

Korine N. Kolivras · Andrew C. Comrie

## Modeling valley fever (coccidioidomycosis) incidence on the basis of climate conditions

Received: 18 February 2002 / Accepted: 28 October 2002 / Published online: 15 January 2003  
© ISB 2003

**Abstract** Valley fever (coccidioidomycosis) is a disease endemic to arid regions within the Western Hemisphere, and is caused by a soil-dwelling fungus, *Coccidioides immitis*. Incidence data for Pima County, reported to the Arizona Department of Health Services as new cases of valley fever, were used to conduct exploratory analyses and develop monthly multivariate models of relationships between valley fever incidence and climate conditions and variability in Pima County, Arizona, USA. Bivariate and compositing analyses conducted during the exploratory portion of the study revealed that antecedent temperature and precipitation in different seasons are important predictors of incidence. These results were used in the selection of candidate variables for multivariate predictive modeling, which was designed to predict deviation from mean incidence on the basis of past, current, and forecast climate conditions. The models were specified using a backward stepwise procedure, and were most sensitive to key predictor variables in the winter season and variables that were time-lagged 1 year or more prior to the month being predicted. Model accuracy was generally moderate ( $r^2$  values for the monthly models, tested on independent data, ranged from 0.15 to 0.50), and months with high incidence can be predicted more accurately than months with low incidence.

**Keywords** Climate and health · Modeling · Valley fever · Coccidioidomycosis · *Coccidioides immitis* · Disease prediction

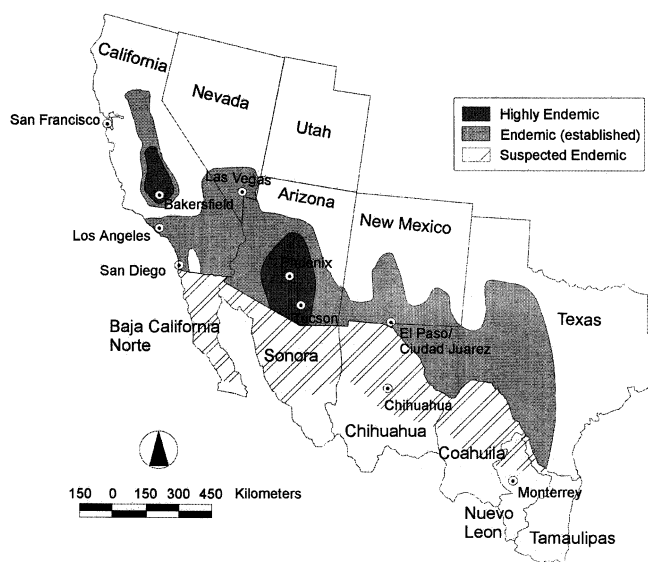
### Introduction

Coccidioidomycosis, commonly known as valley fever, is a disease endemic to parts of the Western Hemisphere.

K.N. Kolivras (✉) · A.C. Comrie  
Department of Geography and Regional Development,  
The University of Arizona, Room 409 Harvill Building,  
Tucson, AZ 85721-0076 USA,  
e-mail: korine@u.arizona.edu  
Tel.: +1-520-621-1652, Fax: +1-520-621-2889

It is found in limited regions in the United States, as well as areas in Central and South America including Mexico and Argentina. The most highly endemic regions within the United States (Fig. 1) include Kern County in the San Joaquin Valley of California (hence the name valley fever) and Pima, Pinal, and Maricopa counties of Arizona (Maddy 1965). Major urban areas within the endemic zone include Bakersfield, California, and Phoenix and Tucson, Arizona. The disease is caused by *Coccidioides immitis* (*C. immitis*), a soil-dwelling fungus that is sensitive to climate conditions. While the basic relationships between climate conditions and valley fever incidence are generally acknowledged, they have received little study and are poorly understood (Kolivras et al. 2001).

Valley fever infections begin in the lung, when the fungus becomes airborne and is inhaled by a host. Humans, other mammals especially dogs and cattle, and

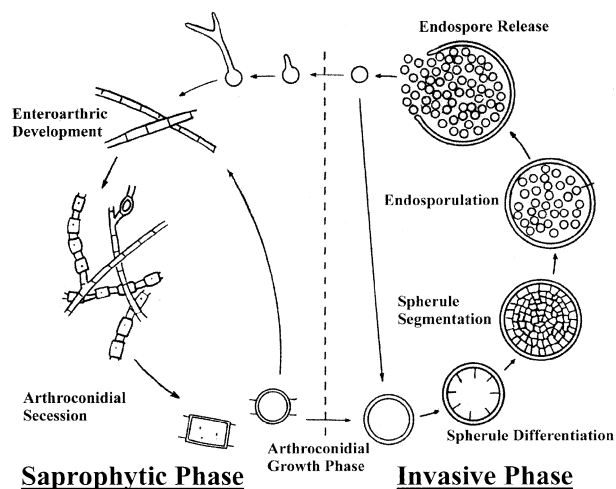


**Fig. 1** Areas of the United States and northern Mexico that are considered endemic for valley fever. (Adapted from Kirkland and Fierer 1996)

