damage to western livestock producers. On average, grasshoppers annually consume 21-23% of western rangeland forage, at an estimated loss of $400 million (Hewitt & Onsager 1983). For example, in 1985-86, Wyoming treated approximately 6.5 million acres at a cost of $22.75 million to private, state and federal interests. Due in part to restructuring of the state cost-share program, some 9 million acres of infested rangeland were left untreated in 1987, resulting in the loss of 225,000 tons of air-dried forage.

Rangeland grasshopper management advising is a specific task that attempts to properly manage grasshopper infestations in order to minimize financial losses. It requires making accurate predictions about the amount of forage that will be consumed by grasshoppers and deciding whether any insecticide application is morally or economically justifiable. However, the decision whether to use insecticides or other control measures is a complex task because of the multiplicity of relevant factors, such as maintaining minimal inputs for profitable ranching in the western United States, preserving natural enemies for chronic control of grasshoppers (Joern & Gaines 1990), safeguarding biodiversity including beneficial grasshoppers, and protecting environmental and human health.

2.2 Behavioral Prediction with an Incomplete Causal Model

A causal model for the behavior of a physical system is a model of the interactions among the components of the system that is capable of predicting or explaining the system's behavior. In domains with complete causal theories, predictions about the behavior of the systems are made by looking at the state of the components within the system and applying the causal model. For example, knowing the voltage(s) applied to an electrical circuit and the causal model of the circuit - characteristics of each component and their connections - allows one to predict the circuit's behavior.

However, many domains lack a complete causal theory. Causal models may be incomplete in any of the following ways:

1. Imprecisely parameterized (i.e., the nature of the interactions is known only in a general sense),
2. Incompletely parameterized (i.e., the components are known to interact, but the nature of the interactions is unknown),
3. Incompletely connected (i.e., it is not known which components interact), or
4. Incompletely constructed (i.e., missing components relevant to the behavior of the system).

Although many of the components of a rangeland grasshopper ecosystem are known, the causal model is incomplete in that the interactions among the components are only partially understood. While model-based reasoning can play a role in grasshopper management, there is a general recognition that the interactions affecting grasshopper population dynamics are too poorly understood and too complex to permit precise prediction through numerical simulation (Lockwood & Lockwood 1991; Pimm 1991; Allen & Hoekstra 1992). Grasshopper populations are extremely labile, and a multiplicity of biotic and abiotic factors regulate their densities. Based on a simplified rangeland habitat comprised of just 10 grasshopper species, 10 plant species, four soil types, and 10 predators, Lockwood (1996) calculated that 10,560 two- and three-factor interactions end with grasshoppers. More realistic estimates of diversity suggest as many as 175 million interactions end with grasshoppers. If even 1% of these were ecologically relevant, the number of interactions would be far too great to simulate. Thus, while simulations of grassland ecosystems can provide insight into their dynamics (Fedra 1991; Rodell 1978), it is not feasible to devise models adequate for accurate prediction of the consequences of treatment options.

Despite the fact that rangeland ecology is poorly understood and very complex, entomologists experienced in rangeland management routinely provide useful advice to ranchers. The decisions of pest managers addressing this task appears to be largely based on a set of synthetic, prototypical cases that take into account the diversity of rangeland conditions, productivities, and vegetation types, the enormous range of weather conditions, and site-specific elements (*e.g.*, honey production, wildlife management, water development, etc.). These cases do not necessarily correspond to specific real world experiences, but express the essential features of past management experiences that define prototypical instances in which particular management practices are optimized. For example, the most authoritative guide for management of African (Desert) Locust infestations consists of a collection of discrete cases compiled as a reference for workers (Pedgley 1981)[2]. Specific cases have been found to be an important knowledge source in

---

[2]Although not explicitly setting forth any systematic or formal means of case-matching, (Pedgley 1981) is premised on the assumption that forecasting locust population dynamics and appropriate management strategies can be based on comparisons with specific cases.

a variety of problem domains in which precise general rules are unavailable or inadequate (Klein & Calderwood 1988).

A second form of knowledge used by experts in rangeland grasshopper management consists of general rules that appear to constrain management decisions under specific circumstances (*e.g.*, there is no point in controlling grasshoppers once the adults have laid eggs, as the majority of the damage is already done and the next generation is assured).

Finally, while the causal model of rangeland ecology is insufficient for providing accurate predictions about rangeland grasshoppers, there are some aspects of grasshopper population ecology for which there is sufficient information to apply mechanistic models (*e.g.*, the development of grasshoppers through the nymphal stage is a relatively well-defined function of temperature). Figure 1 sets forth the most important of the qualitative causal constraints that influence forage consumption. M+ indicates monotonically increasingly relations (*e.g.*, increasing the level of pesticides in the system causes an increase in grasshopper attrition), while M- indicates monotonically decreasing relations (*e.g.*, increasing grasshopper attrition causes a decrease

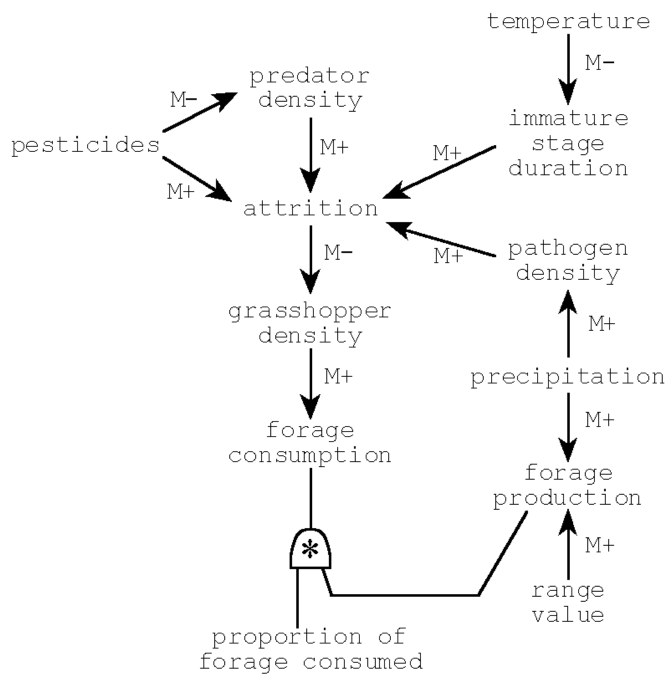

Figure 1: Qualitative relations in rangeland ecosystems.

in grasshopper density). The proportion of forage consumed is determined by dividing the forage consumption by the forage production. Other information available for modeling rangeland grasshopper ecosystems includes the following:

1. The developmental stages of grasshoppers, including
   a. The average length or developmental rate of each stage adjusted according to temperature.
   b. The proportion of lifetime consumption that occurs at each stage.
   c. The attrition rate at each stage adjusted according to precipitation.
2. Some species of grasshoppers, termed *nymphal overwintering*, hatch late in the *growing season*, hibernate during the winter as nymphs, and complete their development during the following growing season. Others, termed *egg overwintering* species, overwinter as eggs, then hatch, lay eggs and die within a single growing season.
3. The significant production of forage at a location occurs during a specific portion of the growing season, termed the *critical forage growing period*, for that location.

In summary, rangeland grasshopper management advising is a specific task arising within rangeland management that requires making accurate predictions about the behavior of a physical system with an incomplete causal theory. The absence of a complete and accurate model necessitates integrating a variety of individually incomplete knowledge sources, including both empirical and model-based knowledge.

2.3 Process Description of Expert Problem Solving

The ability of entomologists and pest managers to provide meaningful advice further indicates that other sources of knowledge can compensate for the absence of a complete model of rangeland ecosystems. To explicate these knowledge sources and also problem-solving methods, a protocol analysis of problem solving by several experts in rangeland grasshopper management at the University of Wyoming was performed. For each expert, several problem-solving episodes were transcribed in which the expert responded to a simulated telephone inquiry by a rancher. These "solve-aloud" problem-solving episodes illustrate the elicitation of relevant case facts by the expert, the formation and discrimination among tentative hypotheses, and expert explanations. Based on this protocol analysis, the following process description of expert problem solving for this task was developed:

1. Determine the relevant facts of the infestation case, such as grasshopper species, developmental phases,[3] and density, from information provided by the

---

[3]During their lifetime, grasshoppers progress through three developmental stages: egg, nymph, and adult. The nymphal stage usually consists of five instars separated by molts. The eight **developmental phases** of a grasshopper's lifecycle are defined as follows: 1 = egg; 2 = nymphal instars 1 through 3; 3 = instars 2 through 4; 4 = instars 3 through 5; 5 = instars 4 through 5; 6 = pre-egg laying adults; 7 = adults; and 8 = dead grasshoppers.

user. This requires **rule-based reasoning** using rules such as, "if grasshoppers are observed to have brightly colored wings or make a clicking sound in flight, then they are bandwinged adults that overwintered as nymphs."

2. Determine whether the grasshopper infestation is a potential problem. The infestation is not a problem if:
   a. The current date is outside of the "growing season" when forage needed for livestock grows. This requires **rule-based reasoning** to determine whether the date is outside of the growing season, given the historical growing season for the location and the date.
   b. The size of the infestation is small. This requires **rule-based reasoning** to determine whether the infestation size is below a minimum threshold.
   c. The majority of the grasshoppers overwinter as nymphs. This requires **rule-based reasoning** to determine whether the majority of the grasshoppers observed have brightly-colored wings or make a clicking sound in flight, and are therefore, adult bandwinged grasshoppers that overwintered as nymphs.
   d. The majority of the grasshoppers are in inappropriate phases. This requires **rule-based reasoning** to determine whether the majority of the grasshoppers are at such an early developmental phase that the extent of the infestation cannot be predicted with reasonable certainty or at such a late developmental phase that a significant proportion of lifetime forage

consumption and egg-laying have already occurred, making insecticide application pointless.

3. If the infestation is potentially a problem, determine whether grasshopper consumption will lead to competition with livestock for available forage.
    a. Estimate the proportion of available forage that will be consumed by each distinct grasshopper population (*i.e.*, nymphal overwintering, egg overwintering) using **case-based reasoning**. For each distinct grasshopper population (*i.e.*, subcase):
        i. Determine the prototypical case that most closely matches the current subcase. This requires **model-based reasoning** to assist matching by aligning the developmental phases of the prototypical case and the subcase.
        ii. Adapt the consumption estimate predicted by the prototypical case based on the featural differences between the prototypical and current subcase. This requires **model-based reasoning** to account for the influence of each feature on consumption.
    b. Total the forage loss estimates for each subcase to predict the overall proportion of available forage that will be consumed by grasshoppers.
    c. Compare grasshopper consumption with the proportion of available forage needed by livestock.
4. If there will be competition, determine what possible treatment options should be excluded. This requires **rule-based reasoning** using rules such as "wet

conditions preclude the use of malathion; environmental sensitivity precludes all chemical treatments."

5. If there are possible treatment options, for each one provide an economic analysis by estimating both the first-year and long-term savings.
    a. Estimate the first-year savings using **model-based reasoning** to determine the proportion of forage which would be saved given the efficacy of the treatment type, the developmental phases of the grasshoppers at the time of treatment, and the proportion of lifetime consumption by grasshoppers at each phase.
    b. Estimate the long-term savings using **rule-based reasoning** to determine if the majority of the grasshoppers will begin laying eggs before treatment can be applied given the developmental distribution of the grasshoppers at the time of treatment. If the majority of grasshoppers will not begin laying eggs, use **statistical reasoning** to determine the decreased probability of infestation in subsequent years given the Markov transitional probabilities for the infestation location and the effect of the treatment type on beneficial control agents.

In performing this process, human experts exhibit several desirable characteristics:

1. **Speed**. Human experts can provide useful advice very quickly. This suggests, consistent with the process description, that human experts can use highly compiled knowledge in the form of prototypical cases and rules.
2. **Graceful degradation**. Human experts can use, but do not require, highly precise information of the type required for accurate model-based reasoning. Less accurate information may degrade the quality of advice an expert can give, but doesn't preclude useful advice. In the worst case, human experts can provide plausible advice based merely on the location of the rangeland and the date.
3. **Explanations in terms of a causal model**. Although the speed and graceful degradation of human expert performance suggest that experts can use compiled knowledge, they can also readily provide causal explanations for their conclusions. Moreover, entomologists can generate causal predictions of the effects of incremental variations on case facts. This behavior strongly suggests that they have access to causal models that can assist in explanation and in adaptation of prototypical cases.
4. **Opportunism**. Human experts can use a variety of different strategies to solve a single given problem depending on the available information. Human experts don't address the subgoals that arise in decision-making in an invariant

order, but adapt their problem-solving behavior to the particular facts of a given case.

In summary, rangeland grasshopper management typifies a task in which the absence of a complete and accurate model necessitates integrating a variety of individually incomplete knowledge sources. The next chapter describes CARMA, a multiple-paradigm computer system for rangeland management.

Chapter 3

CARMA: A Multiple-paradigm Advisory System

CARMA (CAse-based Range Management Adviser) is a system for advising ranchers about the best response to rangeland grasshopper infestations. The protocol analysis described in Chapter 2 indicated that a solution should consist of a treatment recommendation supported by an explanation in terms of causal, economic, and pragmatic factors, including a numerical estimate of the proportion of forage consumed and a cost-benefit analysis of the various treatment options. Because this advice can't be produced by any individual reasoning technique, the focus of the CARMA project has been on integrating the multiple problem solving paradigms used by human experts.

Figure 2 shows the goal-structure that CARMA attempts to satisfy during a consultation, including a treatment recommendation as the top-level goal. Figure 3 shows an overview of CARMA's components, including the consultation manager and its tasks, and the reasoning modules and information required to complete the tasks. Sections 1 through 5 of this chapter describe CARMA's use of different reasoning paradigms to implement the process description of entomological problem solving set forth in Chapter 2. The last section summarizes how model-based reasoning and case-based reasoning are integrated in CARMA, and how CARMA emulates the four desirable expert characteristics (*i.e.*, speed, opportunism, graceful degradation, and explanations in terms of a causal model.)

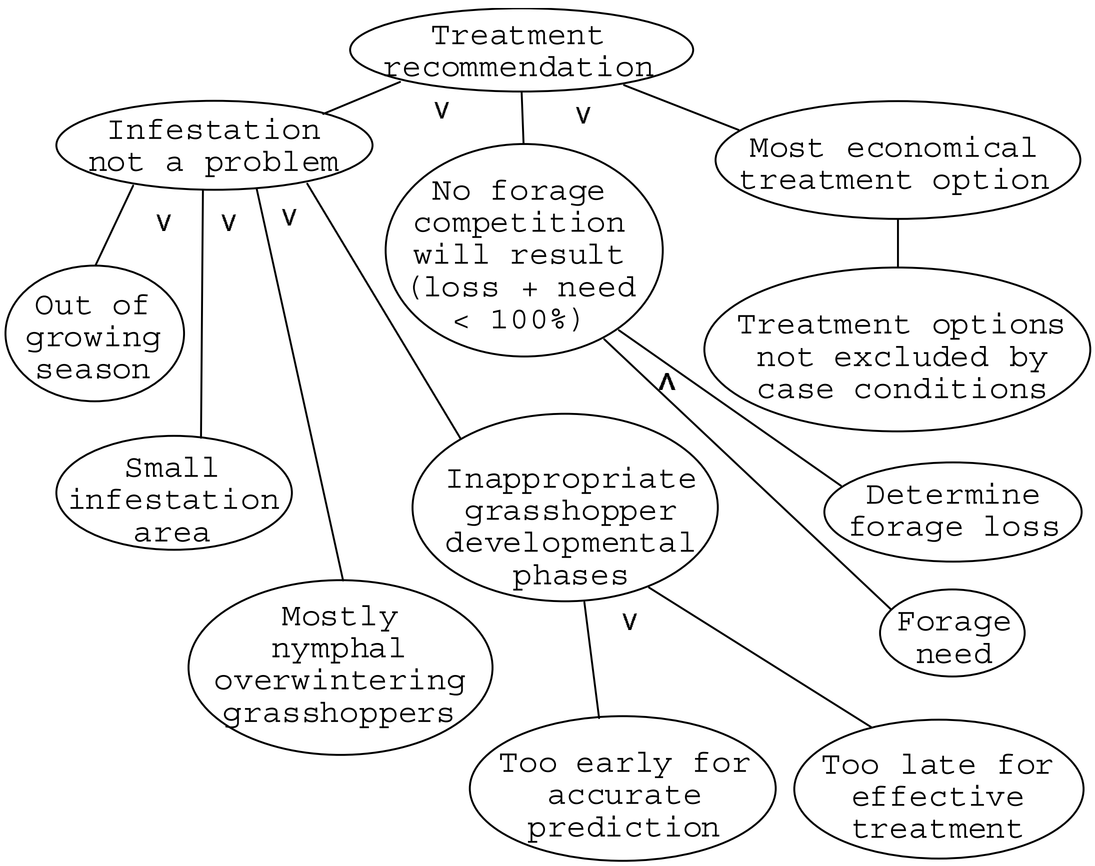

Figure 2: The goal structure that CARMA attempts to satisfy during a consultation. "Treatment recommendation" is the top-level goal. Subgoals separated by ∨'s mean that satisfying any of the subgoals results in satisfying the parent goal. Subgoals separated by ∧'s mean that all subgoals must be satisfied in order to satisfying the parent goal.