reptiles are all susceptible to the disease. The majority of the people infected (about 60%) either present no symptoms, or experience mild, cold-like symptoms (Smith et al. 1946). Some may endure a variety of flu-like conditions, including fever, coughing, and chest pain, which usually appear after an incubation period of 10 days to 3 weeks (Smith et al. 1946; Stevens 1995; Mardo et al. 2001). Of those infected by *C. immitis*, about 1% develop the disseminated form of the disease when the fungus spreads beyond the lungs (Einstein and Johnson 1992). Disseminated valley fever can express itself with a wide variety of conditions, including joint damage, skin lesions, and potentially fatal meningitis. Those with mild symptoms may recover quickly, while people with disseminated disease may experience chronic symptoms for an extended period of time. Within the U.S., there are approximately 100,000 new infections each year with about 1% of infections resulting in death (Valley Fever Center for Excellence 2001). This number is simply an estimate, however, since most infected people do not become ill enough to seek medical treatment, and their cases are therefore not reported. While most infected people do not seek medical care, the treatment of serious cases can be costly. Treatment within the U.S. costs approximately \$9 million per year, and results in a loss of about 1,000,000 person-days of labor (Pappagianis 1980).

Occupation is a factor shown to affect exposure to the fungus, while factors including gender, age, ethnicity, and immune status have been linked to the likelihood of experiencing disseminated disease. Although anyone living in or spending time in the endemic region can be exposed, those working outside, including agricultural and construction workers (Johnson 1981) and archaeologists (Werner and Pappagianis 1973) are more likely to be exposed to the fungus than those working in other occupations and, on average, experience a higher incidence of disease. People under the age of 5 and over the age of 50 years are more susceptible than others to developing a disseminated case, as are men and people of African-American or Filipino descent (Pappagianis 1988). Immunosuppressed individuals, including those who are HIV-positive, organ transplant patients, diabetics, and women in their third trimester of pregnancy, are also more susceptible than others to disseminated disease (Valley Fever Center for Excellence 2001).

A brief summary of the lifecycle of the fungus is useful for understanding the link between climate conditions and valley fever incidence. *C. immitis* is considered to be a dimorphic fungus (Fig. 2), meaning that its lifecycle consists of two different phases (Fiese 1958). In the soil, *C. immitis* exists in the saprophytic phase. Given proper conditions, slender filaments of cells, called hyphae, grow as a saprophyte in the upper part of the soil. When the soil dries, alternating cells in the hyphae die and the remaining viable detached cells become arthroconidia (spores). Some portions of the live fungus remain in the soil, while other spores may become airborne due to a natural (generally wind) or anthropogenic disturbance of



**Fig. 2** *Coccidioides immitis* exists in both saprophytic (left) and parasitic (right) phases. (Adapted from Fiese 1958)

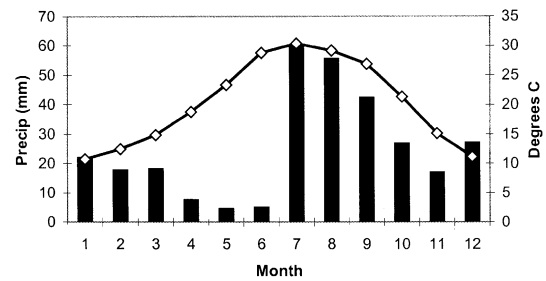
the upper portion of the soil. Once a host inhales a spore, the invasive phase of *C. immitis* begins and the fungus reproduces within the lungs (Fig. 2). The infection in the invasive phase is generally contained in the pulmonary system, but in some cases it disseminates to other parts of the body. In some cases, the death of the host, such as a rodent, returns spores to the soil, thus completing the cycle and beginning the saprophytic phase.

*C. immitis* responds to the moisture content and temperature of soil, therefore a relationship exists between climate conditions and valley fever incidence (Kolivras et al. 2001; Hugenholtz 1957; Maddy 1957, 1958). However, little research examining the role of climate variability in the occurrence of valley fever has been performed since the 1950s and 1960s. Most studies anecdotally mention the presence of a link between climate and incidence, but they do not quantitatively examine that relationship. No modeling studies attempting to predict incidence on the basis of climate conditions have been conducted. Kolivras et al. (2001) reviewed the existing literature and found that, of the studies during that period, only a few compared climate and incidence data. In particular, the study by Hugenholtz (1957) looked for a correlation between such information, but only 14 years of incidence data were available at that time.

The role of precipitation in the lifecycle of *C. immitis* is twofold: the fungus requires moisture to complete its lifecycle, but a period of dry conditions enables the hyphae to break apart and develop fungal spores that may become airborne (Pappagianis 1980). After rain, the fungus grows rapidly until the soil dries or until competitors stifle its growth (Reed 1960; Maddy and Coccozza 1964). After the soil dries and spores have formed, wind or anthropogenic disturbances liberate the fungal spores, which may then be dispersed and cause infections if they are inhaled. Therefore, past research has led to the development of a hypothesis that states that a cycle of wet and dry conditions is necessary for outbreaks of the disease to occur.

Temperature also plays a vital role in the growth of *C. immitis* through surface soil sterilization. It is hypothesized that during prolonged periods of hot, dry conditions, for example during summer when little rain is received, the surface of the soil is partially sterilized and many competitors are removed, but *C. immitis* spores remain viable below the surface (Maddy 1965; Reed 1960). When rain falls, conditions in the surface soil eventually approach the ideal for the growth of the fungus. It is thought that *C. immitis* then returns to the surface layer, which contains few competing organisms, and grows fairly rapidly in this ideal environment (Maddy 1957). A subsequent dry period then allows the fungus to encyst and become airborne, and infections to occur.

Because of data availability, this study focuses on Pima County, which is located in south central Arizona in the Sonoran Desert (Fig. 1). Pima County, Arizona (essentially the greater Tucson region) has one of the highest rates of valley fever in the world, and is experiencing a rapid growth of people susceptible to the disseminated form of the disease, including the elderly and immunosuppressed people (Galgiani 1999). Characteristic of much of the endemic area, Pima County receives low annual precipitation (approximately 300 mm in Tucson), which is coupled with a wide range in diurnal and seasonal temperatures (Fig. 3). The region is characterized by a bimodal precipitation pattern, with winter/spring (December through March) and late summer (July through September) rainfall peaks separated by dry periods (Sheppard et al. 2002). Winter precipitation is received mainly as a result of frontal systems that enter the southwestern portion of the United States, with characteristic soaking rains that may last several days. Following the northward retreat of frontal systems in spring is a dry foreshummer period (April through June) in which insolation and temperatures are high owing to a lack of cloud cover under the overlying subtropical anticyclone. Summer precipitation occurs as a result of the North American monsoon, and is characterized by intense thunderstorms with high spatial and temporal variability in precipitation (Sheppard et al. 2002). Monsoon circulation is usually in place around the beginning of July, and typically lasts through mid-September. Following the end of the monsoon pattern, a relatively dry period occurs in October and November until the beginning of winter precipitation, typically in December (Fig. 3). These average patterns show large variability from year to year, and are affected in part by climate fluctuations such as the El Niño–Southern Oscillation (Sheppard et al. 2002). The climate of Pima County is therefore conducive to high valley fever incidence when considered within the framework of the fungus's response to climate conditions given the annual cycle of wet and dry conditions. Other relatively time-invariant environmental factors within the endemic area, including soil type and salinity, also provide an environment favorable to the growth of the fungus (Kolivras et al. 2001). However, this study focuses on the covariability and potential predictive association between climate conditions and disease incidence.



**Fig. 3** Average monthly precipitation (bars) and temperature (line) at Tucson International Airport, Tucson, Arizona (1961–1990)

The broad aims of this research are twofold. Given the paucity of information in the literature, the first goal is to improve our understanding of the basic relationships between climate and valley fever through exploratory data analyses, including bivariate correlation analyses and compositing analyses of antecedent climate conditions. Using the understanding of climate and valley fever gained through the exploratory data analyses, our second goal is to develop monthly multivariate models to predict valley fever incidence on the basis of current or forecast climate conditions. Through this step, we also aim to improve understanding of the response of *C. immitis* to changes in climate conditions through an examination of the climate variables best related to, and the best predictors of, valley fever incidence.

## Data and methods

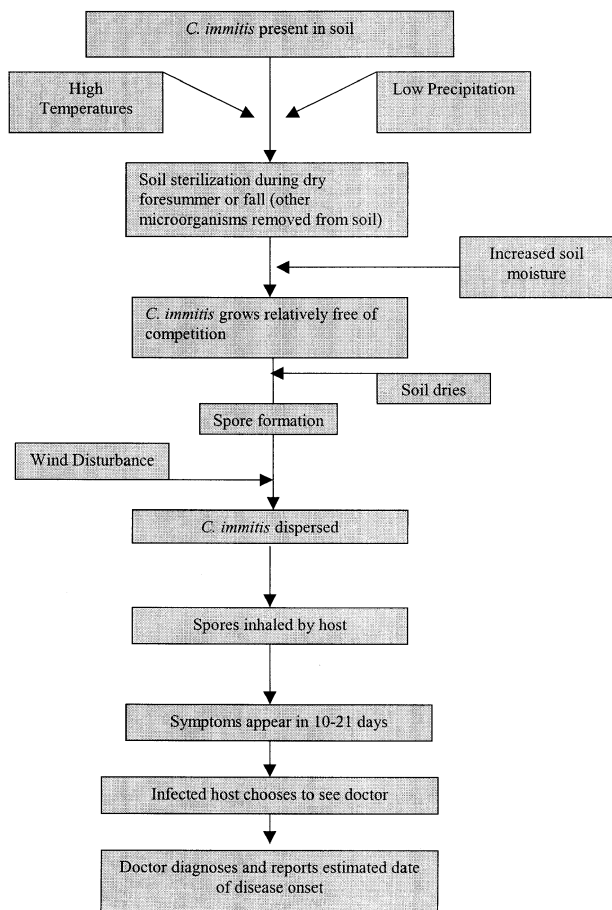
### Data

Monthly valley fever incidence data, by month of estimated onset of disease symptoms, for Pima County for 1948–1998 were obtained from the Arizona Department of Health Services (ADHS). Valley fever became a reportable disease in 1995 in states where it is endemic, but incidence data have been gathered for over 50 years in Arizona. Compliance is expected to be fairly good, although incidence data only represent sufferers who become ill enough to seek medical care. Monthly data were used in order to examine patterns at sub-annual time scales. Although some incidence data are available at weekly time scales, the lag between infection and reporting makes the use of such data inappropriate.