3.1 Determining Relevant Case Features

CARMA provides advice by reasoning about the relevant features of an infestation case (*e.g.*, the species, density, and developmental phases of the grasshoppers). These features are inferred by rules from information provided by the user through

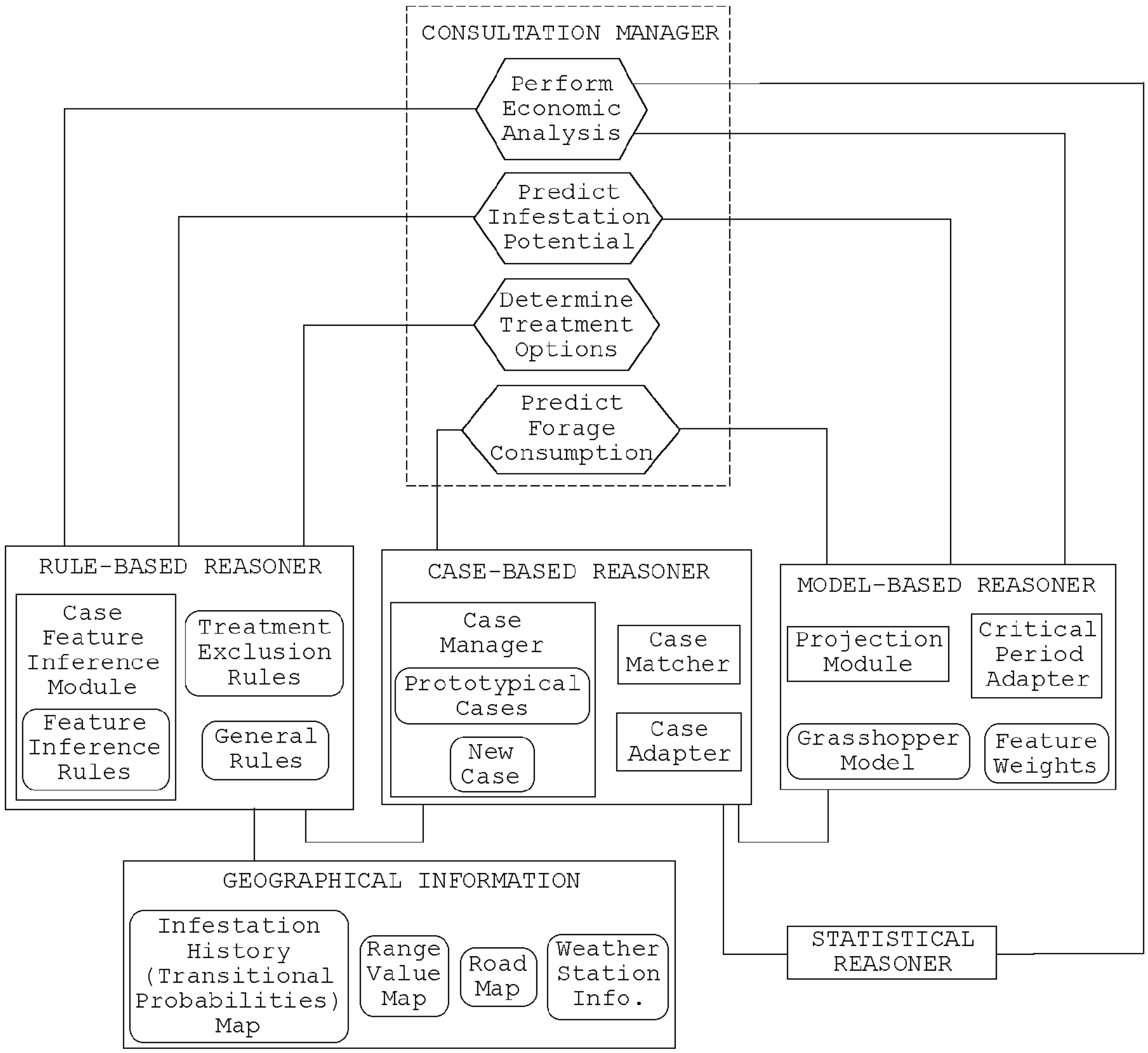

Figure 3: Organizational Overview of CARMA's Components. Hexagons represent tasks, rectangles represent modules, ovals represent information, and lines represent information paths. The order of consultation steps is not shown.

window-based interface procedures. CARMA makes use of multiple levels of rules for inferring each case feature, ordered by the certainty or the accuracy of each rule. The rules are applied in succession until either the user can provide the necessary information or a default value is chosen. For example, if the value of the case feature "total number of grasshoppers per square yard" is unknown to the user, CARMA instructs the user to estimate the number of grasshoppers that would be present in 18 square-foot circles. If the user can't provide this information, the system attempts to infer this feature using a rule that grasshopper density is equal to 1.5 times the number of grasshoppers seen hopping away with each step taken by the user in the field. Otherwise, the value defaults to the statewide historic average of four grasshoppers per square yard. By applying rules in the order of their certainty, CARMA reasons with the best information available.

A typical interface window for determining the observed grasshopper type distribution appears in Figure 4. It includes the options "Why" for describing why this information is important to the consultation, "Help" for advising the user about the various window features and their operations, "How To" to explain the proper procedure for

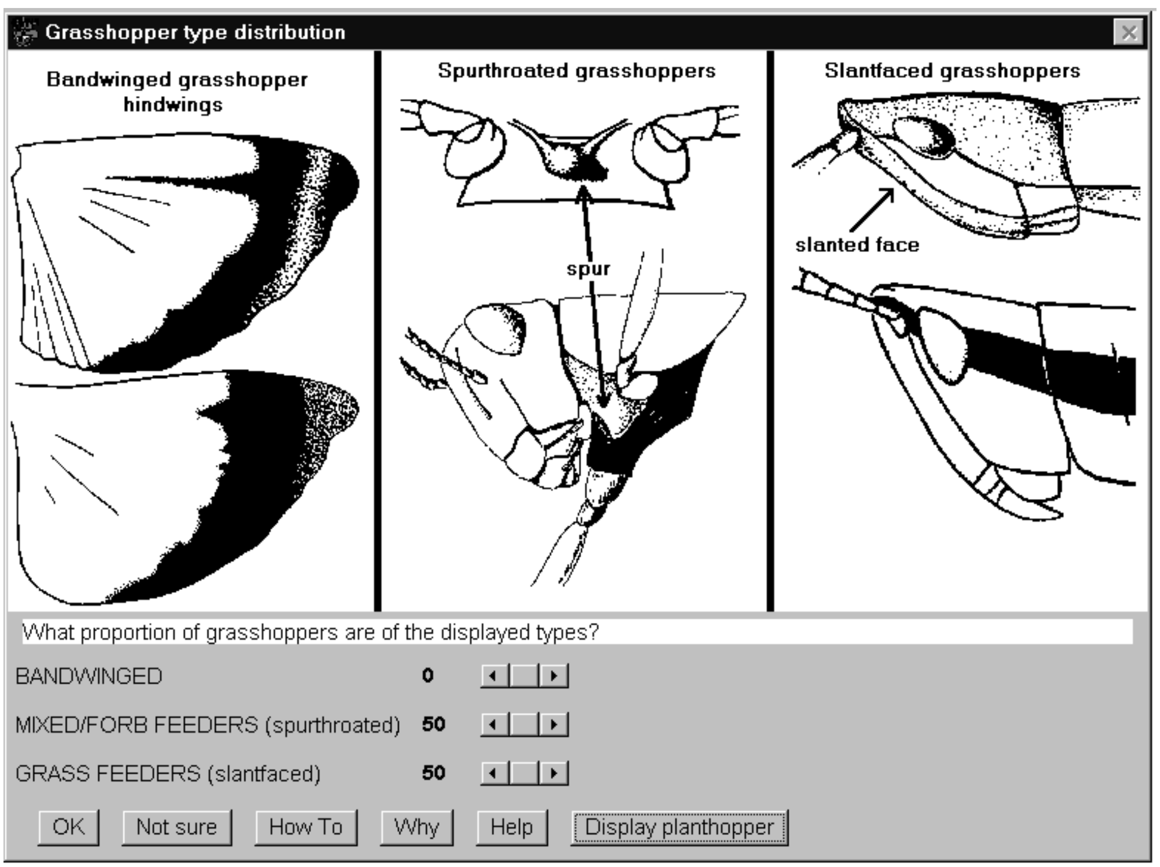

Figure 4: Interface window for determining the observed grasshopper type distribution.

gathering the required information, "Not sure" to trigger the selection of an alternative rule for inferring the feature, and "OK" to indicate that the user has chosen an answer. "Display planthopper" shows a small insect that the user should not accidentally mistake for a grasshopper.

Figure 5 shows an input window that asks the user to provide the infestation location by clicking on a map of Wyoming's major roads, towns, and county borders. CARMA uses this location to retrieve the historical values for the site including infestation history, range value, temperatures, and precipitation.

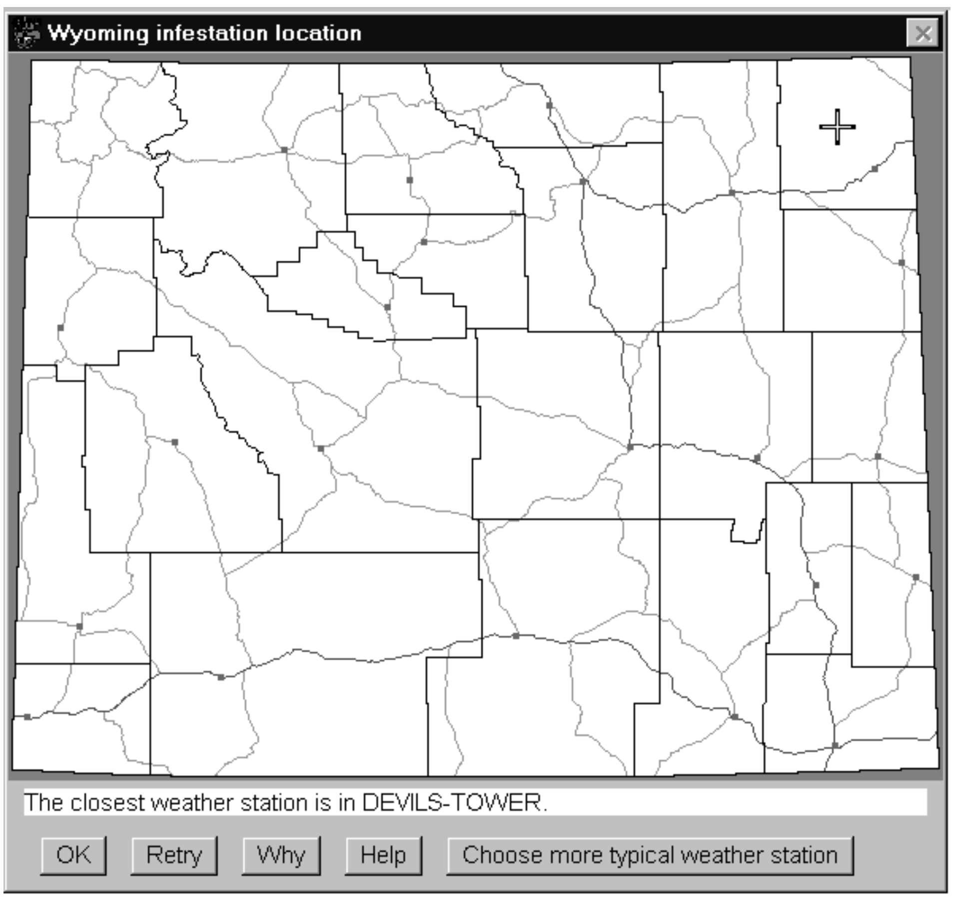

Figure 5: Interface window for determining infestation location.

Since a complete case specification is not always required to give advice, CARMA fills in the facts of a new case opportunistically. This means that CARMA asks the user for information only when the corresponding case feature is required for the reasoning process to continue. At the earliest point at which a decision can be made, the case-feature inference process halts, advice is given, and the consultation is completed. This minimizes the amount of input required for CARMA to make a decision, thereby accelerating consultations. For example, the date and location of an infestation may indicate that it is too early to assess the severity of a grasshopper infestation. In such cases, CARMA advises the user to rerun the consultation at a later time without prompting for further information.

3.2 Determining Infestation Potential

CARMA's first step in advising a rancher about the best response to a grasshopper infestation is deciding whether a potential problem exists. CARMA determines that an infestation is not a problem and terminates a consultation if it discovers any of the following facts:

1. The date is outside of the growing season.
2. The size of the infestation is too small to be viable.
3. The majority of the grasshoppers overwinter as nymphs.
4. The developmental phases of the majority of the grasshoppers are too early for accurate prediction or too late for effective treatment.

The sections that follow describe why these conditions lead to the termination of a consultation, and the methods used to make these decisions.

3.2.1 Outside of the Growing Season

Consumption by grasshoppers is only damaging if it occurs within the growing season of rangeland vegetation when forage needed for livestock grows. If the date of the current infestation is earlier or later than the historical growing season for the area, any grasshoppers that are present will not cause appreciable damage, so no action should be taken. This decision is made by comparing the current date with the historical growing season for the area.

3.2.2 Small Infestation Size

The size of an infestation is an indicator of its viability and hence its future damage potential. A small infestation size indicates either an isolated hatching area or a population with little viability. Infestations smaller than 500 acres are considered unlikely to lead to a significant infestation.

3.2.3 Grasshoppers Overwintering as Nymphs

A tract of rangeland invariably contains multiple grasshopper species. Although virtually all species have only one generation per year, the timing of life-history events and consumption characteristics vary greatly. Specifically, grasshoppers overwintering as nymphs divide their consumption between two growing seasons and consume far less during the growing season than grasshoppers overwintering as eggs. If the majority of the grasshoppers have the former life history, CARMA determines that little forage loss will occur unless densities are extraordinarily high.

To determine the proportion of nymphal-overwintering grasshoppers, CARMA uses **case factoring** to split the overall population of a case into subcases according to life history (*e.g.*, overwintering type). The overall grasshopper population is initially divided into three observed categories: bandwinged (*i.e.*, grasshoppers having brightly-colored wings or make a clicking sound in flight); forb or mixed grass/forb feeders (*i.e.*, grasshoppers having a round head with a spur throat); and grass feeders (*i.e.*, grasshoppers having a slanted face or pointed head, or a round head with no spur throat). If the grasshoppers are part of the bandwinged category, CARMA concludes that the grasshopper population is nymphal-overwintering. Otherwise, the population is determined to be egg-overwintering. For example, the new case set forth in Table 1 is split into two subcases, SubcaseA and SubcaseB, based on life history.

3.2.4 Inappropriate Grasshopper Phases

To provide a meaningful consultation, the majority of grasshoppers must be sufficiently developed to determine the extent of the infestation with some certainty (the infestation potential of a very young population of grasshoppers can fluctuate drastically based on the weather and disease and is therefore much less predictable), but immature enough that a significant proportion of lifetime forage consumption remains, making insecticide application or biological control economically feasible. Rule-based reasoning is used to end consultations involving grasshopper populations whose average developmental phase is less than 2.5 (*i.e.*, too early) or equal to 7.0 (*i.e.*, too late).

3.3 Determining Forage Competition

If a grasshopper infestation is potentially a problem, CARMA estimates forage consumption using a library of prototypical cases. This forage consumption estimate is used to predict whether grasshopper consumption will lead to competition with livestock for available forage.

3.3.1 Prototypical Infestation Cases

The protocol analysis indicated that pest managers estimate forage consumption by comparing new cases to prototypical infestation scenarios. These prototypical cases differ from conventional cases in two important respects. First, the prototypical cases are not expressed in terms of observable features (*e.g.*, "Whenever I take a step, I see four grasshoppers with brightly colored wings fly"), but rather in terms of abstract derived features (*e.g.*, "Approximately six nymphal overwintering grasshoppers in the adult phase per square yard"). Second, the prototypical cases are extended in time, representing the history of a particular grasshopper population over its lifespan. Each prototypical case is therefore represented by a "snapshot" at a particular, representative point in time selected by the entomologist. In general, this representative point is one at which the grasshoppers are at developmental phases in which treatment is feasible. An example prototypical case appears as Case4 in Table 1.

3.3.2 Case-Based Prediction of Forage Consumption

As previously mentioned, CARMA uses case factoring to split the overall population of a case into subcases according to life history (*i.e.*, overwintering as nymphs or eggs). To predict overall forage loss, CARMA totals forage loss predictions for each subcase. The following sections detail how CARMA predicts forage loss by using a

	Case4	New case		Case4 after projection
		SubcaseA	SubcaseB	
Overwintering type	nymph	nymph	egg	nymph
Feeding types	grass 10% mixed 90%	grass 50% mixed 50%	grass 100%	grass 10% mixed 90%
Average phase	2.0	3.0	7.0	3.0
Density	27.0	36.0	4.0	24.0
Proportion of lifetime consumption in critical period	92.7	86.0	12.4	92.7
Date	June 8	June 14		June 15
Precipitation	normal	dry		normal
Temperatures	normal	cool		hot
Infest. history	high	high		moderate
Range value	low	moderately high		low
Total area infested	12000	9800		12000
Forage loss	60% (high)	?		60% (high)

Table 1: Case examples.

causal model to assist case-based reasoning in three different ways: temporal projection; featural adaptation; and critical period adaptation.

3.3.2.1 Temporal Projection

To predict the forage loss of a subcase, CARMA first retrieves all prototypical cases whose life history (*i.e.*, overwintering type) matches that of the subcase. The prototypical case whose weighted featural difference from the new case is least (as described in section 4.1) is selected as the best match. Since prototypical cases are extended in time but are represented at a particular time, CARMA temporally projects the best matching prototypical case forwards or backwards to align its average developmental phase with that of the new subcase. This requires using the model to simulate grasshopper attrition, which depends on developmental phase, precipitation, and developmental rate (which in turn depends on temperature) throughout the interval of the projection. CARMA assumes that the grasshoppers within a developmental phase are evenly distributed throughout the "developmental days" (*e.g.*, normally one week long but adjusted based on temperatures) within that phase. Therefore, CARMA breaks the distribution into daily populations, projects the populations the required number of days (adjusting the density each day based on attrition), then regroups the daily populations into their new developmental phases. Attrition rates are adjusted by scalars (one scalar for precipitation = "wet" and another for precipitation = "non-wet") that are learned via the algorithms described in Section 4.2. A graphic example of temporal projection appears in Figure 6. Further details of temporal projection appear in Appendix A.