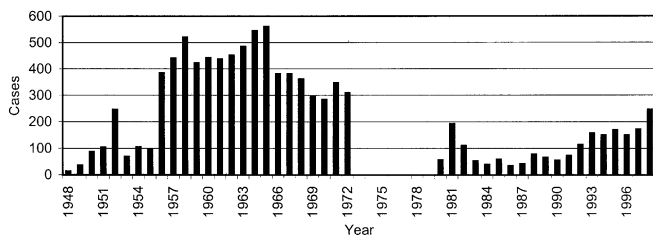
Ideally for this study, we would analyze the relationship between valley fever and climate using actual fungal count data from the soil or air rather than incidence data. Unfortunately, fungal count data are currently unavailable for several reasons: The fungus is very difficult to isolate in the soil, and the culturing process requires special laboratory biosafety facilities and is very time-intensive. As a result, there are no time-series of spore data amenable to climatic analysis. Instead, we use incidence data, which are several steps removed from the effect of climate (Fig. 4). Following an airborne dispersal of *C. immitis* spores in which a host becomes infected, if symptoms appear and the host's condition becomes severe, which may take weeks or months, the infected person will visit a doctor. The physician then reports to Arizona Department of Health Services (ADHS) the estimated date of disease onset, which may be quite uncertain or approximate.

There are concerns about the quality of the incidence data too. Figure 5 illustrates the data record, and several points require comment. As shown by the graph, there are no available data for





**Fig. 4** Incidence data are several steps removed from the effects of climate conditions on fungal growth



**Fig. 5** Annual cases of valley fever, Pima County, Arizona (1948–1998, data unavailable 1973–1979). Much of the temporal variability may be related to changes in reporting

1973–1979, decreasing the 51-year data record to 44 years. Perhaps the major problem in the data record is the lack of a consistent reporting standard and diagnostic criteria over time. The method of reporting cases of valley fever to the ADHS by doctors has changed over the past 50 years because of variation in physician diagnosis and case definition, and the use of laboratory verification in recent years (K. Komatsu, personal communication 2000). Overreporting may explain the very high number of cases during the late 1950s and 1960s, but interannual variability in the period of record may still be dominated by reporting changes and changes in disease classification codes. The data from 1980 to 1998 are considered to be more trustworthy than the entire data record since annual variation is less extreme during that time. Also, in the mid-1990s, reporting techniques were standardized

when the disease became a reportable illness through a standard case definition. In addition, the number of susceptible people has changed over this period, disease awareness has increased, and soil disturbance due to development has varied as well. These factors play a role in the unevenness of the time series. There may be a difference between urban/rural reporting; however, given that the majority of Pima County's population is within the Tucson metropolitan area, such a reporting difference would not greatly affect the data in this study.

Monthly climate data for southeastern Arizona (Climate Division 7, which includes Pima County) were obtained from the National Climatic Data Center (NCDC). In addition to temperature and precipitation, we used the Palmer Drought Severity Index (PDSI) as a proxy for soil moisture since an appropriate measure for Pima County was not available for 1948–1998. A negative PDSI value indicates dry soil conditions, while a positive value indicates moist conditions. The balance between wet and dry conditions in the soil is likely to affect the growth and distribution of *C. immitis*, and therefore PDSI was deemed a useful variable to include in the study. One caveat with the use of PDSI is that temporal autocorrelation is intrinsic to the index. The method used to create the value means that the index does not change rapidly with changes in temperature and precipitation patterns. Rather, the index changes slowly and smoothly over a period of several months, much like soil moisture conditions (Guttman 1991). Finally, the index is not currently forecast into the future, and would therefore not be useful in a model incorporating forecast climate conditions. For these reasons, PDSI was used for exploratory analyses, but not included in model development. Other climate data were acquired from NCDC for an individual station, Tucson International Airport, which is located in the south-central part of the city and provides data representative of the larger urban area. Since 98% of Pima County's population resides in the Tucson metropolitan area (United States Census Bureau 2001), it was acceptable to apply Tucson station climate data to countywide valley fever data. The station data included average daily maximum, minimum, and dew point temperatures and average daily wind speed. These daily data were averaged to produce monthly values for comparison to the monthly incidence data.

#### Exploratory data analysis

To understand the basic relationships between valley fever and climate, and to determine the most appropriate climate variables to include in the multivariate predictive model, an exploratory data analysis was performed in two steps. Initially, bivariate comparisons of climate variables and incidence were performed. Then, the climate conditions leading up to a month with particularly high or low incidence were examined through a compositing analysis. The results of the exploratory portion of the study guided the development of multivariate regression models to predict monthly incidence using antecedent climate conditions.

The monthly climate variables included in the bivariate analyses were total precipitation, average, minimum, and maximum temperatures, dew point temperature, average wind speed, and the PDSI. For these analyses, valley fever incidence data from 1980 to 1998, standardized by the mid-year Pima County population estimate (United States Census Bureau 2001), were used. As previously mentioned, these data are considered to be more reliable than the entire long-term record.

The analyses were performed using lags of 1–24 months in order to determine the timing of the influence of climate variables on incidence. The lags accounted for a delay in the impact of climatic conditions on the growth and dispersal of *C. immitis*. Incidence in a particular month was compared to each of the climate variables in the preceding months, up to a period of 2 years. The relationship was examined visually with scatterplots, and by calculation of Pearson's correlation coefficients between variable pairs.

It is likely that some correlations are due to chance, but many are significant at the 95% level, and some are significant at the

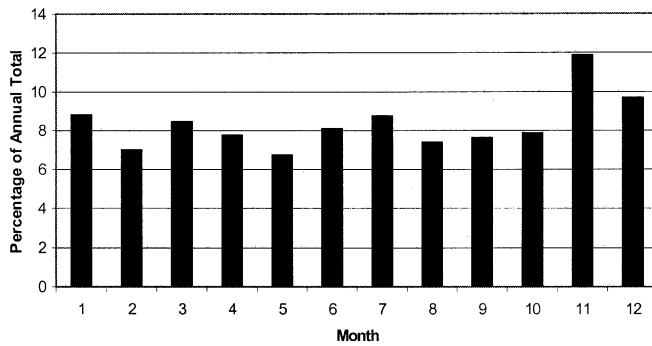


Fig. 6 Mean monthly percentage of total annual incidence

99% level. Fourteen significant correlations (5% of 288) would be expected to be due to chance in our comparison of 12 months of incidence data with up to a 24-month lag in climate variables, and in our analysis more than 14 significant values were found when correlating incidence and all climate variables.

In order to use the entire record of incidence data (1948–1998) for the composite analyses (the second portion of the exploratory data analysis), the raw case counts were transformed to account for the changes in reporting methods over the period. During any one particular year, it is likely that the same reporting standards were used. Therefore, month-to-month variability in any one year is likely to be relatively precise, even if the raw data are inconsistent over longer periods. Incidence in each month was expressed as a percentage of the respective year's annual total (e.g., January 1983 as a percentage of total incidence in 1983) (Fig. 6). Deviation from the mean monthly percentage of the annual total was then calculated for each month (e.g., January 1983 percentage above or below average percentage for all Januaries), and the ten highest and lowest of these transformed incidence deviations were identified for each month. Climate conditions throughout the period of record (1948–1998, excluding 1973–1979) were averaged by month. Differences from mean climate conditions were then calculated and composited (averaged by month) for temperature and precipitation for the 48 months preceding these high and low incidence deviations. The composite values were graphed, and the relative incidence values were visually compared to climate values. This procedure added to the information found during the bivariate analysis, and allowed for the development of better-informed models. Antecedent above or below-average climate conditions were compared to similarly above or below-average valley fever incidence (e.g., are high January incidences preceded by hotter/cooler and wetter/drier than average conditions?). Auto-correlation in the data from month to month was found to be low, and it is unlikely that a reporting bias explains our results since months with both high and low incidence were analyzed at the monthly time scale.

#### Modeling overview

To improve our understanding of multivariate relationships between climate variables and incidence, and to explore the potential for forecasting disease outbreaks, multiple linear regression models were developed for each month. Candidate input variables were selected from the results of the exploratory data analyses, and screened by principal-components analysis to avoid multicollinearity. Variables included in the model development portion of the study were temperature and precipitation. The variables included in model development also incorporated a number of interaction terms that were developed for each month, in which precipitation and temperature variables selected from the exploratory data analyses were multiplied to allow for complex relationships. PDSI was useful during the exploratory analyses, but was not used for model development.

The 44-year time-series of incidence data for each month was examined for outliers, which were found to be artifacts resulting from the data transformation process. Six different months had one outlier each in the 1948–1998 time series, which were excluded from model development; no other outliers were identified. The models were designed to predict deviation from mean incidence (explained above), and were cross-validated on independent data.

Eleven to fourteen initial variables for each month were selected for model development. Some of those variables were highly correlated with one another; therefore a principal-components analysis was conducted for each month in order to avoid multicollinearity and increase parsimony. The original variables were reduced to between five and eight components for each month, which explained 72% to 78% of the variance. The highest-loading variable in each component, as well as those variables that were not highly loading in any component but it was logical to include given findings reported in the literature, were entered into the modeling procedure.

The monthly models were initially developed on all data using a backward stepwise regression procedure to reduce the variables to those that were statistically significant ( $\alpha = 0.10$ ). The relatively small number of years ( $n = 44$ ) for model building made standard cross-validation techniques, in which a subset of the data is set aside for testing, difficult to use. Therefore, a standard jack-knife (leave-one-out) cross-validation technique was employed. After a monthly model had been developed on all data using the backward stepwise procedure, the selected variables were forced into individual non-stepwise models in which data for 1 year were left out. This process was repeated so that each year was left out of the process one time. Each instance of the model was then used to predict the year that was omitted. This resulted in  $n = 44$  independent data points for validation from the 44 similar, but not quite identical models.

## Results and discussion

### Bivariate analysis

Incidence in late summer (July, August, and September) is negatively correlated with precipitation in the summer months immediately before (Table 1). High monsoonal precipitation may decrease the likelihood of fungal spores becoming airborne, decreasing incidence in the months that follow. Incidence during the same period in late summer is also positively correlated with precipitation during the winter and early spring (March and April). The relationship implies that these soaking rains may provide the moisture needed for the fungus to grow within the soil. This pattern repeats itself at longer time scales. Summer precipitation at a 1 year lag is negatively correlated with incidence during the following late summer and fall, while February precipitation is positively correlated with incidence during spring and summer 1 year later.