For example, the prototypical case that best matches SubcaseA is Case4, as shown in Table 1. Because the developmental phase of Case4 before projection is earlier than that of SubcaseA, Case4 must be projected forwards in time, causing grasshoppers to be removed from the population due to attrition *(i.e.*, 27.0 grasshoppers per square yard before projection to 24.0 grasshoppers per square yard after projection).

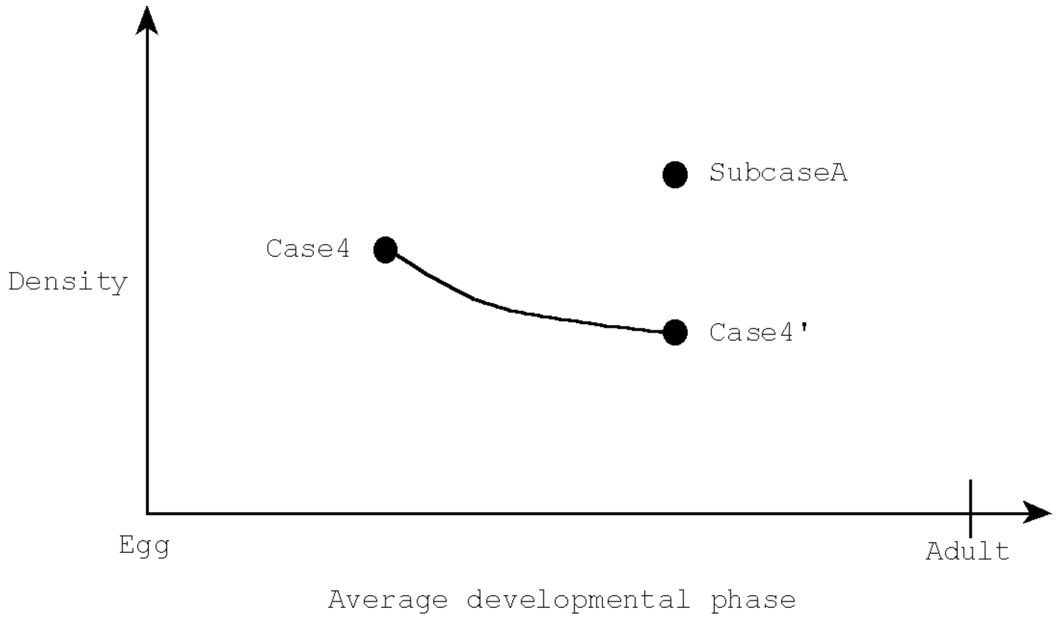

Figure 6: Projection of a prototypical case from Case4 to Case4' to align its average developmental phase with SubcaseA.

Temporal projection aligns developmental phases but not necessarily dates. For example, the date of Case4 after projection is later than the date of SubcaseA due to a number of possibilities, including the hatch date of Case4 was later than that of SubcaseA, or the developmental rate of grasshoppers in Case4 was slower than the rate in SubcaseA. As a result, the average developmental phase of the grasshoppers in SubcaseA on June 14 is the same as that of Case4 one day later on June 15. These dates are used in critical period adaptation, which is described in section 3.3.2.3.

3.3.2.2 Featural Adaptation

The forage loss predicted by the best matching prototypical case, **`FL(PC)`**, is modified to account for any featural differences between it and the subcase based on the influence of each of the **n** features on consumption as represented by a list of featural adaptation weights **A** (*i.e.*, $A_1 ... A_n$). This results in a predicted forage loss for the new subcase, **FL(NC)**, *i.e.*,

$$FL(NC) = FL(PC) + \sum_{i=1}^{n} A_i \times QFD(i)$$

where **QFD(i)** is the quantitative difference for feature **i** between the new subcase and prototypical case . For example, a lower temperature value means lower forage losses, because lower temperatures tend to slow development, increasing grasshopper attrition. Thus, the forage loss estimate predicted by Case4 (60%) must be adapted downward to account for the fact that temperatures in SubcaseA (**cool**) are lower than in Case4 (**normal**). In determining the quantitative feature difference between the new subcase and prototypical case for qualitative features such as temperature, CARMA computes a simple difference, *i.e.*,

$$Q(NC, i) - Q(PC, i)$$

where **Q(NC,i)** and **Q(PC,i)** are the quantitative values for feature **i** in the new subcase and prototypical case, respectively. For quantitative features such as density, proportion of lifetime consumption in the critical period, and total area infested, a proportional difference is used, *i.e.*,

$$\frac{Q(NC,\ i)\ -\ Q(PC,\ i)}{Q(PC,\ i)}\ .$$

Adaptation weights are set using a hill-climbing algorithm that optimizes CARMA's predictive accuracy on training instances (discussed in section 4.2). The weights used in featural adaptation constitute a linear approximation of the function from derived case features to consumption amounts in the neighborhood of each prototypical case.

3.3.2.3 Critical Period Adaptation

As previously mentioned, consumption is only damaging if it occurs during the growing season of a rangeland habitat. However, there is a *critical forage growing period* within the growing season, when forage losses caused by grasshoppers can not be fully replaced by forage growth. The forage loss predicted by a prototypical case must be adapted if the proportion of the lifespan of the grasshoppers overlapping the critical period differs between the new case and the prototypical case. This process, termed *critical period adaptation,* is a specific featural adaptation that requires determining the proportion of lifetime consumption occurring in the critical period based on the developmental phases of the new and prototypical cases that fall within the critical period and the proportion of lifetime consumption occurring in these developmental phases. The forage loss estimate is then adjusted based on the featural adaptation weight for the critical period and the difference in the proportion of lifetime consumption in the critical period between the new and prototypical cases. Details about determining the proportion of lifetime consumption occurring in the critical period appear in Appendix B.

A graphic example of critical period adaptation appears in Figure 7. Because grasshopper development in SubcaseA is ahead of that in Case4 (SubcaseA's developmental phase on June 14 corresponds to Case4's developmental phase on June 15), CARMA determines that Case4 applies to more of the critical period than SubcaseA because it will only reach Day 1 of developmental phase 3 by the beginning of the critical period (June 17), while SubcaseA will already reach Day 8 of developmental phase 3. CARMA uses a model of grasshoppers' rate of consumption at each developmental phase to calculate the proportion of lifetime consumption occurring after the beginning of the critical period and before the end of the critical period. For example, only 86.0% of SubcaseA's consumption occurs during the critical period, whereas 92.7% of Case4's consumption occurs within this period. The quantitative feature difference for critical period adaptation is computed as a proportional difference, therefore CARMA adjusts the initial consumption estimate by (86.0 - 92.7) / 92.7 = -0.072 multiplied by the adaptation feature weight for critical period.

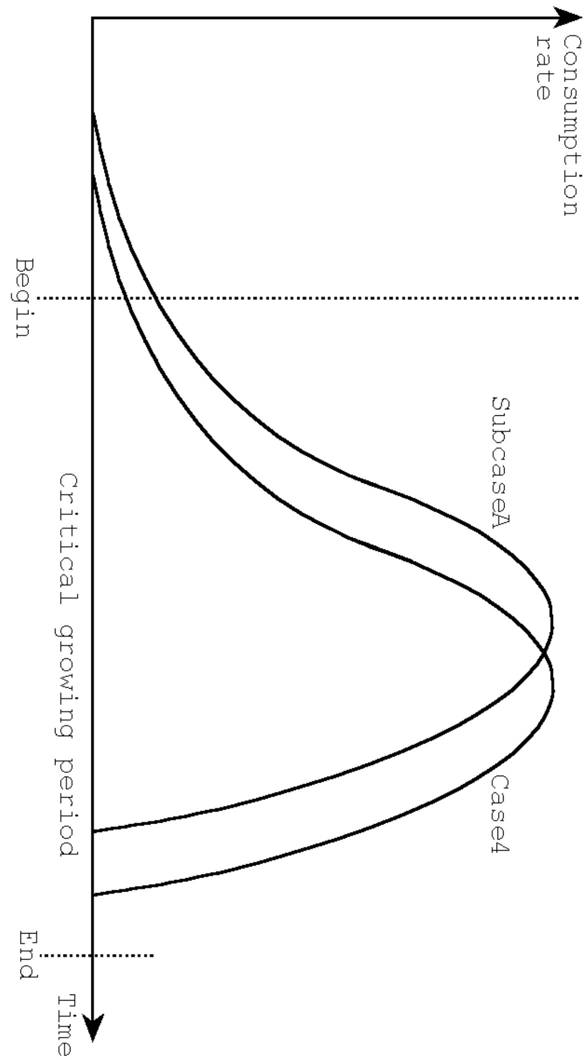

Figure 7: Critical period adaptation from Case4 to SubcaseA.

After adaptation, the consumption predictions for each subcase (*i.e.*, populations of grasshoppers with distinct feeding patterns) are summed to produce an overall consumption estimate. In the given case, the sum of predicted consumption of the two subcases is 90% (86.5+3.4). Because of variability resulting from the imprecise nature of rangeland ecosystems, this prediction is converted to the qualitative range, **`high`**, meaning that approximately 60 to 100% of the available forage will be lost. An interface window explaining estimated forage loss is shown in Figure 8. It gives both aggravating and mitigating factors (*i.e.*, factors tending to increase vs. reduce the forage loss estimate).

The natural language explanation is produced using conventional template instantiation techniques. First, the explanation generator creates the natural language representation of pertinent qualitative feature values using simple lookup tables (*e.g.*, the text string for feature value high-mod is "moderately high"). The text string are then combined with the explanation template. For example, the template for the first sentence in the forage loss explanation is

> <"From the information you have provided, it is estimated that the grasshoppers will consume a" qualitative-forage-loss-string "percentage of the forage available for the year or approximately" quantitative-forage-loss-range-string "%.">

If the proportion of available forage that will be lost to grasshoppers and the

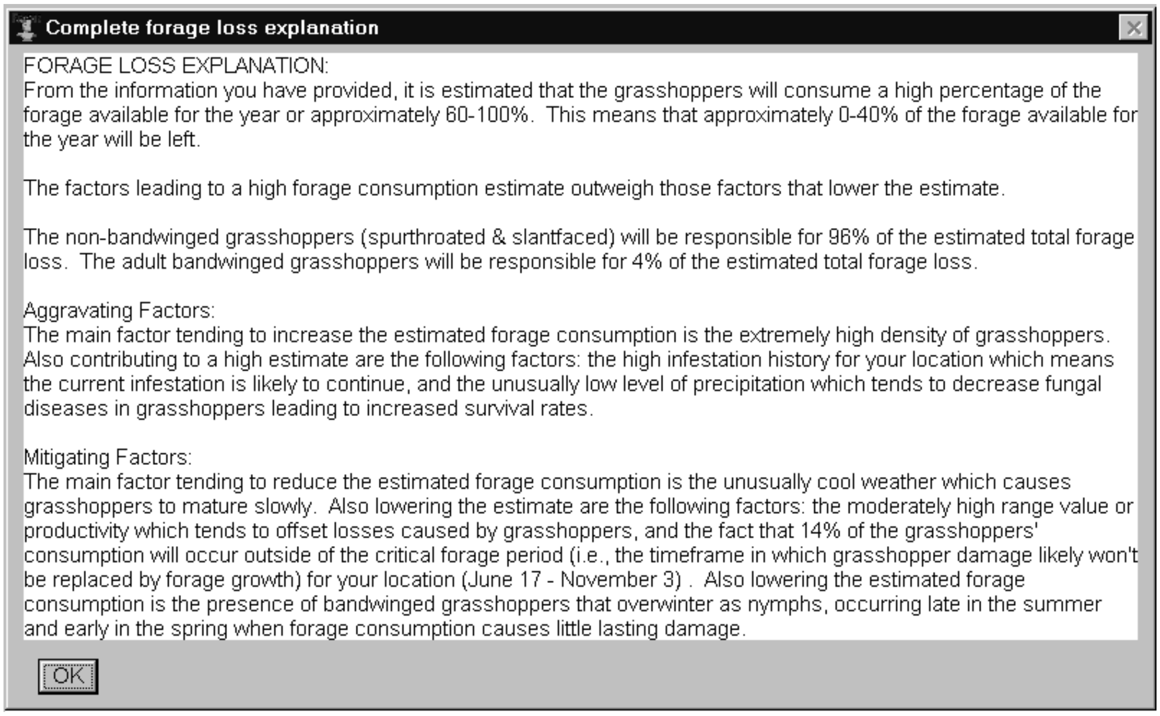

Figure 8: Interface window explaining estimated forage loss.

proportion needed for livestock (and wildlife) exceeds 100% of the forage available, CARMA concludes that competition will occur. In this example, competition is possible and the consultation should continue if the proportion of available forage needed by livestock is greater than 40%. For example, if forage need is 60%, the expected year-long competition should range from 0% (*i.e.*, (40+60)-100) to 20% (*i.e.*, (60+60)-100). A typical interface window explaining estimated forage competition is shown in Figure 9.

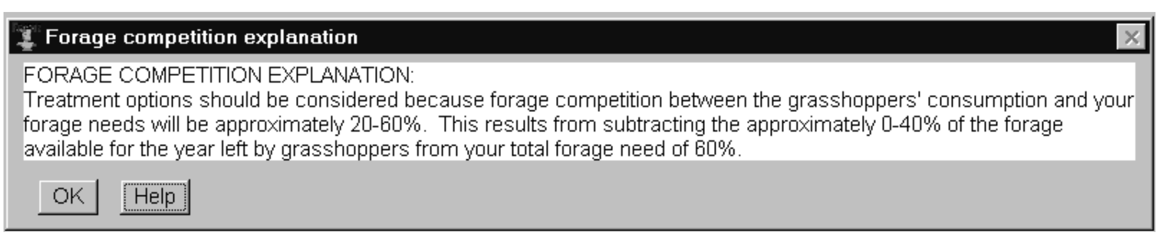

Figure 9: Interface window explaining estimated forage competition.

3.4 Determining Treatment Options

If there will be competition, CARMA applies a set of rules to determine what possible treatment options are excluded by the conditions of the case. Some of the information necessary for determining exclusion is already known from the case features (*e.g.*, the presence of grasshoppers in the first nymphal instar suggests an ongoing hatch, thereby excluding malathion and carbaryl bait from consideration). Other conditions must be determined from further user input (*e.g.*, "Will it be hot at the time of treatment?" If so, exclude malathion.). An interface window explaining the selection of acceptable treatments appears in Figure 10. The explanation includes the rules that were used to exclude treatments. This explanation is also derived using standard template-instantiation techniques.

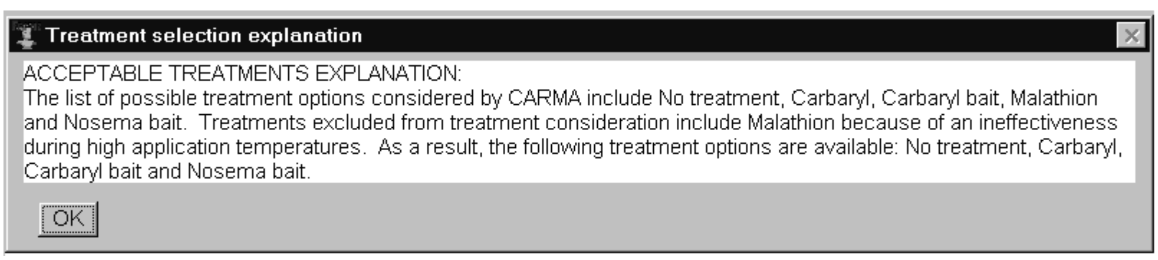

Figure 10: Interface window explaining the selection of acceptable treatments.

3.5 Treatment Recommendation

For each possible treatment option, CARMA provides estimates of the reduced probability of future reinfestation and current-year and long-term savings. From the estimated savings, CARMA recommends the treatment or treatments that are most economical. A typical treatment recommendation window including estimates of future reinfestation and economic savings appears in Figure 11. Notice that this analysis includes "no treatment" as an option.