Average air temperatures in July and August are positively associated with incidence in Pima County in autumn (Table 2). There is a positive relationship between incidence in late winter and spring, and temperature in the preceding 7–9 months. Incidence in other months appears to be affected less by average temperature, although temperatures in December through February are positively correlated with incidence with a lag of 14–19 months. An analysis similar to that of average air temperature was conducted with minimum and maxi-





imum air temperatures and incidence (not shown). Minimum air temperatures in July and August are positively correlated with incidence in the months that follow, particularly early fall and winter. Maximum temperatures in July, August, and September appear to positively affect incidence in the short term, in the 1 or 2 months that follow. Higher than normal maximum temperatures in summer may lead to increased evaporation and below-normal soil moisture, thereby allowing the fungus to become airborne and infections to occur in the next few months. This finding also fits well with previous research that associates high temperatures with soil sterilization. Given the results of our correlation analysis, it appears that extreme summer temperatures in particular are important, and lead to higher than normal incidence in the following winter.

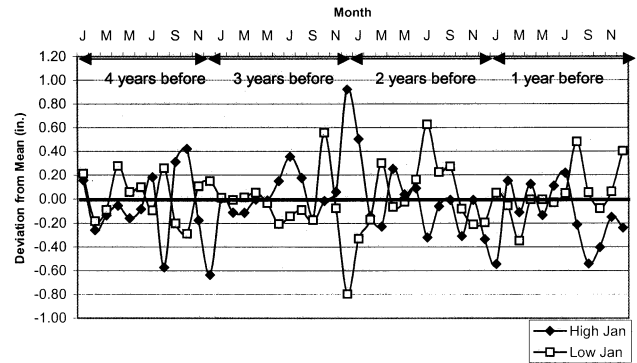
Average dew point temperatures in the first 7 months of the year are significantly associated with incidence in only 1 or 2 months with few clear, consistent patterns (not shown). It was expected that dew point temperature, as an indicator of moisture content in the air, would affect the ability of the fungus to become airborne; a high moisture content in the air would translate to somewhat moist topsoil. However, the few seemingly spurious high correlations did not indicate any clear association between dew point temperature and incidence.

At least for this temporal scale, no relationships between wind speed and incidence were significant (not shown). It is more likely that individual, daily wind events, such as very high gusts, affect incidence rates of valley fever. Gust data and maximum sustained wind speed were not analyzed in this study because they occur on timescales much shorter than 1 month (daily or hourly); their possible importance suggests they should be examined in future studies.

The PDSI value has a lagged negative influence on incidence in every month in which there is an apparent relationship. Incidence in winter and spring is correlated with PDSI in summer and fall, a pattern that repeats itself over longer periods as well (Table 3). In the short term, PDSI is likely negatively correlated with incidence because greater soil moisture prevents the fungus from becoming airborne. Conversely, if PDSI values are near zero or negative, the soil is likely to be dry and more infections may occur. PDSI is negatively correlated with incidence on longer time scales as well. Incidence in fall is negatively correlated with PDSI at lags of 8–24 months, as well as in the 3 months immediately preceding.

### Composite relationships

Months with a high (low) relative incidence are often immediately preceded by lower (higher) than average precipitation. During the summer (June through September), months with a high (low) relative incidence are characterized by higher (lower) than average precipitation for much of the previous 12–36 months. Although



**Fig. 7** Precipitation composites (average deviation from mean precipitation 1948–1998, 1973–1979 excluded) leading up to Januarys with high and low percentage of total annual valley fever incidence. The rightmost point represents December immediately prior to the high/low Januarys, the second rightmost point represents November 2 months prior, etc

deviation from mean precipitation is highly varied prior to a month with high relative incidence, in most monthly graphs (not shown) a pattern appears 24 months earlier that indicates above-average precipitation. This again points to the long-term importance of moisture enabling the fungus to grow abundantly in the soil.

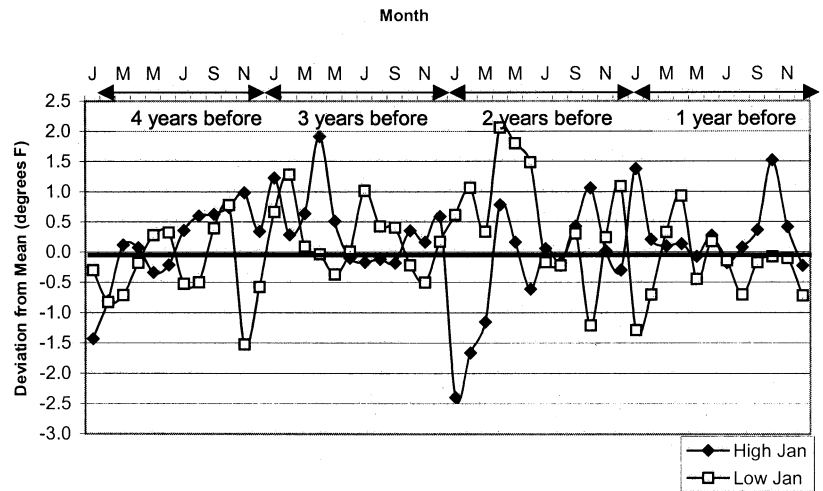
The January graph (Fig. 7) provides an example of the complexity of the composite analysis for precipitation. The pattern is highly varied both for months with high and low incidence. Some expected patterns are visible however. In December and January 2 years prior to a January with a high percentage of total annual incidence, above-average precipitation is received, while a drying trend is present in the September and October immediately preceding it. In contrast, low-incidence Januarys are preceded by dry winters 2 years earlier, and possibly by wet summers 18 months before. Therefore, an opposite pattern is apparent for months with high and low incidence.