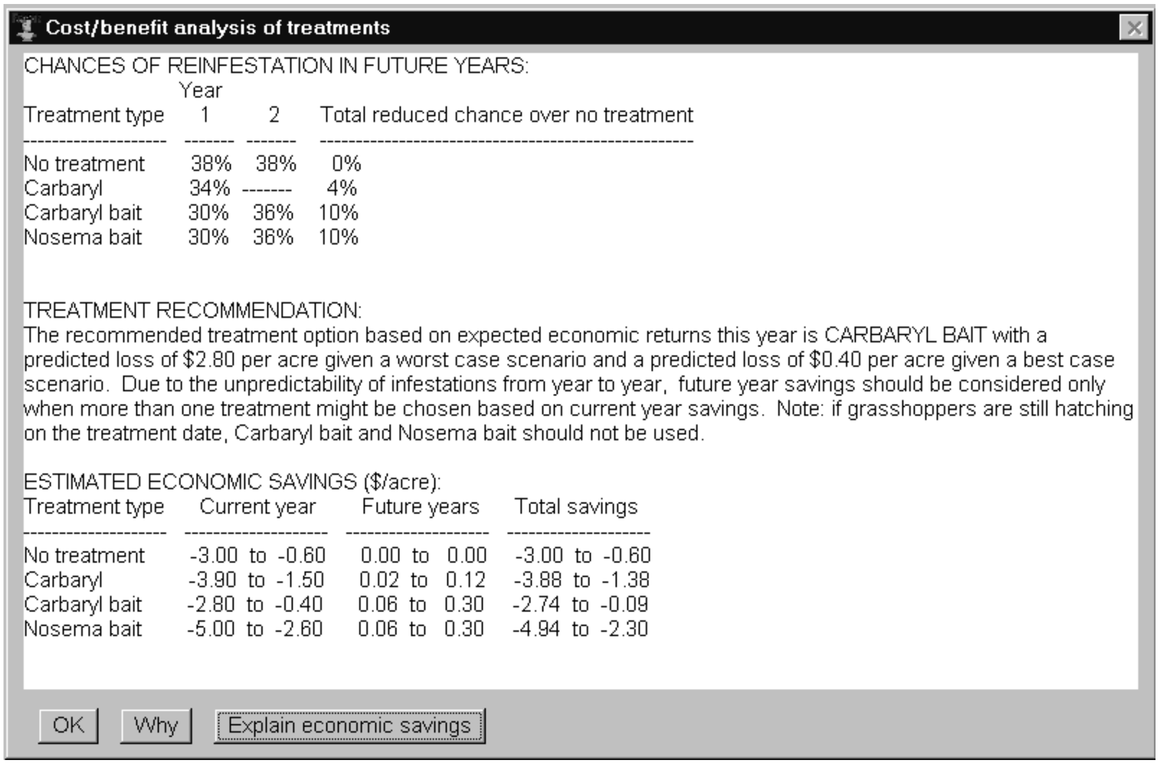

Figure 11: Interface window that recommends a treatment option and provides estimates of future reinfestation and economic savings. For carbaryl the dashed line in its second year reinfestation probability indicates no reduction in infestation probability that year over no treatment.

3.5.1 Reduced Probabilities of Future Reinfestation

CARMA uses statistical reasoning and the historically derived Markov transitional probabilities for the infestation location to calculate for each treatment type the total reduced probability of future reinfestation. First, CARMA determines whether the grasshoppers will begin laying eggs before the treatment date. If the developmental distribution of the grasshoppers at treatment is dominated by adults, CARMA determines that too many eggs will already be laid, and no reduction in the probability of future reinfestation will result from treatment because eggs are not affected by treatment.

If few eggs will have been laid, CARMA calculates the yearly reinfestation probabilities for each treatment type based on the historical Markov transitional probabilities as long as the probability of infestation with treatment for the year is significantly lower that the probability of infestation without treatment (*i.e.*, until the benefits of treatment have ended). The total reduced probability of future reinfestation for each treatment is calculated by summing each yearly difference between the probabilities of infestation without and with treatment.

Because the number of grasshoppers that may emerge in future years is often not directly proportional to the number of eggs laid the current year (*i.e.*, grasshopper densities are dependent on a great number of factors such that, under ideal conditions, grasshoppers are capable of expanding or growing from a low population one year to a very high population the next), transitional probabilities are adjusted only slightly based on the efficacy of treatments in reducing the number of eggs laid. The transitional probabilities are reduced further for those treatments capable of preserving beneficial organisms. For example, treatments such as carbaryl bait are designed to be consumed specifically by grasshoppers and are therefore unlikely to affect biological control agents such as birds and insects. Conversely, sprays such as malathion blanket an entire area and hurt beneficials indiscriminately.

A greater reduction in the transitional probabilities is made for treated infestations whose total area is quite large, because treatment will tend to reduce the chance that grasshoppers from previously untreated areas will migrate into the treated area. More details about determining future infestation probabilities appear in Appendix C.

3.5.2 Economic Analysis

For each possible treatment option, CARMA provides estimates of current-year and long-term savings. Each analysis involves a range that indicates best to worst case estimates (negative values indicate a loss). A typical interface window explaining (*i.e.*, showing a trace of) the savings calculations appears in Figure 12.

3.5.2.1 Current-year Savings

For each possible treatment option, CARMA estimates the current-year savings as

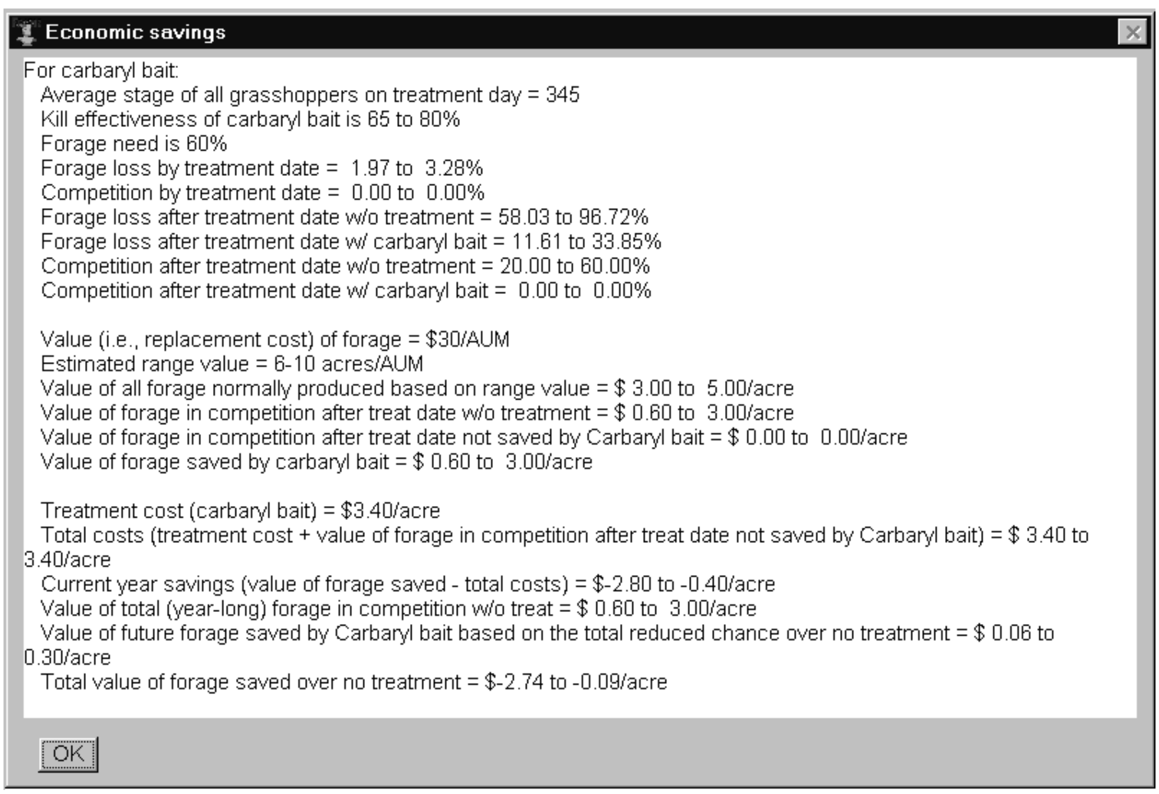

Figure 12: Interface window explaining the economic savings calculations.

the difference between the value of forage in competition saved by treating and the treatment cost. CARMA first computes the amount of pre-treatment forage loss. This is done by temporally projecting the developmental distribution of each subcase forwards to the user-provided treatment date (often a week or more from the current date). In a manner similar to determining the percentage of lifetime consumption occurring within the critical period, CARMA applies a model of grasshoppers' rate of consumption at each developmental phase to each subcase to calculate the proportion of lifetime consumption occurring before the treatment date. This proportion is used to scale the year-long forage loss estimate, resulting in the pre-treatment loss. The pre-treatment forage loss estimates for each subcase are summed to produce the total pre-treatment forage loss. Next, CARMA estimates the amount of post-treatment forage loss without treatment by subtracting pre-treatment forage loss from total forage loss. For example, if total forage loss is estimated to be 60-100%, and pre-treatment forage loss is estimated to be 2.0-3.3% (*i.e.*, approximately 20% of the grasshoppers' total consumption will occur before the treatment date), then the post-treatment forage loss will be 58.0-96.7% (because 80% of the grasshoppers' lifetime consumption must occur after the treatment date).

For each option, CARMA estimates the amount of post-treatment forage loss with treatment according to the expected efficacy of the treatment and the post-treatment forage loss without treatment. For example, the insecticide carbaryl bait is usually 65 to 80% effective. If the estimated post-treatment forage loss without treatment is 58.0-96.7%, then at best carbaryl bait should prevent 80% of the 58.0% loss, and at worst prevent 65% of the 96.7% loss, resulting in a 11.6 to 33.9% post-treatment forage loss.

CARMA calculates the year-long forage loss for each option by summing pre- and post-treatment forage loss. Year-long competition resulting from a treatment option is calculated by comparing year-long forage loss resulting from the option and forage need. The proportion of forage in competition saved is simply the proportion of forage in competition without treatment minus the proportion of forage in competition with treatment. For example, if pre-treatment forage loss is 2.0-3.3% and post-treatment forage loss is 11.6-33.9%, the year-long forage loss for the option is 13.6-37.2%. Given a forage need of 60%, the year-long competition with treatment ranges from (13.6+60)-100 = -26.4 to (37.2+60)-100 = -2.8, which is less than zero, resulting in no competition. If the year-long forage in competition without treatment is 20-60%, and the treatment option will result in no competition, then the expected forage in competition saved by treating is 20-60%.

With the per-unit forage value and range value estimates provided by the user, CARMA estimates the current-year savings for an option to be the value of forage in competition that is saved minus the cost of the treatment. In this example, the per-unit forage value is $30/AUM (*i.e.*, an animal unit month - the amount of forage necessary to support a cow and calf for one month) and the estimated range value (or productivity) is 6-10 acres/AUM. Therefore, the current-year savings ranges from:

$$20\% \times \frac{\$30}{AUM} \times \frac{AUM}{10\ \ acres} = \frac{\$0.60}{acre}$$

to

$$60\% \times \frac{\$30}{AUM} \times \frac{AUM}{6\ \ acres} = \frac{\$3.00}{acre}\ .$$

3.5.2.2 Long-term Savings

CARMA calculates the savings for future years for each treatment type as the value of year-long (*i.e.*, total) forage in competition without treatment (taken from the first year calculations) times the total reduced probabilities of future reinfestation.

Based on the current year savings, CARMA recommends the treatment that is estimated to save the most under a worst case scenario and the treatment that is estimated to save the most under a best case scenario. Usually, the worst and best scenarios produce the same recommended treatment. Following the treatment recommendation, the consultation is complete.

3.6 Multiple-Paradigm Reasoning in CARMA

CARMA implements the process description of entomological problem-solving by combining a variety of distinct reasoning paradigms. In particular, CARMA uses model-based reasoning in three different ways to assist case-based reasoning for the purpose of predicting forage loss. First, a model of grasshopper attrition is used in temporal projection to simulate the attrition that would have occurred during the interval between the developmental distributions of the new case and the prototypical case. Second, featural adaptation constitutes a linear approximation of the function from derived case features to consumption amounts in the neighborhood of each prototypical case. Finally, critical period adaptation modifies the prediction estimate to take account of any difference in overlap between grasshopper lifespans and the critical forage growing season.

CARMA's implementation of the process model emulates the four characteristics of human expert performance mentioned above: speed, opportunism, graceful degradation, and causal explanations. CARMA is fast because, like a human expert, it can use compiled knowledge in the form of cases and rules, rather than relying entirely on computation-intensive simulations. CARMA is opportunistic in that it can recognize when no more information is required from the user (*e.g.*, when no accurate prediction can be made, or when it is too late in the season for treatment to be economical).

Graceful degradation is achieved by CARMA in two ways. First, CARMA uses multiple levels of rules ordered by certainty to infer case features. Thus, precise information can be used if available, but the absence or inaccuracy of the information does not cause a catastrophic fall-off in accuracy. Second, CARMA's use of CBR (*i.e.*, case-based reasoning) means that incrementally less precise information will lead to incrementally less accurate matching and adaptation, but not a catastrophic inability to provide plausible predictions and advice.

Finally, CARMA is capable of providing causal explanations, notwithstanding its use of CBR, based on a knowledge of what constitutes variations from normal rangeland conditions. CARMA orders features based on their importance to forage consumption and the magnitude by which they vary from normal. An explanation is generated by describing the features that most aggravate and mitigate forage loss using standard template-instantiation techniques.

Chapter 4

Learning Match and Adaptation Weights

CARMA uses two sets of weights in case-based reasoning: match weights (used to assess the similarity between cases) and featural adaptation weights (used to adapt the consumption predicted by the best matching prototypical case in light of any featural differences between it and the subcase). General domain knowledge, such as the identifying characteristics and developmental phases of grasshoppers, can be provided by the domain expert. By contrast, match and featural adaptation weights must be acquired by the system itself.

4.1 Match Weights

Match weights are set by determining the *mutual information gain* between case features and qualitative consumption categories in a given set of training cases, since recent research has indicated that this is often the most accurate measure of featural importance for matching (Wettschereck & Dietterich 1995). Separate match weights are computed for each grasshopper overwintering type for seven case features:

`precipitation`, `temperature`, `range value`, `infestation history`, `average developmental phase`, `density`, and `feeding type`.

Quantitative features, such as `density`, are converted to qualitative values for computation of mutual information gain, since small quantitative variations seemed to have little effect on matching. The matching feature difference between two individual feature values is determined by finding the difference between the positions of the values in an ordered qualitative feature value list. For example, range value can equal one of the qualitative values in the ordered set {`low`, `low-moderate`, `moderate`, `high-moderate`, and `high`}, so that the matching feature difference between `low` and `high` is four, the maximum possible difference. The similarity of two cases is determined by summing each individual feature difference multiplied by the corresponding match weight.

4.2 Adaptation Weights

Featural adaptation weights are set by a hill-climbing algorithm, `AdaptWeights`, that incrementally varies the list of adaptation weights **`A`** to minimize the *root-mean-squared error* (RMSE), *i.e.*,

$$\sqrt{\frac{1}{n} \times \sum_{i=1}^{n} [PFL\,(C_i,P,M,A) - ExpertPred\ \ (C_i)]^2}$$

for prototypical case library P and match weights M, where `PFL`(C_i, P, M, A) is

CARMA's predicted forage loss and `ExpertPred`(C_i) is an expert's prediction of consumption for each training case C_i in the training set T. The algorithm for **`AdaptWeights`** is as follows:

function **AdaptWeights** (T, P, M)

$I =$ *initial increment (or adjustment value)*

$D_{\min} =$ *minimum improvement value*

$\boldsymbol{A} =$ *initial list of global adaptation weights*

$D' = RMSE(T, P, M, \boldsymbol{A})$

$D = \infty$

loop until $(I < I_{\min})$ ***do***

    loop until $(|D' - D| < D_{\min})$ ***do***

        $D = D'$

        $\boldsymbol{A}' = \boldsymbol{A}$ *with the element* A_i *of* $\boldsymbol{A}$ *for which* RMSE $(T, P, M, \boldsymbol{A})$

            is least if A_i *is replaced by* $(A_i \pm I)$

        $D' =$ RMSE $(T, P, M, \boldsymbol{A}')$

        if $(D' < D)$ ***then*** $A = A'$

        else $D' = D$

    $I = I \div 2$

return $\boldsymbol{A}$

Separate adaptation weights are computed for each grasshopper overwintering type for eight case features: `precipitation`, `temperature`, `range value`, `infestation history`, `average developmental phase`, `density`, `feeding type`, `proportion of lifetime consumption in the critical period`, and `total area infested`. CARMA can learn featural adaptation weights in either of two modes: *global*, in which a single set of weights are acquired for the entire case library; or *case-specific*, in which separate weights are acquired for each prototypical case.

In computing the featural adaptation weights, qualitative case features (*e.g.*, precipitation = "Dry") are converted into quantitative values based on the position of the value in an ordered qualitative feature value list. An adaptation feature difference is computed as the difference between the quantitative feature values of the two cases. The consumption prediction of the matching prototypical case is adjusted by the sum of the adaptation feature differences multiplied by the adaptation weights for each feature.

Chapter 5

Evaluation of CARMA

The design of CARMA's forage consumption prediction component is based on the hypothesis that integrating model-based and case-based reasoning can lead to more accurate forage consumption predictions than the use of either technique individually. This hypothesis is based on the observation that neither the causal model nor the empirical data available for rangelands are individually sufficient for accurate prediction. To test this hypothesis, the configuration of CARMA with the highest predictive accuracy had to be determined. Sections 1 through 3 detail tests of CARMA and the subsequent modifications made to CARMA based on these findings that lead to the final configuration. The testing of the integration of model-based and case-based reasoning is described in section 4 and discussed in section 5. A comparison of case-specific and global adaptation weights is given in section 6. CARMA demonstrates that integrating various reasoning paradigms can lead to a useful advising system. CARMA's advice is evaluated in section 7 by comparing it to the advice given by Wyoming entomologists and pest managers. Section 8 summarizes the findings.

5.1 The Initial System Configuration ($CARMA_A$)

CARMA was originally implemented as a purely case-based reasoning system using the PROTOS case-based reasoning shell (Porter, Bareiss, & Holte 1990). However, this implementation proved to be a poor model of expert problem solving in this domain. PROTOS is designed to produce a diagnostic category as a solution. However, the protocol analysis indicated that a solution should consist of a treatment recommendation supported by an explanation in terms of causal, economic, and pragmatic factors, including a numerical estimate of the proportion of forage consumed and a cost-benefit analysis of the various treatment options. The rule-based and model-based steps of expert problem solving set forth in Chapter 2, which are necessary for such solutions of this nature, can't be accommodated within a purely case-based reasoning approach. The focus of the CARMA project therefore turned to integrating the multiple problem solving paradigms used by human experts.

Since the protocol analysis indicated that pest managers estimate forage consumption by comparing new cases to prototypical infestation scenarios, a set of prototypical cases was elicited from an entomologist who participated in the protocol analysis. An initial set of prototypical cases was obtained by asking the expert what stereotypical situations were used as a standard for comparison with the problem situations addressed by the expert in the protocol analysis. Additional prototypical cases were obtained by presenting the expert with a wide range of artificial problems and asking the expert to identify stereotypical situations that would be most relevant to forage consumption predictions in those situations. The resulting prototypical case library

(ProtoL) contained eight cases - two nymphal overwintering cases and six egg overwintering cases. A separate experimental set, Set1, consisted of 15 cases generated by the same expert as ProtoL. These initial cases contained qualitative rather than quantitative forage loss predictions. This initial configuration of CARMA is termed *CARMA*$_A$.

5.1.1 Tests of CARMA$_A$

CARMA$_A$ was initially tested to estimate the performance of its consumption prediction module and to determine the relative contributions of three model-based adaptation techniques (*i.e.*, featural adaptation, temporal projection, and *critical period adjustment*[4]) to predictive accuracy. The evaluation of CARMA$_A$ was complicated by the absence of empirical data against which to measure CARMA$_A$'s predictions. Therefore, expert human judgments were used as an external standard. Set2 containing 48 test cases was thus created with randomly generated features and forage loss predictions estimated by a second entomologist.

To determine the contribution of model-based knowledge, an ablation study was performed in which the full CARMA$_A$ consumption prediction module was compared to CARMA$_A$ minus the model-based components and to two different inductive methods: decision-tree induction using ID3 (Quinlan 1986) and linear approximation using QR

---

[4]Critical period adjustment was initially an adaptation technique separate from featural adaptation that received no benefit from training. It was equivalent to critical period adaptation with a featural adaptation weight of 1.0 such that any forage loss outside of the critical period was discounted.

factorization (Hager 1988) to find a least-squares fit to the feature values and associated predictions of the training cases.

CARMA$_A$ was tested using ProtoL as its case library. CARMA$_A$'s global feature weights, used both in case matching and in adaptation, were tuned using an early version of the `AdaptWeights` hill-climbing algorithm to optimize CARMA$_A$'s overall predictive accuracy through the ProtoL case library on Set1. The ablated versions of CARMA$_A$ used the same global feature weights and case library as the full system. ID3 and linear approximation were given ProtoL and Set1 as training instances.

The accuracy of each approach was tested by comparing its forage loss prediction for each case in Set2 with the prediction of the expert. The qualitative difference between two forage loss predictions was calculated as the number of categories by which the predictions differ in the ordered set {`low`, `low-moderate`, `moderate`, `high-moderate`, and `high`}, so that `low` differs from `high` by four categories, the maximum possible qualitative difference. The results, which appear in Table 2, include the mean qualitative error per test case (i.e., the mean qualitative difference between the prediction of the approach and the expert over all the test cases) and the estimated mean quantitative error per test case[5] (based on a scale of 0% to 100% for quantitative forage loss predictions). The performance of CARMA$_A$ is shown in row two. Rows three and four show, respectively, CARMA$_A$ with featural adaptation removed (CARMA$_A$ - CPA) and CARMA$_A$ with critical period adjustment and temporal projection removed

---

[5]Estimated mean quantitative error is a conservative estimate based on a conversion of Set2 forage loss predictions from qualitative to quantitative values.

(CARMA$_A$ - CPA, P). The performance of *factored nearest-neighbor prediction* (factored-NN), *i.e.*, CARMA$_A$ with projection, featural adaptation, and critical period adjustment removed,[6] is shown in row five.

	Mean qualitative error	Estimated mean quantitative error
CARMA$_A$	0.42	14.9
CARMA$_A$ - FA	0.79	21.6
CARMA$_A$ - CPA, P	0.83	22.3
CARMA$_A$ - FA, CPA, P (factored-NN)	0.85	22.0
ID3	1.00	25.2
Linear approximation	1.15	30.1

Table 2: Summary of CARMA$_A$ test results. P, FA, and CPA represent temporal projection, featural adaptation, and critical period adjustment, respectively.

CARMA$_A$'s average qualitative error was 0.42. Removal of the various model-based adaptation techniques significantly degraded CARMA$_A$'s performance. CARMA$_A$'s error rate was almost doubled by removal of featural adaptation (0.79), removal of critical period adjustment and temporal projection (0.83), or by removal of all three (0.85). CARMA$_A$ was not tested with case factoring disabled. However, ID3's performance on unfactored cases, 1.00, was lower than CARMA$_A$'s performance with all model-based

[6]Under this approach, cases are first factored into populations with distinct overwintering types, nearest-neighbor prediction (1-NN) as described in Cover & Hart (1967) is performed for each population, and the resulting consumption predictions for all populations are summed.

reasoning other than case factoring disabled, suggesting that case factoring is also an important requirement for performance in this domain.

Featural adaptation assumes that the function for forage consumption can be approximated by a linear equation in the neighborhood of prototypical cases. Given the large contribution of featural adaptation to $CARMA_A$'s performance, it seems reasonable to wonder whether the forage consumption function can be globally approximated by a linear equation. However, the performance of linear approximation (1.15) indicates that a linear function for consumption as a function of case features is a poor predictor.

5.1.2 Initial findings

The most important weakness of the $CARMA_A$ implementation of forage consumption prediction module was that it used a single set of global feature weights for both matching and featural adaptation. Even if the consumption function can be approximated by a linear function in the neighborhood of prototypical cases, as assumed in featural adaptation, it doesn't follow that the same linear function is appropriate for all prototypical cases. Indeed, the observed poor performance of global linear approximation, shown in Table 2, suggested that linear approximations, and therefore feature weights, should be specific to individual prototypical cases. Moreover, while it is plausible that feature weights for matching should be the same as feature weights for adaptation, this hypothesis needed to be tested. Thus, two important research issues were to compare case-specific and global featural adaptation weights and to test the effect of separating match and adaptation weights.

Another important limitation of this initial evaluation was that the consumption predictions associated with Set2 were produced by a different entomologist than the entomologist from whom the prototypical cases (ProtoL) and Set1 were elicited. As a result, there may have been inconsistencies between the testing set and the library of prototypical cases.

5.2 Greater Weighting Flexibility and Multiple Experts ($CARMA_B$)

Based on the concerns from testing $CARMA_A$, match and adaptation weights were separated. This new configuration of CARMA was called *$CARMA_B$*. As with the initial tests, expert human judgments were used as an external standard against which to measure $CARMA_B$'s predictions. A complication introduced by the use of expert human judgments as an evaluation standard is the possibility that in making consumption predictions human experts fail to use of all aspects of the model of grassland ecology. To test this possibility, we performed an ablation study similar to the tests in section 5.1 in which we tested the effect on prediction accuracy of removing each form of adaptation knowledge from $CARMA_B$.

To obtain a representative sample of expert opinions, questionnaires were sent to 20 entomologists and pest managers recognized for their work in the area of grasshopper control. Each expert received 10 hypothetical cases (located at the northern Wyoming border) randomly selected from a complete set of 20 cases. The descriptions of the 20 cases contained at least as much information as is typically available to an entomologist from a rancher seeking advice. The questionnaire asked the expert to make several predictions about the case, including the predicted quantitative forage loss. A total of 13 recipients of the questionnaire responded. The resulting experimental case sets consisted of 13 sets of expert responses containing 10 cases each (the *Expert Sets*), and eight sets filled in by Wyoming experts containing 10 cases each (the *Wyoming Expert Sets*).

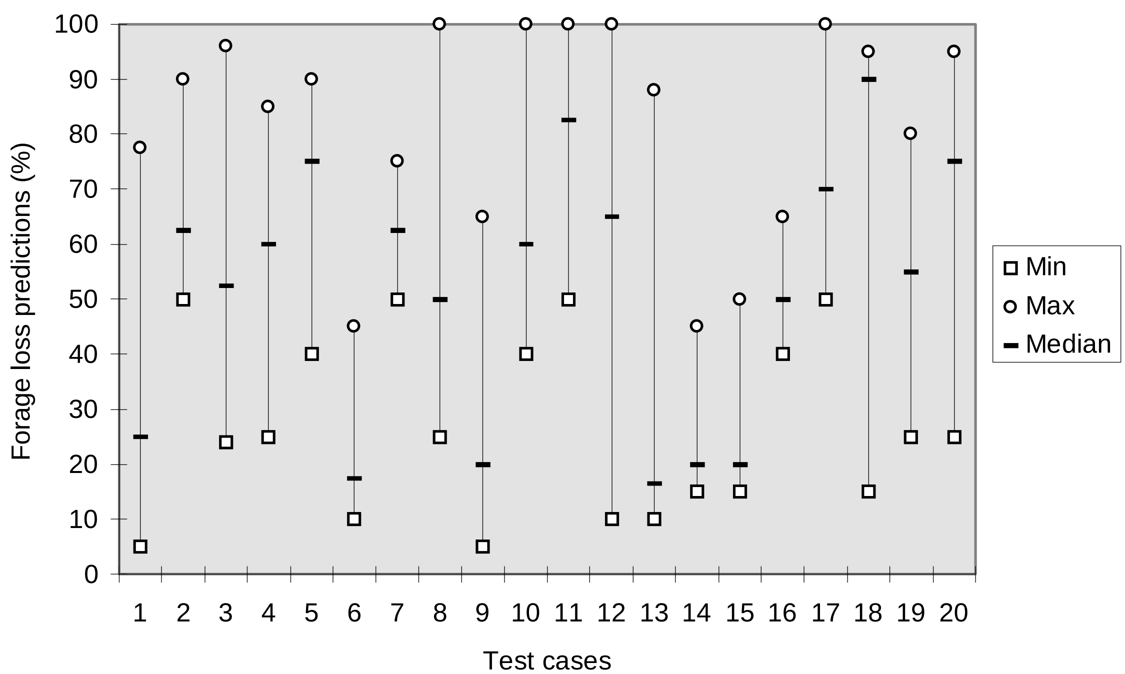

Figure 13: Range of forage loss predictions by 13 experts on 20 test cases.

5.2.1 Expert Prediction Variation

Suprisingly, there was a very wide variation in consumption predictions among the experts over the set of 20 cases. Figures 13 and 14 show the variation in consumption predictions among the 13 experts (from 25 to 90%) and 8 Wyoming experts (from 5 to

70%), respectively. As expected, the eight Wyoming experts were more consistent in their forage consumption predictions. However, while the predictions of the eight Wyoming experts are not as widely ranging as the entire set of 13 experts, it was not determined whether this reduced variation resulted solely from a smaller set size (i.e., other combinations of eight experts would result in less variation than the Wyoming experts), or that a greater consensus results from Wyoming experts familiar with Wyoming cases. Nonetheless, the very wide variation in consumption predictions among the experts furthered concerns (from $CARMA_A$) that inconsistencies (or inaccuracies) are likely when CARMA is trained on the cases predicted by one expert but tested on cases from another expert.

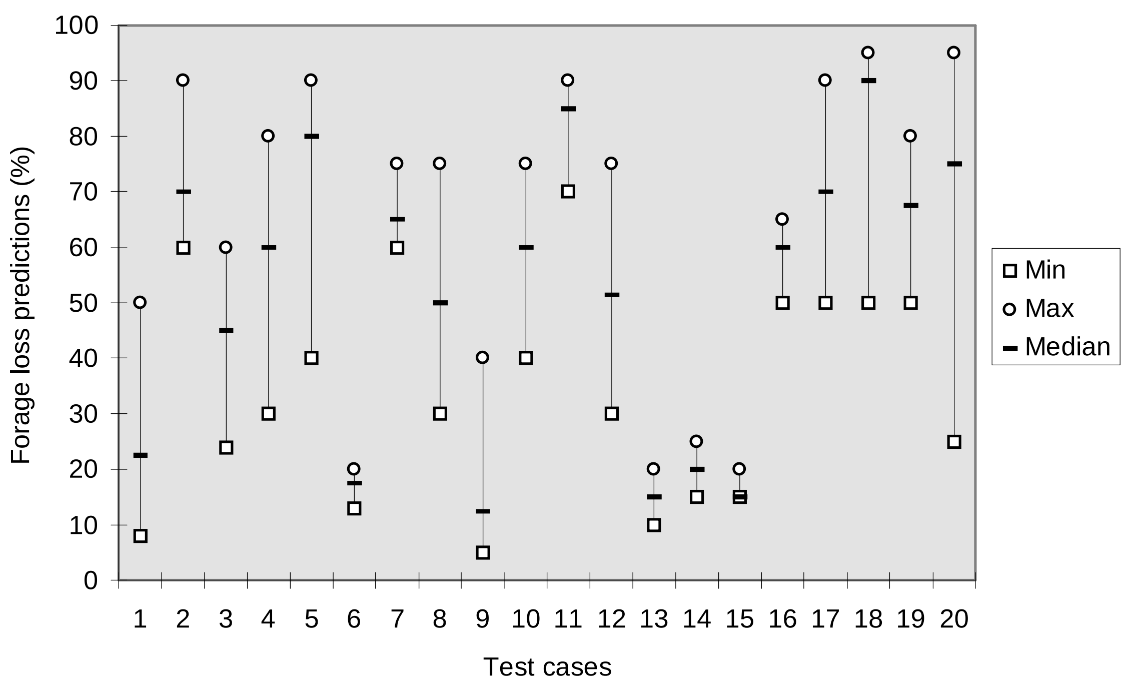

Figure 14: Range of forage loss predictions by eight Wyoming experts on 20 test cases.

5.2.2 Experimental Design

In an effort to avoid inconsistencies between training and testing sets, each predictive method was tested using a series of leave-one-out tests in which a set of cases (S) from a single expert was split into one *test case* (C) and one *training set* (S - C). The methods were trained on the forage loss predictions of the training set and tested on the test case. This method was repeated for each case within the set (S). The forage loss predictions (between 0% and 100%) represent the proportion of available forage that would otherwise be available for livestock but will instead be consumed by grasshoppers. CARMA_B was tested using a protocol under which each set of training cases is used as CARMA's library of prototypical cases. This protocol is implemented in `LeaveOneOutSpecificTest` and `LeaveOneOutGlobalTest`, which perform the leave-one-out tests for the specific and global adaptation weights schemes, respectively. Both procedures call `AdaptWeights`, the hill-climbing algorithm described above. `LeaveOneOutSpecificTest` calls `AdaptWeights` with a prototypical case library containing only one case.

```
function LeaveOneOutSpecificTest (S)
  for each case Ci ∈ S do
    P = S − Ci                ;prototypical cases
    M = list of global match weights for set P according to info.gain
    for each prototypical case Pj ∈ P do
      T = P − Pj              ;training set
      A(Pj) = AdaptWeights (T,{Pj},M)
    Di = (PredictForageLoss (Ci,P,M,A) − ExpertPred (Ci))^2
  return √Avg (D)
```

function **LeavOneOutGlobalTest** (S)

for *each case* $C_i \in S$ ***do***

$\quad P = S - C_i$ *;prototypical cases*

$\quad M =$ *global match weights for set* P *according to info .gain*

$\quad A =$ **AdaptWeights** (P, P, M)

$\quad D_i =$ (**PredictForageLoss** (C_i, P, M, A) $-$ **ExpertPred** $(C_i))^2$

return $\sqrt{\text{Avg}\,(D)}$

5.2.3 Ablation Experiment

As with CARMA_A to determine the contribution of the various forms of model-based adaptation to CARMA_B's predictive accuracy, an ablation experiment was conducted in which the performance of the full CARMA_B system was compared to CARMA_B's performance with various adaptation mechanisms disabled on both the Expert Sets and Wyoming Expert Sets. The second column of Table 3 shows CARMA_B's average root-mean-squared error using case specific weights (CARMA_B-specific). Columns three and four show CARMA_B-specific with, respectively, projection and critical period adjustment removed, and column five shows CARMA_B with featural adaptation removed. The performance of factored nearest-neighbor prediction (factored-NN) is shown in column six.

	Specific weights			No featural adaptation		Global weights	
	Full	Minus projection	minus CPA	minus featural adaptation	minus FA, P, CPA (factored-NN)	Full	minus CPA
Expert sets	22.6	21.3	17.3	29.8	20.5	23.2	18.7
Wyoming expert sets	24.3	22.5	17.6	30.1	21.2	27.1	21.1

Table 3: CARMA_B's average percentage root-mean-squared error across 13 Expert Sets and eight Wyoming Expert Sets with various adaptation methods removed.

These data show that full $CARMA_B$-specific actually performs worse than factored-NN on both the Expert Sets and Wyoming Expert Sets. Removing featural adaptation makes performance still worse, while removing projection improves $CARMA_B$'s performance beyond the full configuration. Removing critical period adjustment is necessary before $CARMA_B$ performs better than factored-NN.

Columns seven and eight show $CARMA_B$ using global weights ($CARMA_B$-global). As with $CARMA_B$-specific, $CARMA_B$-global was more accurate with critical period adjustment removed. However, $CARMA_B$-global minus critical period adjustment, while more accurate than NN, is less accurate than $CARMA_B$-specific with critical period adjustment removed. Each configuration of $CARMA_B$ was more accurate on the Expert Sets than on the Wyoming Expert Sets.