The temperature composite graphs are also highly varied, but some overall patterns are apparent. The graphs are summarized for the sake of space constraints, with an analysis of the January graph provided as an example. Months with a high percentage of total annual incidence are often preceded by higher than average temperatures. This factor is likely related to soil moisture, as well as the soil sterilization hypothesis involving summer temperatures. High temperatures increase evaporation, leaving the soil dry and the fungus able to become airborne. Some months with a low percentage of total annual incidence are preceded by lower than average temperatures; however the pattern is not as consistent as that of high-incidence months. Approximately 2 years prior to a month with high incidence, the composite graphs for some months show below-average temperatures. This decrease in temperature coincides with the period of increased precipitation that allows the fungus to grow in higher than average numbers.

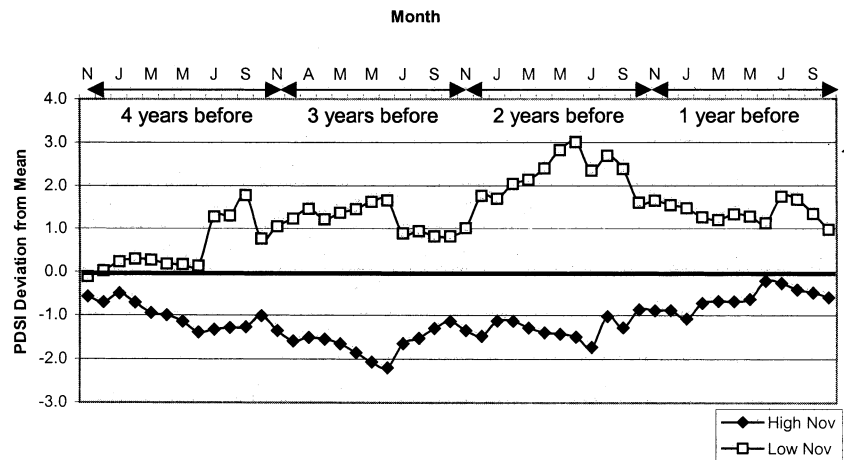




**Fig. 8** Temperature composites (average deviation from mean temperature 1948–1998, 1973–1979 excluded) leading up to January are also highly varied and complex. The rightmost point represents December immediately prior to the high/low Januarys, the second rightmost point represents November 2 months prior, etc



**Fig. 9** Palmer Drought Severity Index (*PDSI*) composites show that average deviation from mean *PDSI* conditions fluctuate very little prior to November with either a high or a low percentage of total annual valley fever incidence (1948–1998, 1973–1979 excluded). The rightmost point represents October immediately prior to the high/low Novembers, the second rightmost point represents September 2 months prior, etc



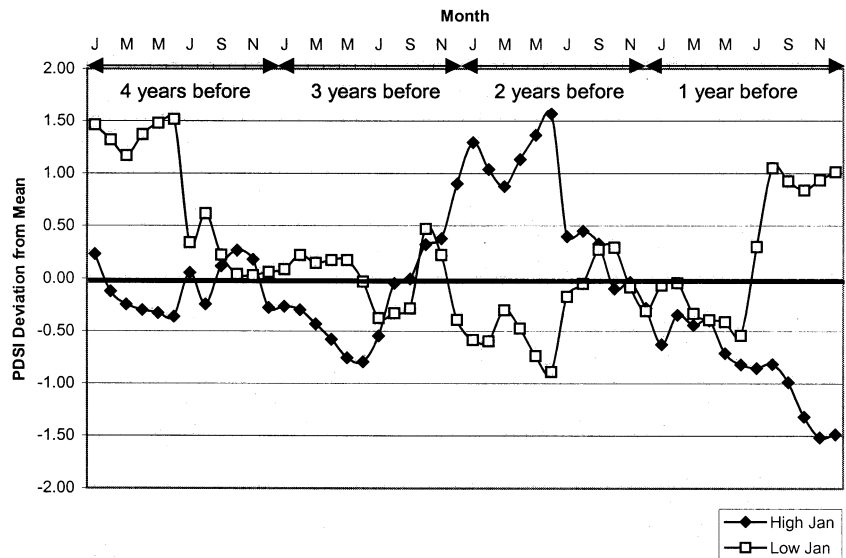
The composite graph for Januarys (Fig. 8) with a high percentage of total incidence illustrates the above points. Temperatures in October are above average and, along with below-average precipitation (Fig. 7), soil moisture conditions likely allow fungal spores to become airborne more easily. Winter months (January, February, and March) 2 years prior to a January with high incidence experience below-average temperatures and receive above average precipitation, as indicated by Fig. 7. These conditions in the soil may foster an environment conducive to the growth of the fungus. For low-incidence Januarys, there is an inconsistent temperature pattern over the preceding 18 months, although the April–June dry foreshummer period nearly 2 years before seems unusually warm.