In summary, the ablation experiment tentatively showed that case-specific adaptation weights led to better performance than global adaptation weights. The experiment also showed that featural adaptation increased predictive accuracy, but projection and critical period adjustment decreased accuracy. This suggested that projection and critical period adjustment do not accurately reflect the problem-solving behavior of human experts in this predictive task. However, strong arguments could be made for the inclusion of both of these model-based techniques. First, the expert who generated the initial prototypical case library (ProtoL) continued to claim that critical period adjustment (*i.e.*, discounting forage loss outside of the critical period) was a valid technique. Second, removing temporal projection from CARMA would create severe implementation difficulties, because it would require a much larger prototypical case library so that the developmental phase of a new case would always be properly matched. Simply discounting developmental phases would be quite counterintuitive for the following reason. It was already known that grasshopper density is typically the most important case feature. However, the importance of density is not completely independent of the developmental phase. For example, it is known that 30 grasshoppers in the third nymphal instar are much more potentially damaging than 30 grasshoppers in the first instar since grasshoppers in the third instar consume more forage due to their larger size and are less likely to be lost due to attrition. In this example it is important to know that grasshopper densities decrease over time as grasshoppers progress through each developmental phase.

The fact that $CARMA_B$ is more accurate on the Expert Sets than the Wyoming Expert Sets is also peculiar. A reasonable assumption is that the best predictions about Wyoming grasshopper infestations come from experts within Wyoming as they typically have the most experience with these infestations. Since the goal of the CARMA project is provide the best advice about Wyoming grasshopper infestations, $CARMA_B$ clearly should be tailored to emulate the predictions of the Wyoming experts rather than the entire set of experts. In addition, since many model-based parameters within $CARMA_B$ are based on conditions typical within Wyoming (*i.e.*, conditions most familiar to Wyoming experts) it seems that $CARMA_B$'s predictive component is initially biased (before training) towards Wyoming experts, and would be most likely through training to resemble their predictions than those of other experts. $CARMA_B$'s inaccuracy on the Wyoming Expert Sets is perhaps just a side effect of an incorrectly configured forage consumption prediction module.

Unfortunately, a full comparison between $CARMA_B$ and $CARMA_A$ is not possible because $CARMA_B$ was evaluated according to leave-one-out tests on the predictions of one expert, while $CARMA_A$ was evaluated by training it on the predictions of one expert, then testing it on the predictions of another expert. In addition, $CARMA_B$'s test results are in terms of root-mean squared error, while $CARMA_A$'s results are in terms of estimated mean quantitative error (normally much lower than the root-mean squared error). Nonetheless, it appears that $CARMA_B$'s predictive accuracy improved substantially over $CARMA_A$'s accuracy since root-mean-squared error is typically much higher than mean error, especially the conservative mean quantitative error estimates for $CARMA_A$. This improvement suggests that match and adaptation weights should indeed be separated. These findings warranted more tests of specific and global adaptation weights and the various model-based techniques, improvements to critical period adjustment and temporal projection, and improvements to the predictive accuracy on the Wyoming test cases.

5.3 Learning Model-based Adaptation Parameters (CARMA)

In tests of $CARMA_{B,}$ removing both temporal projection and critical period adjustment led to an increase in predictive accuracy, even though both seem to be model-based knowledge that is used by experts. It was hypothesized that each expert views grasshoppers' developmental progress over a lifetime and critical period consumption somewhat differently, such that the experts' predictions are simply inconsistent with temporal projection and critical period adjustment in their initial hard-coded form. Therefore, CARMA was revised to include a technique called *model-based parameter learning* (MPL) in which parameters are included in the hill-climbing training to fully optimize temporal projection and critical period adjustment on the training cases.

In this configuration, called simply CARMA, critical period adjustment is replaced by adding `proportion of lifetime consumption in the critical period` as an adaptation feature and learning its adaptation weight. Temporal projection is included in MPL by learning parameters that scale the attrition values. Learning zero (0) values for both the critical period adapation weight and the projection parameters is equivalent to disabling critical period adaptation and temporal projection, respectively, while learning values of 1 is equivalent to fully enabling both adaptation techniques.

As described in section 5.3.1 CARMA was first tested to determine whether MPL could lead to an improvement in predictive accuracy over $CARMA_B$. During these tests, it was discovered that doing temporal projection after matching (instead of before where it had been done in $CARMA_A$ and $CARMA_B$) might lead to improvements. Tests of this approach are described in section 5.3.2. Ablation tests were then performed to test the effect on predictive accuracy of removing each form of adaptation knowledge from CARMA. Tests comparing the configuration of CARMA with the highest predictive accuracy with purely model-based and purely empirical reasoning are described in section 5.4.

These tests use the same data sets as $CARMA_B$ (*i.e.*, the Expert Sets and the Wyoming Expert Sets) plus a set of 20 cases representing the median of the experts' prediction on each case (the *Median Set*), and a set of 20 cases representing the median of the Wyoming experts' prediction on each case (the *Wyoming Median Set*). The median sets were derived in an effort to capture the expertise evident in the individual sets in one set. However, because of the extreme variability in experts' opinions for each case, it is unlikely that the median results in anything very meaningful. In fact, taking the median of the predictions for each case should cause a loss of variability from case to case, and in effect cause a complex function curve to be tremendously flattened. Leave-one-out tests were performed on these sets using the same methods (algorithms) as the tests of CARMA_B.

5.3.1 Model-based Parameter Learning Experiment

An initial experiment was performed in which CARMA_B minus critical period adjustment was compared to: 1) CARMA and 2) CARMA with projection parameter learning disabled (i.e., CARMA_B with critical period adjustment consumed by featural adaptation). The purpose of this experiment was to determine if model-based parameter learning critical period featural adaptation weight could increase predictive accuracy beyond simply disabling critical period adjustment. Table 4 shows the average root-mean-squared error of CARMA_B minus critical period adjustment (CARMA_B - CPA), CARMA minus projection parameter learning (CARMA - PPL), and CARMA, using both case-specific and global adaptation weights. The third column of Table 4 shows the performance of various CARMA configurations on the Wyoming Experts Sets - the sets upon which CARMA needs to be most accurate. Column four shows the performance on the Expert Sets. The results for the median sets appear in columns five and six.

	Configuration of CARMA	Wyoming expert sets	Expert sets	Wyoming Median Set	Median Set
Specific weights	$CARMA_B$ - CPA	17.6	17.3		
	CARMA - PPL	13.9	17.3	8.6	10.6
	CARMA	13.6	15.9	9.7	11.4
Global weights	$CARMA_B$ - CPA	21.1	18.7		
	CARMA - PPL	17.6	18.9	10.4	12.0
	CARMA	15.0	15.1	10.6	10.8

Table 4: Average percentage root-mean-squared error across eight Wyoming Expert Sets, 13 Expert Sets, the Wyoming Median Set and the Median Set for various configurations of CARMA.

These data show that for the Wyoming Expert Sets CARMA minus projection parameter learning outperforms $CARMA_B$ minus critical period adjustment by 21.0% for case-specific adaptation weights and 16.6% for global adaptation weights. This suggests that replacing critical period adjustment by critical period adaptation is a desirable change. On the Expert Sets, neither configuration is noticeably better.

Adding projection parameter learning further improves CARMA's predictive accuracy on the Wyoming Expert Sets by 2.2% for specific weights and 14.8% for global weights, and on the Expert Sets (8.1% - specific and 20.1% - global). Both results support the addition of projection parameter learning.

As expected, all configurations of CARMA (*i.e.*, with and without projection parameter learning and with global or case-specific weights) are more accurate on the

Wyoming Expert Sets than the Expert Sets. CARMA's relatively high accuracy on the median sets (8.6% to 10.6% for the Wyoming Median Set and 10.6% to 12.0% for the Median Set) suggests that taking the medians of the experts results in more easily predictable sets.

5.3.2 Match Projection Experiment

In $CARMA_A$ and $CARMA_B$ prototypical cases are temporally projected immediately *before* case matching takes place. This temporal projection is initially limited to 2 weeks in the match phase to bias matches towards cases with similar developmental phases. Before the adaptation stage, any remaining misalignment in developmental phases is eliminated by continuing temporal projection until the developmental phases are completely aligned. Temporal projection results in a modification of the densities and developmental distributions within the prototypical cases. However, since match weights are set by determining the mutual information gain between case features and qualitative consumption categories of prototypical cases in their *unprojected* form, using match weights derived from unprojected prototypical cases on projected prototypical cases could cause inconsistencies such that the incorrect prototypical case would match the new subcase. To test this hypothesis, configurations of CARMA were compared in which temporal projection was performed before and after case matching.

Table 5 shows the average root-mean-squared error of CARMA with projection occurring before matching (CARMA) and CARMA with projection after matching (CARMA - MP), using both case-specific and global adaptation weights. Columns three and four show the performance of various CARMA configurations on the Wyoming Experts Sets and Expert Sets, respectively.

The results show that delaying temporal projection until after case matching improves CARMA's performance for both the Wyoming Expert Sets (2.2% - specific and 5.3% - global) and the Expert Sets (0.6% - specific and 1.3% - global). Although tentative due to the marginal improvement, the results suggest that slight inconsistencies may result when projecting prototypical cases before matching and that projection should therefore be performed afterwards instead.

	Configuration of CARMA	Wyoming Expert Sets	Expert Sets	Wyoming Median Set	Median Set
Specific weights	CARMA	13.6	15.9	9.7	11.4
	CARMA - MP	13.3	15.8	10.1	11.2
Global weights	CARMA	15.0	15.1	10.6	10.8
	CARMA - MP	14.2	14.9	10.0	10.8

Table 5: Average percentage root-mean-squared error across eight Wyoming Expert Sets, 13 Expert Sets, the Wyoming Median Set and the Median Set for CARMA with and without match projection.

5.3.3 Ablation Experiment

To determine the contribution of the various forms of model-based adaptation to CARMA's predictive accuracy, the last in the series of ablation experiments was performed in which the performance of CARMA was compared to CARMA's performance with various adaptation mechanisms disabled. The second row of Table 6 shows CARMA's average root-mean-squared error using case specific weights

(CARMA-specific). Row three shows CARMA-specific with projection removed and row six shows CARMA with featural adaptation removed. Rows four and five show CARMA using global weights (CARMA-global). The performance of factored nearest-neighbor prediction (factored-NN) is shown in row seven.

	Configuration of CARMA	Wyoming expert sets	Expert sets	Wyoming Median Set	Median Set
Specific weights	CARMA	13.6	15.9	9.7	11.4
	CARMA - P	15.6	17.7	10.9	10.5
Global weights	CARMA	15.0	15.1	10.6	10.8
	CARMA - P	15.4	17.7	10.7	11.3
No featural adaptation	CARMA - FA	23.4	21.5	29.0	24.1
	CARMA - FA, P (factored-NN)	21.1	20.7	22.8	21.0

Table 5: Average percentage root-mean-squared error across 13 Expert Sets and 8 Wyoming Expert Sets for various configurations of CARMA.

These data show that for the Wyoming Expert Sets and Expert Sets using both specific and global adaptation weight methods the full CARMA configuration is the best. Removing projection and/or featural adaptation from CARMA leads to a noticeable decrease in performance.

In summary, the ablation experiment showed that projection and featural adaptation each increased predictive accuracy. In testing the contributions of model-based and case-based knowledge to predictive accuracy CARMA was therefore tested using both projection and featural adaptation.

5.4 CARMA vs. Empirical and Model-based Approaches

To test the hypothesis that integrating model-based reasoning and case-based reasoning can lead to more accurate predictions than the use of either technique individually, CARMA's empirical and model-based knowledge components were separated, tested in isolation, and compared to the performance of the full CARMA system under both global and case-specific adaptation weight modes.

CARMA's empirical component was evaluated by performing leave-one-out-tests for a nearest-neighbor approach and two other inductive approaches that used CARMA's empirical knowledge: decision tree induction using ID3[7] and linear approximation.

The predictive ability of CARMA's model-based component in isolation was evaluated by developing a numerical simulation based on CARMA's model of rangeland ecology. This simulation required two forms of knowledge implicit in CARMA's cases: the forage per acre based on the range value of the location, and the forage typically eaten per day per grasshopper for each distinct grasshopper overwintering type and developmental phase. The steps of the numerical simulation are as follows:

1. Project each grasshopper population back to beginning of the growing season.
2. Simulate the density and developmental phases for each overwintering type through the end of the critical period growth season based on the precipitation and

[7]ID3 classified cases into 10 qualitative consumption categories representing the midpoints (5, 10, 15, ... , 95) of 10 equally sized qualitative ranges. ID3's error was measured by the difference between the midpoint of each predicted qualitative category and the expected quantitative consumption value.

temperature given in the case.

3. Calculate the forage eaten per day per acre based on the grasshopper density per acre and the forage eaten per day per grasshopper for each overwintering type and developmental phase as affected by temperature.

4. Convert the total forage consumed to the proportion of available forage consumed based on the forage per acre.

The effect of temperature on consumption (as a result of changing metabolism rate) was represented by multiplying a coefficient (determined from a lookup table indexed by temperature) by the forage eaten per day per grasshopper for each overwintering type. The numerical simulation was trained by hill-climbing on temperature-based coefficients to maximize the predictive accuracy on the training cases.

	Predictive method	Wyoming expert sets	Expert sets	Wyoming median set	Median set
Specific weights	CARMA	13.6	15.9	9.7	11.4
	CARMA - MP	13.3	15.8	10.1	11.2
Global weights	CARMA	15.0	15.1	10.6	10.8
	CARMA - MP	14.2	14.9	10.0	10.8
Empirical only	CARMA - FA, P (factored-NN)	21.1	20.7	22.8	21.0
	ID3	34.9	30.5	35.2	32.8
	Linear appr.	25.6	28.0	11.9	13.3
Model-based only	Numerical simulation	29.6	28.8	27.9	27.6

Table 6: CARMA's average percentage root-mean-squared error across 13 Expert Sets, eight Wyoming Expert Sets, the Expert Median Set and the Wyoming Expert Median Set compared with purely empirical and purely model-based approaches.

The accuracy of each approach was tested using leave-one-out testing for each of the eight Wyoming Expert Sets and the 13 Expert Sets. The results, which appear in Table 6, include the root-mean-squared error for each of the methods.

5.5 Discussion of Integration

The results of the integration experiment provide initial confirmation for the hypothesis that integrating model-based and case-based reasoning through model-based adaptation leads to more accurate forage consumption predictions than the use of either technique individually. The root-mean-squared error for CARMA-specific on the Wyoming Expert Sets (13.6) is 35.5% lower than for the nearest-neighbor approach (21.1) and 46.9% lower than for linear approximation (25.6). The error rates for the other approaches on this data set were higher than for nearest-neighbor and linear approximation: numerical simulation (29.6) and ID3 (34.9). This initial confirmation is tentative because the low level of agreement among experts and absence of any external standard gives rise to uncertainty about what constitutes a correct prediction. However, this validation problem appears to be an inherent property of the domain of rangeland pest management.

Consumption prediction can be viewed as approximating a function from derived case features to consumption predictions (a *consumption function*). Prototypical cases constitute representative points in feature space for which function values are known. The prototypical cases can be used to induce a representation of the function as a decision tree (*e.g.*, ID3) or a numerical function (*e.g.*, linear approximation). The poor performance of

ID3 and linear approximation suggests that the biases of these inductive methods are poorly suited to the consumption prediction task. The high performance of linear approximation on the median sets (11.9 - Wyoming Median Set and 13.3 - Median Set) suggests that taking the median of the predictions for the expert sets causes the complex consumption function curve to be tremendously flattened, and as a result it is much more easily predicted by linear approximation.

Numerical simulation can be used to derive individual values for the function. However, the incompleteness of available models of rangeland ecology limits the accuracy of this approach.

A pure nearest-neighbor approach implicitly assumes that the consumption function is constant in the neighborhood of prototypical cases. CARMA's model-based adaptation approach uses a model of rangeland ecology to attempt to approximate the consumption function in the neighborhood of individual prototypical cases. For example, projection consists of simulation through the temporal interval necessary to align the developmental phases of two cases. Although the model may be insufficient in itself for accurate consumption prediction, it may greatly improve the accuracy of nearest-neighbor prediction.

5.6 Trainability of Case-specific vs. Global Adaptation Weights

The poor performance of linear approximation (25.6 as compared to 21.1 for the nearest-neighbor approach) indicates that no single linear function can accurately predict consumption. Thus, it is unlikely that a single linear function is sufficient to adapt the consumption prediction of every case. CARMA-specific does not depend on the assumption that the consumption function can be approximated by a single linear equation in the neighborhood of every prototypical case. It was therefore hypothesized that CARMA-specific would outperform CARMA-global because the latter depends on the assumption that the consumption function can be approximated by a single linear equation in the neighborhood of every prototypical case.

However, the fact that case-specific adaptation weights exhibit only marginally better predictive accuracy than global weights in the leave-one-out tests suggests overfitting on a small data set. The poor performance of case-specific adaptation is likely the result of an insufficient number of training cases to properly set the adaptation weights of every prototypical case. For individual expert sets of 10 cases each, leave-one-out tests train on nine cases and test on one, as described in section 5.2.1. The adaptation weights for the each training case are determined by maximizing the predictive accuracy on the other eight training cases. In the test cases, eight of the features used in adaptation have variable or nonconstant values (i.e., `precipitation`, `temperature`, `range value`, `infestation history`, `average developmental phase`, `density`, `feeding type`, and `proportion of lifetime consumption in the critical period`) so eight separate adaptation weights must be learned. These variables can be viewed as giving rise to eight linear equations with eight unknowns. However, based on the poor performance of linear approximation, the forage consumption function is obviously not linear. Combining the eight adaptation weights with the projection parameters further complicates the issue.

For such a complicated, non-linear function, a broader coverage of the feature values is necessary. However, eight training cases can't possibly come close to covering the 16 possible values for the qualitative features (e.g., `precipitation`, `temperature`, `range value`, and `infestation history` can take on a total of 16 possible different values), let alone the possible interdependent combinations of these values. Coverage of the quantitative feature values must also be considered. As a result, it is likely that adaptation weight learning (both global and case-specific) requires a much broader coverage of the features values than is present in only 10 test cases, particularly for case-specific weights.

In an effort to determine the potential accuracy of the various configurations of CARMA were they to be trained on a more ideal number of cases, *trainability* (i.e., the ability of a configuration to be fully trained on a set of cases) tests were performed on the different expert sets. For each expert set, the configurations of CARMA were trained on the entire set of 10 cases, then tested on the same set of cases. The results appearing in Table 7 include the results on both the Wyoming Expert Sets and the Expert Sets. The measure of trainability and the increase in error from training on the entire sets to performing leave-one-out tests are shown.

The results of the trainability tests show that for each configuration of CARMA, case-specific adaptation weights are capable of greater accuracy than global weights on both the Wyoming Expert Sets and Expert Sets. However, in the leave-one-out tests on the Wyoming Expert Sets, the accuracy of case-specific weights increases only marginally over global weights. On the Expert Sets the performance of case-specific weights is usually lower than global weights. The subpar performance of case-specific weights is particularly evident by the tremendous increase in error from the trainability tests to the leave-one-out tests (93% to 147%) as compared to global weights (52% to 67%). The rise in error for the case-specific method suggests an insufficient coverage of feature values due to a low number of training cases, which leads to an overfitting of the adaptation weights. For global adaptation, overfitting appears to be less of a problem. In more general terms, it appears that CARMA-specific requires a larger number of cases to accurately set the adaptation weights than CARMA-global.

		Wyoming Expert Sets			Expert Sets		
CARMA predictive method	Adapt. weight method	Trainability	Leave-one-out	Increase in error (%)	Trainability	Leave-one-out	Increase in error (%)
CARMA	specific	5.5	13.6	147	6.5	15.9	145
	global	9.9	15.0	52	9.8	15.1	54
- MP	specific	5.5	13.3	142	6.5	15.8	143
	global	9.3	14.2	53	9.8	14.9	52
- PPL	specific	7.2	13.9	93	8.0	17.3	116
	global	10.6	17.6	66	11.3	18.9	67
- PPL, MP	specific	6.6	15.2	130	7.5	16.1	115
	global	9.4	15.7	67	10.9	18.0	65
- P	specific	7.0	15.6	123	8.0	17.7	121
	global	9.9	15.4	56	11.3	17.7	57

Table 7: Average percentage root-mean-squared error for various configurations of CARMA through both trainability tests and leave-one-out tests across eight Wyoming Expert Sets, 13 Expert Sets, the Wyoming Median Set and the Median Set.

5.7 CARMA as a Useful Advising System

Producing accurate predictions about grasshopper forage consumption through the integration of model-based and case-based reasoning in CARMA is a very important step in providing useful advice about grasshopper infestations. As described in Chapter 2, a complete consultation resulting in a treatment recommendation requires integrating the forage consumption module with the remaining expert problem solving steps. The design of CARMA was based on the hypothesis that integrating various reasoning paradigms can lead to a useful advising system. To test this hypothesis, CARMA-specific with temporal projection occurring after match projection was trained on the Wyoming Median Set and its treatment recommendations were compared to those of the eight Wyoming experts on the 20 test cases. Table 8 shows the number of times the experts chose each treatment type for each test case. CARMA's selection is marked by a *.

The results in Table 8 show that CARMA's treatment recommendations are fairly indistinguishable from the experts, failing to match at least one expert in only one of 20 test cases (test case 3). This suggests that CARMA is a reasonable model of the experts. However, the results also show that CARMA's treatment recommendations match the majority selection only 70% of the time (14 / 20). If "no treatment" was selected each time, which CARMA does for 90% of the test cases, it would still match the majority selection 60% of the time.

As had been previously noted, training CARMA on the median forage loss predictions of the Wyoming experts causes a loss in variability from case to case and a flattening of the forage loss function curve. The average forage loss prediction in the

Wyoming Expert Sets in 51.1%, while the average forage need is 51.5%, meaning that on average grasshoppers will compete with livestock for only 2.6% of the total available forage. This low level of competition would only be treated in economically justifiable

	Number of times each treatment was recommended				
Test case	No treatment	Malathion	Carbaryl	Carbaryl bait	Nosema bait
1	**3***		1		
2	1*		**3**		
3	0*	**2**		**2**	
4	1*	**3**	1		
5	**5***				
6	**4***				
7	1*	**2**			
8	**4***				
9	**4***				
10	1*	**3**			
11	**4***				
12	**4***				
13	**5***				
14	**3***	1			
15	**3***				
16	1	**2***			
17	1	**3***		1	
18	2*	**3**			
19	**2***	**2**			
20	**3***	1		1	
Totals	52	22	5	4	0

Table 8: Summary of the treatments recommended by Wyoming experts for each of the 20 test cases as compared to CARMA. For each case the most recommended treatment is boldfaced. CARMA's recommendations are indicated by a *.

circumstances. Many of the cases do not permit the selection of any economical treatment because the treatment conditions have limited the number of treatment options (*e.g.*, in some of the test cases the user wishes to preclude all toxins, which leaves only high-cost pathogen baits) to those that are more expensive than purchasing replacement forage or renting additional land. As such, no treatment should be expected to be the prevalent selection when training on the Wyoming Median Set. In summary, the test results demonstrate that by integrating various reasoning paradigms CARMA's treatment recommendations fall within those of the Wyoming experts 95% of the time. Based on the Wyoming entomologists and pest managers as the measure of expertise, CARMA has tentatively achieved the level of a useful advising system. A better evaluation of CARMA's treatment recommendations could probably be made by training and testing CARMA on each of the individual experts separately.

5.8 Summary

This chapter has shown tentative confirmation of two main hypotheses. First, the tests of CARMA's forage consumption prediction component demonstrate that integrating model-based and case-based reasoning can lead to more accurate forage consumption predictions than the use of either technique individually. Second, the integration of various reasoning paradigms in CARMA can lead to a useful advising system. Other interesting results were discovered during the testing of the various configurations of CARMA. First, it was discovered that performance is improved by separating match and adaptation weights. Case-specific adaptation weights seemed to be more appropriate than global weights, but the number of test cases was insufficient to fully test this hypothesis.

The results showed that inconsistencies between a training set and a testing set are likely when the two are generated by different experts. The wide range of expert opinions further suggests that it would be extremely difficult to train on one expert and then duplicate the predictions of another expert. Learning model-based parameters (i.e., attrition scalars and critical period adaptation weight) to fit experts' views led a an improvement in CARMA's ability to match the forage loss predictions of the Wyoming experts. The tests also suggested that temporal projection of prototypical cases may need to occur after case matching to be consistent with match weights set from information gain.

Chapter 6

Related Work

The two areas of related research most relevant to the CARMA project are the application of artificial intelligence to natural resources problem-solving, and the integration of multiple artificial intelligence techniques.

6.1 AI in natural resources

Two principal single-paradigm approaches to designing knowledge-based systems for natural resources management have been followed. One approach has been to use rule-based reasoning to attempt to model the process of expert human reasoning, *e.g.*, Beck, Jones, & Jones (1989) and Gupta & Suryanto (1993). An alternative approach has been to derive the expert's final answer without attempting to duplicate the expert's reasoning process (Batchelor & McClendon 1989). For example, Rodell (1978) describes a large-scale numerical simulation model for grassland ecosystems. However, there is a growing recognition that no single reasoning technique is sufficient *per se* for complex natural resource problems. More recent approaches have used neural networks, *e.g.*, Ehrman, Clair, & Bouchard (1996) and Slutz & Derr (1994). Neural networks are an instance of supervised concept learning with performance similar to ID3 and nearest-neighbor. If neural networks were to be tested without the advantage of model-based knowledge, results similar to ID3 and nearest-neighbor are probable (Mooney *et al.* 1989; Weiss & Kapouleas 1989).

Several approaches combine rules and models, using at each stage in the reasoning process the technique that is most appropriate (Tao *et al.* 1991; Stone & Schaub 1990). This permits each reasoning technique to compensate for the weaknesses of the other. For example, a rule-based system modeled after an expert might be improved by the addition of model-based knowledge that the expert lacks (Jones, Jones, & Everett 1987; Beck, Jones, & Jones 1989). Conversely, a model-based system might become easier to use (especially for a novice) when combined with rules that interpret model results, *e.g.*, COMAX (Lemmon 1986).

CARMA represents a continuation of the trend toward incorporation of multiple reasoning paradigms to more effectively model human expert performance and to compensate for incomplete or uncertain knowledge. There is psychological evidence that much of human problem-solving uses past cases (Klein & Calderwood 1988). In particular, it is our observation that entomologists and pest managers reason with prototypical cases in rangeland grasshopper management. CARMA illustrates how case-based reasoning can be integrated with other reasoning paradigms for natural resource management.

6.2 Integration of CBR and MBR

Several previous research projects have investigated the benefits of integrating case-based reasoning with model-based reasoning (MBR). However, these projects have generally assumed the existence of a correct and complete causal model. For example, CASEY (Koton 1988) performed medical diagnosis using model-based reasoning to assist both case matching and case adaptation. However, CASEY presupposed both the existence of a complete causal theory of heart disease and complete explanations of each case in terms of that theory. Because the causal model in CARMA's domain is insufficient for accurate prediction and the causal explanations associated with cases are incomplete, the assumptions underlying CASEY's matching and adaptation strategies are inapplicable to CARMA's domain.

Rajamoney and Lee (1991) used a different approach to integrating case-based reasoning with model-based reasoning, termed *prototype-based reasoning*. This approach uses a library of prototypes to decompose problems into familiar subproblems. Model-based reasoning is applied to the subproblems, a consistent composition of the subproblems is determined, and model-based reasoning is applied to determine the behavior of the resulting simplified model. As with CASEY, this approach presupposes a complete and correct (though not necessarily tractable) causal model. Similarly, Goel and Chandrasekaran's (1989) use of device models to adapt design cases presupposes that the device models are complete and correct.

Feret and Glascow (1993) describe an alterative approach under which model-based reasoning is used for "structural isolation" (*i.e.*, identification of the structural components of a device that probably give rise to the symptoms of a fault). Cases are indexed by these tentative diagnoses, which are then refined using case-based reasoning. This approach, while appropriate for diagnosis, is ill-suited for behavioral prediction in the absence of faults. CARMA's use of model-based reasoning for case matching and adaptation represents an alternative approach to integrating CBR and MBR appropriate for domains characterized by an incomplete causal model.

6.3 Integration of CBR and RBR

Several projects have combined case-based reasoning and rule-based reasoning (RBR). PROTOS (Porter, Bareiss, & Holte 1990) uses rules to reason about the degree of equivalence between features in different cases in a technique called *knowledge-based matching*. In effect, PROTOS attempts to infer matching of abstract features from nonmatching observable features. An ablation study demonstrated that the use of rules for matching contributes significantly to PROTOS's performance (Porter, Bareiss, & Holte 1990). CARMA's technique of inferring abstract features in a new case in order to establish a match with a prototypical case is similar to this approach. CARMA differs in that its prototypical cases have only abstract features, whereas PROTOS' past cases are described in terms of observable features.

CABARET (Rissland & Skalak 1989) and GREBE (Branting & Porter 1991) are architectures designed to permit either rules or cases to apply to problem-solving goals. GREBE uses rules to improve case matching by inferring case facts and reformulating open-textured terms. CABARET's agenda mechanism uses heuristics to choose dynamically between rule-based reasoning and case-based reasoning. GREBE and CABARET demonstrate that integration of CBR and RBR can lead to high performance in very complex domains. CARMA is similar to these systems in that it permits both CBR and RBR to apply to high-level goals (unlike PROTOS and CASEY, which use RBR only to assist case matching and adaptation). CARMA differs from CABARET and GREBE in that the process model of rangeland pest management specifies the particular goals to which CBR and RBR apply, so in this domain (unlike the legal domains of CABARET and GREBE) it is not necessary to choose dynamically between the two techniques for each goal that arises during problem solving.

ANAPRON (Golding & Rosenbloom 1991) combines cases and rules to solve problems by using cases to represent exceptions to predictions made by the rules. However, this approach is not applicable to domains such as rangeland ecosystems, where cases and models are the main predictive components.

Chapter 7

Contributions and Future Work

This dissertation has presented an approach for integrating multiple knowledge sources for the purpose of providing accurate predictions about the behavior of physical systems whose causal theory is incomplete. This approach was used in the construction of CARMA, a system for rangeland grasshopper management advising. The research contributions of the CARMA approach and various possibilities for future research suggested by the development of CARMA are summarized in this chapter.

7.1 Contributions

This research has both theoretical and practical contributions. The theoretical contribution is a general approach for combining multiple knowledge sources (specifically case-based and model-based knowledge) for the purpose of providing accurate predictions about the behavior of physical systems whose causal theory is incomplete. Such an approach is important because many biological, ecological, and other natural systems are characterized by incomplete models and insufficient empirical data for accurate predictions. Modelling expertise in such domains requires integrating the incomplete knowledge sources. This integration is shown in the context of rangeland grasshopper management advising, a specific task arising within rangeland management that requires predictions about a biological system characterized both by an incomplete model and insufficient empirical data for accurate use of empirical techniques.

The practical contribution is a demonstration of how the general approach of integration can be applied to the task of natural resources management advising. This approach has been implemented in CARMA, demonstrating that integrating various reasoning paradigms can lead to a useful advising system.

7.1.1 Integrating Individually Incomplete Knowledge Sources

Chapter 2 introduced rangeland grasshopper management advising, a specific task arising within rangeland management that requires making accurate predictions about the behavior of a physical system with an incomplete causal theory. The absence of any complete reasoning technique necessitates integrating a variety of individually incomplete knowledge sources.

CARMA is a system that applies the general approach for integrating incomplete knowledge sources to the rangeland grasshopper management advising task. Specifically, CARMA predicts the amount of the forage consumption a grasshopper infestation will cause through model-based-adaptation, a specific integration technique for combining the incomplete causal model with case-based reasoning. CARMA differs from previous systems for natural resources management in two important respects. First, CARMA represents an improvement over single paradigm approaches (*e.g.*, rule-based reasoning, numerical simulation models, and neural networks) by having the power to exploit multiple, individually incomplete knowledge sources. Second, CARMA differs from other multiple-paradigm approaches (*e.g.*, rules and models) by using past cases, perhaps the core of expert problem-solving in much of natural resources management, as well as many other domains.

CARMA differs from previous integrations of case-based reasoning and model-based reasoning in two ways. First, several efforts have assumed the existence of a correct and complete causal model, *e.g.*, Koton (1988) and Rajamoney and Lee (1991), that is not present in many complex biological, ecological, and other natural systems. Second, Feret and Glascow's (1993) approach to identifying structural components of a device that probably give rise to the symptoms of a fault is ill-suited for complex behavioral prediction in the absence of faults.

Several projects have combined case-based reasoning and rule-based reasoning (RBR): Porter, Bareiss, & Holte (1990), Rissland & Skalak (1989), Branting & Porter (1991) and Golding & Rosenbloom (1991). However, these approaches are not applicable to domains such as rangeland ecosystems, where cases and models are the main predictive components.

7.1.1.1 Evaluation

The effectiveness of integrating model-based and case-based reasoning in CARMA to provide more accurate forage consumption predictions than the use of either technique individually was evaluated by comparing CARMA's empirical and model-based knowledge components to the full consumption prediction module under both global and case-specific adaptation weight modes. The tests of CARMA's forage consumption prediction component demonstrate that integrating model-based and case-based reasoning can lead to more accurate forage consumption predictions than the use of either technique individually. Other interesting results discovered during the testing of the various configurations of CARMA included the need to separate match and adaptation weights and the possibility that case-specific adaptation weights are more appropriate than global adaptation weights in this domain.

The results showed that inconsistencies between a training set and a testing set are likely when the two are generated by different experts, particularly based on the wide range of expert opinions in this domain. The range of opinions indicates that rangeland grasshopper management advising may typify a task in which there is no external standard of correctness, and that experts' predictions vary according to a number of factors such as experience, ethical values, risk aversion, and institutional perspectives.

Learning model-based parameters (i.e., attrition scalars and critical period adaptation weight) led to an improvement in CARMA's ability to match the forage loss predictions of the Wyoming experts. The tests also suggested that temporal projection of prototypical cases may need to occur after case matching to be consistent with match weights set from information gain.

7.1.2 CARMA: A Useful Advising System

CARMA represents a practical application of the general integration approach to the task of natural resources management advising. In doing so, CARMA demonstrates that integrating various reasoning paradigms can lead to a useful advising system. CARMA's advice (*i.e.*, treatment recommendations) was evaluated by comparing it to the advice given by entomologists and pest managers. Test results showed that CARMA's treatment recommendations were fairly indistinguishable from the Wyoming experts, failing to match at least one expert only 5% of the time.

7.2 Future Work

The implementation of CARMA represents an integration of multiple knowledge sources for the purpose of making predictions about the behavior of physical systems whose causal theory is incomplete. CARMA demonstrates that such an integration can lead to a high level of performance at the rangeland grasshopper management task. This section discusses issues that should be addressed in an effort to improve CARMA, and possible extensions of CARMA to other domains.

7.2.1 Comparison of Case-specific and Global Adaptation Weights

CARMA allows two alternative approaches to learning adaptation weights: case-specific and global. While case-specific weights produced higher predictive accuracy than global weights on the Wyoming experts, the relatively small test sets seemed to cause both types of weighting methods to suffer from overfitting. To perform a better comparison of both methods, leave-one-out tests should be extended to larger test sets. In doing so, the effect of incremental variations in test set size needs to be assessed. It is hypothesized that larger test sets in this domain are less linear, and therefore favor case-specific adaptation weights.

7.2.2 Comparison of Wyoming and Non-Wyoming Experts

The best predictions about Wyoming grasshopper infestations are arguably produced by Wyoming experts as they typically have the most experience with these infestations and the conditions typical to Wyoming. As a result Wyoming experts should provide more consistent predictions than other experts. This consistency could be tested by comparing the eight Wyoming experts' predictions to all other combinations of eight experts.

Since CARMA's goal is to provide the best advice about Wyoming grasshopper infestations, CARMA was tailored to emulate the predictions of the Wyoming experts, which it does better than the other experts. It should be determined whether CARMA's higher performance on the Wyoming experts is due to their experience with Wyoming conditions (which are modelled in CARMA) or mere coincidence.

7.2.3 CARMA As a Useful Advising System

CARMA demonstrates that integrating various reasoning paradigms can lead to a useful advising system in that CARMA's treatment recommendations are fairly indistinguishable from the Wyoming experts, failing to match at least one expert's treatment recommendation only 5% of the time. However, CARMA's recommendations match the majority selection only 70% of the time. It should be determined whether

CARMA deviates from the majority selection because its forage consumption prediction module was trained on the Wyoming Median Set (which could be determined by training and testing CARMA on individual experts) or because some experts use different treatment selection rules than CARMA.

7.2.4 Extensions of CARMA

Although CARMA is designed to advise ranchers about rangeland grasshopper infestations in Wyoming, the integration approach used in CARMA could be easily extended to grasshopper infestations in other states and possibly other countries. Such an extension would require modifying the case-base and model to include information such as other grasshopper species, range values, and historical weather conditions. This extension might also require the use of different model-based techniques. Taking this idea a step further, CARMA could be tailored to produce advice about cropland rather than rangeland grasshopper infestations, or even other pest species without varying the general technique of integrating empirical and model-based knowledge. Although CARMA is not easily extendable to predicting the behavior of all physical systems with an incomplete causal model, the general approach to combining appropriate knowledge sources is definitely applicable to any domain insufficiently described by any single reasoning technique. The emphasis of developing a system that provides predictions in such domains is on determining the domain specific knowledge sources and how they should be integrated.

7.2.5 Coping with Wide Variation in Expert Opinions

Because of the wide variation in expert opinions, the case library of the fielded version of CARMA was restricted to Wyoming experts, who arguably have more experience with Wyoming cases and who are more consistent in their predictions. However, a relatively wide range of opinions is present even among the Wyoming experts. In fact, geographical location may influence the experts' predictions less than such factors as experience, ethical values, risk aversion, and institutional perspectives. As a result, capturing the expertise of all the experts in one data set (e.g., by taking the medians of the experts' predictions) is probably impossible. Clark (1991) describes a solution in which multiple inconsistent expert opinions are incorporated in a system in which the user can select a single expert from among a set of experts. Applying this method to CARMA, someone needing control recommendations for a grasshopper outbreak on a wildlife refuge or rangeland managed by the Nature Conservancy might select cases reflecting the judgements of an expert with high environmental values. On the other hand, a rancher wishing to control grasshoppers that might compete with livestock for valuable forage might select the cases of an expert who tends to avoid risky economical recommendations. Match and adaptation values associated with the cases of different experts would have to be loaded along with these cases.

Bibliography

Allen, T. F. H., and Hoekstra, T. W. 1992. *Toward a Unified Ecology*. New York, N.Y.: Columbia University Press.

Batchelor, W., and McClendon, R. 1989. Two approaches to knowledge engineering in an insect pest management expert system. In *Proceedings of the International Summer Meeting of the ASAE/CSAE*, volume 89-7081, 1-16.

Beck, H.; Jones, P.; and Jones, J. 1989. Soybug: An expert system for soybean insect pest management. *Agricultural Systems* 30(3):269-286.

Branting, L. K., and Porter, B. W. 1991. Rules and precedents as complementary warrants. In *Proceedings of the Ninth National Conference on Artificial Intelligence*. Anaheim: AAAI Press/MIT Press, pp. 3-9.

Clark, P. 1991. *A Model of Argumentation and its Application in a Cooperative Expert System*. PhD thesis, Strathclyde University, Glasgow, UK.

Cover, T. and Hart, P. 1967. Nearest neighbor pattern classification. In *IEEE Transactions on Information Theory* 14:50-55.

Ehrman, J. M., Clair, T. A., and Bouchard, A. 1996. Using neural networks to predict ph changes in acidified eastern Canadian lakes. *AI Applications* 10(2):1-8.

Derr, V. E., and Slutz, R. J. 1994. Prediction of el nino events in the pacific by means of neural networks. *AI Applications* 8(2).

Fedra, K. 1991. Interactive modeling of environmental impacts. In *Proceedings of Automatic Control: Triennial World Congress*, 521-527.

Feret, M. P., and Glasgow, J. I. 1993. Hybrid case-based reasoning for the diagnosis of complex devices. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, 168-175. Washington, D.C.: AAAI Press/MIT Press.

Goel, A., and Chandrasekaran, B. 1989. Use of device models in adaptation of design cases. In *Proceedings of the Second DARPA Case-Based Reasoning Workshop*. Morgan Kaufmann, pp. 100-109.

Golding, A., and Rosenbloom, P. 1991. Improving rule-based systems through case-based reasoning. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, 22-27. Anaheim: AAAI Press/MIT Press.

Gupta, C., and Suryanto, H. 1993. A knowledge-based system for insecticide management for rice crops. *Transactions of the ASAE* 36(2):585-591.

Hager, W. 1988. *Applied Numerical Linear Algebra*. Prentice Hall.

Hewitt, G. B., and Onsager, J. A. 1983. Control of grasshoppers on rangeland in the United States: a perspective. *Journal of Range Management* 36:202-207.

Joern, A., and Gaines, S. B. 1990. Population dynamics and regulation in grasshoppers. In Chapman, R. F., and Joern, A., eds., *Biology of Grasshoppers*. New York, N.Y.: Wiley. 415-482.

Jones, J.; Jones, P.; and Everett, P. 1987. Combining expert systems and agricultural models: A case study. *Transactions of the ASAE* 30(5):1308-1314.

Klein, G. A., and Calderwood, R. 1988. How do people use analogues to make decisions? In *Proceedings of the DARPA Workshop on Case-Base Reasoning*, 209-218. Clearwater, Florida: Morgan Kaufmann.

Koton, P. 1988. *Using Experience in Learning and Problem Solving*. Ph.D. Dissertation, Massachusetts Institute of Technology. Department of Electrical Engineering and Computer Science.

Lemmon, H. 1986. An expert system for cotton crop management. *Science* 233:29-33.

Lockwood, J., and Lockwood, D. 1991. Rangeland grasshopper (orthoptera: Acrididae) population dynamics: Insights from catastrophe theory. *Environmental Entomology* 20:970-980.

Lockwood, J. A. 1993a. Benefits and costs of controlling rangeland grasshoppers (orthoptera: Acrididae) with exotic organisms: search for a null hypothesis and regulatory compromise. *Environmental Entomology* 22:904-914.

Lockwood, J. A. 1993b. Environmental issues involved in biological control of rangeland grasshoppers (orthoptera: Acrididae) with exotic agents. *Environmental Entomology* 22:503-518.

Lockwood, J. A. 1996. Population ecology of grasshoppers. In Gangwere, S. K., ed., *Bionomics of Grasshoppers Katydids and their Kin*. CAB International, Wallingford, UK. In press.

Mooney, R.; Shavlik, J.; Towell, G.; and Gove, A. 1989. An Experimental Comparison of Symbolic and Connectionist Learning Algorithms. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* (IJCAI-89).

Pedgley, D. 1981. Desert Locust Forecasting Manual (2 volumes). Center for Overseas Pest Research, London.

Pimm, S. L. 1991. *The Balance of Nature: Ecological Issues in the Conservation of Species and Communities.* Chicago: University of Chicago Press.

Porter, B. W.; Bareiss, E. R.; and Holte, R. C. 1990. Concept learning and heuristic classification in weak-theory domains. *Artificial Intelligence* 45(1-2): 229-263.

Rajamoney, S., and Lee, H. 1991. Prototype-based reasoning: An integrated approach to solving large novel problems. In *Proceedings of the Ninth National Conference on Artificial Intelligence*. Anaheim: AAAI Press/MIT Press, pp. 34-39.

Rissland, E. L., and Skalak, D. B. 1989. Combining case-based and rule-based reasoning: A heuristic approach. In *Eleventh International Joint Conference on Artificial Intelligence*, 524-530.

Rodell, C. F. 1978. Simulation of grasshopper populations in a grassland ecosystem. In Innis, G. S., ed., *Grassland Simulation Model*. New York, N.Y.: Springer-Verlag. 127-154.

Stone, N., and Schaub, L. 1990. A hybrid expert system/simulation model for the analysis of pest management strategies. *AI Applications* 4(2):17-26.

Tao, Y.; Thompson, T.; Moser, L.; Waller, S.; Klopfenstein, T.; Ward, J.; and Wilkerson, V. 1991. Hybrid expert system for beef-forage grazing management. *Applied Engineering in Agriculture* 7(2):262-272.

Weiss, S. and Kapouleas, I. 1989. An Empiricial Comparison of Pattern Recognition, Neural Nets, and Machine Learning Classification Methods. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* (IJCAI-89).

Wettschereck, D., and Dietterich, T. 1995. An experimental comparison of the nearest-neighbor and nearest-hyperrectangle algorithms. To appear in *Machine Learning*.

Appendix A

Temporal Projection

After case matching CARMA temporally projects the best matching prototypical case to align its average developmental phase with that of a new subcase (as described in section 3.3.2.1). `AlignProtoCase` returns the prototypical case PC temporally aligned with the new subcase SC by calling `AvgLifetimeDay` to return the average day of the grasshoppers' lifetime developmental distribution based on the temperatures within a case, and `ProjectCase` to project the prototypical case the required number of days. `ProjectCase` calls several procedures. The number of days that grasshoppers spend within a developmental phase is computed by `NumDays` according to the temperatures within a case. `AttritionScalar` returns the amount by which the grasshopper density within a developmental phase should be scaled based on the precipitation within a case. The algorithms for `AlignProtoCase` and `ProjectCase` are as follows:

$\mathit{function}\ \mathbf{AlignProtoCase}\ (PC, SC)$

$avg\_day_{pc} = \mathbf{AvgLifetimeDay}\ (PC)$

$avg\_day_{sc} = \mathbf{AvgLifetimeDay}\ (SC)$

$project\_days = avg\_day_{sc} - avg\_day_{pc}$

$\boldsymbol{return}\ \mathbf{ProjectCase}\ (PC, project\_days)$

```
function ProjectCase (C, project_days)
  D = density distribution of C by phases
  p = number of developmental phases in D
  L = NIL (density distribution of C by lifetime days)
;; Convert the density dist. from phases to days
  y = 1
  for i = 1 to p do
     no_days = NumDays ( D(i), C )
     for j = y to (y + no_days - 1) do
        L(j) = D(i) ÷ no_days
     y = y + no_days
;; Project foreward
  low = 1
  if (project_days ≥ 0) then
     for i = 1 to project_days do
        for j = y to low do
           L(j+1) = L(j) × AttritionScalar (i,C)
        L(low) = 0
        y = y+1
        low = low+1
;; Project backward
  else
     for i = 1 to |project_days| do
        for j = y to low do
           L(j) = L(j+1) ÷ AttritionScalar (i,C)
        L(y) = 0
        y = y-1
;; Convert back from days to phases
  y = 1
  for i = 1 to p do
     no_days = NumDays ( D(i), C )
        D(i) = Σ_{j=y}^{y+no_days} L(j)
     y = y + no_days
  return C
```

Appendix B

Proportion of Lifetime Consumption in the Critical Period

Critical period adaptation requires estimating the proportion of lifetime consumption in the critical period (as described in section 3.3.2.3). `ProportionConsumptionCriticalPeriod` produces this estimate by alternately projecting the grasshopper population of a case to the ends of the critical period (both beginning and end) to determine the development phases at those two points. `ProportionConsumptionCriticalPeriod` calls `ProportionLifetimeConsumptionRemaining` which estimates the proportion of lifetime consumption that remains for the grasshoppers in a case based on the proportion of grasshoppers in each stage and values for the proportion of lifetime consumption occurring in each stage. These algorithms are as follows:

function **ProportionConsumptionCriticalPeriod** (C)

$avg\_day_C = \textbf{AvgLifetimeDay}(C)$

$project\_days\_back = avg\_day_C - \textbf{BeginCriticalDay}(C)$

$project\_days\_forward = \textbf{EndCriticalDay}(C) - avg\_day_C$

$C_{back} = \textbf{ProjectCase}(C, project\_days\_back)$

$C_{forward} = \textbf{ProjectCase}(C, project\_days\_forward)$

$consumption\_after\_begin = \textbf{ProportionLifetimeConsumptionRemaining}(C_{back})$

$consumption\_after\_end = \textbf{ProportionLifetimeConsumptionRemaining}(C_{forward})$

return $consumption\_after\_begin - consumption\_after\_end$

function **ProportionLifetimeConsumptionRemaining** (C)

$total = 0$

$P = developmental\ phases\ in\ distribution\ of\ C$

$n = highest\ developmental\ phase\ in\ P$

for *each phase* $p \in P$ ***do***

$$total = total + \textbf{ProportionGrasshoppersStage}(C, s) + \sum_{i=s}^{n} \textbf{ProportionConsumptionStage}(s)$$

return *total*

Appendix C

Determining Future Infestation Probabilities

As noted in section 3.5.1, CARMA uses statistical reasoning and the historically derived Markov transitional probabilities for the infestation location to calculate the total reduced probability of future reinfestation for each treatment type. If CARMA determines that few eggs will have been laid before the treatment date (*i.e.*, treatment may result in a reduction in the probability of future reinfestation), CARMA calculates the yearly

	Transition type			
Number of years infested in last 50 yrs.	p(I→I)	p(I→U)	p(U→I)	p(U→U)
0	0.00	1.00	0.00	1.00
1-2	0.00	1.00	0.05	0.95
3-4	0.00	1.00	0.10	0.90
5-6	0.05	0.95	0.15	0.85
7-8	0.10	0.90	0.20	0.80
9-10	0.20	0.80	0.25	0.75
11-12	0.25	0.75	0.30	0.70
13-14	0.30	0.70	0.35	0.65
15-20	0.45	0.55	0.40	0.60

Table 9: Transition probabilities based on the number of years infested in the last 50 years. Infested conditions occur when there are at least 8 grasshoppers in the adult phase.

infestation probabilities for each treatment type based on the historical Markov transitional probabilities. The probabilities, which appear in Table 9, include the following transitions from one year to the next: infested to infested, p(I→I); infested to uninfested, p(I→U); uninfested to infested, p(U→I); and uninfested to uninfested, p(U→U).

Probabilities of future infestation are calculated through the following equations

$$
\begin{aligned}
&pU(0)_{treat} = 1\\
&pU(0)_{notreat,density<8} = 1\\
&pU(0)_{notreat,density\geq 8} = 0\\
&pI(0)_{treat} = 0\\
&pI(0)_{notreat,density<8} = 0\\
&pI(0)_{notreat,density\geq 8} = 1\\
&pU(i+1) = pU(i) \times p(U\rightarrow U) + pI(i) \times p(I\rightarrow U)\\
&pI(i+1) = pI(i) \times p(I\rightarrow I) + pU(i) \times p(U\rightarrow I)\\
&\qquad\quad = pI(i) \times p(I\rightarrow I) + (1 - pI(i)) \times p(U\rightarrow I)
\end{aligned}
$$

starting with year 0 (*i.e.*, the current year of infestation) where **pU(i)** and **pI(i)** are the probabilities in year **i** of being uninfested and infested, respectively. If treatment will be applied, the probability of infestation in the current year, $pI(0)_{treat}$, is 0. If no treatment will be applied, the probability of infestation in the current year, $pI(0)_{notreat,\ density<8}$, is 0 if the grasshopper density is less than 8 grasshoppers per square yard (*i.e.*, uninfested), and 1.0 otherwise.

However, these equations do not account for treatments that preserve beneficials that tend to reduce grasshopper numbers. Neither do they account for cases in which a very large area of infestation is treated. Both of these conditions will lead to further reductions in the probability of future infestation. These conditions are factored in by multiplying the probability of infestation in a year by both treatment and infestation area dependent scalars. The scalars for treatment type and total area of infestation appear in tables 10 and 11, respectively.

	Treatment type				
Year	No treatment	Malathion	Carbaryl	Carbaryl bait	Nosema bait
1	1.00	1.00	1.00	0.90	0.90
2	1.00	1.00	1.00	0.95	0.95
3	1.00				

Table 10: Factors by which the probability of going from uninfested (given treatment) to infested are scaled based on the treatment type.

	Total area infested * 10,000 acres									
Year	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9+
1	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00
2	1.00			0.80	0.75	0.70	0.55	0.45	0.40	0.35
3	1.00						0.80	0.70	0.70	0.65
4	1.00									

Table 11: Factors by which the probability of going from uninfested (given treatment) to infested are to be scaled based on the total area of infestation.

The algorithm `ProbabilityInfestationInYear` factors these scalars into the probability equations to determine the probability of an infestation occurring in a year based on the treatment type as follows:

```
function ProbabilityInfestationInYear (Year TreatmentType)
  If (Year > 0) then
    pI(Year − 1) = ProbabilityInfestationInYear ((Year − 1) TreatmentType)
    return pI(Year − 1) × p(I→I) +
           (1 − pI(Year − 1)) × p(U→I) × TreatmentScalar × AreaScalar
  Else
    If (TreatmentType ≠ "NoTreatment") then return 0.0
    Else
      If (DensityAtAdult < 8) then return 0.0
      Else return 1.0
```

The total reduced probability of future reinfestation for the treatment type is calculated by summing each yearly difference between the probability of infestation with and without treatment.