The temporal smoothing inherent to *PDSI* is apparent in most composites, in that *PDSI* does not fluctuate greatly over the 4-year period. Most monthly composites indicate that a month with high (low) incidence is preceded by drier (wetter) than average conditions, as indicated by *PDSI*. For some months, such as November, with above-average incidence (Fig. 9), *PDSI* values fall

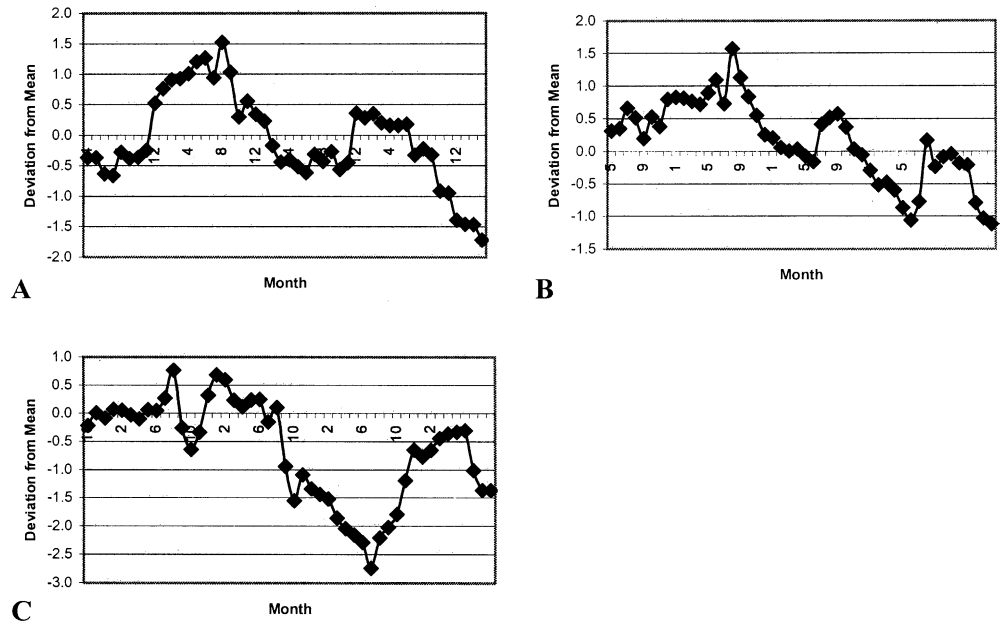
below the mean for the entire 48-month composite. Other months, including January (Fig. 10), show *PDSI* values that fluctuate around the mean in the preceding years. In the January example (Fig. 10), high-incidence Januarys have high *PDSI* values (presumably higher soil moisture) 18 months to 2 years before, followed by a clear drying trend. Low-incidence Januarys show a somewhat contrasting pattern, with dry or moderate conditions leading up to a moist period in the previous 5 or 6 months. Generally, a marked dry period is found about 6 months prior to a month with higher than average incidence. This dry period may allow the fungus to form spores within the soil and be dispersed more easily. The months June through September with above-average incidence are preceded by above average *PDSI* values for almost the entire 48-month composite (not shown). This pattern was not expected, but perhaps indicates that the fungus responds on shorter timescales than *PDSI* variability.

April, May, and October with a high percentage of annual incidence show an interesting moisture pattern that fits well with past findings. In the composite graph for

**Fig. 10** Average deviation from mean PDSI conditions shows high variation prior to January with a high and a low percentage of total annual valley fever incidence (1948–1998, 1973–1979 excluded). The rightmost point represents December immediately prior to the high/low Januarys, the second rightmost point represents November 2 months prior, etc



**Fig. 11a–c** PDSI composites for April (A), May (B), and October (C) with a high percentage of total annual valley fever incidence show similar patterns of moist and dry conditions



all 3 months (Fig. 11), above-average moisture conditions are apparent about 2–3 years prior to a month with high incidence, according to PDSI values, and a reverse pattern exists during some months with a low percentage of annual incidence (not shown). This supports hypotheses in the literature regarding the role of soil moisture in the growth and dispersal of *C. immitis*.

#### Model development and variables

The final variables included in each monthly model are outlined in Table 4, along with the *P* value for each variable. Most of the variables selected by the modeling procedure occur 1 year or more before the month being predicted. It appears that short-term climate conditions are not as important in predicting incidence as long-term

conditions. This is partly counter-intuitive, but may be a result of shorter-term processes being filtered out in the many steps between fungal growth and severe disease incidence (Fig. 4). About 40% of the variables chosen are either winter temperature or winter precipitation over varying periods. It therefore appears that conditions during winter have more of an effect on incidence during any month than conditions during other seasons, and are therefore more useful in prediction. Winter precipitation is more consistent and evaporates less quickly than that from summer thunderstorms (Sheppard et al. 2002). It is characterized by soaking rains rather than intense downpours, and perhaps the resulting greater and more prolonged soil moisture is more important to the *C. immitis* lifecycle than in summer, when rainfall often flows over the surface without soaking into the soil or is rapidly evaporated. The inclusion of winter temperatures in the

**Table 4** Variables and coefficients for each monthly model, ranked by standardized coefficient

Month	Variables	Standardized ( $\beta$ ) coeff.	<i>P</i>
January	Apr–Jun temp, 1.5 years prior	–0.305	0.032
	Dec–Jan precip, 1 year prior	–0.297	0.043
	Dec temp, 1 year prior	–0.294	0.045
	Oct temp, 1 year prior	0.283	0.042
February	Sep–Jan precip trend	0.389	0.02
	Nov–Dec temp, 2 years prior	0.330	0.04
March	Feb–Mar precip, 2 years prior	0.568	0.00
	Nov–Dec precip, 1 year prior	–0.524	0.01
	Feb–Mar precip 2 years prior $\times$ Mar–Jun temp, 2 years prior	–0.493	0.00
	Dec–Mar precip trend, 1 year prior	0.480	0.01
	Aug–Nov precip	–0.342	0.00
	Jan temp, 2 years prior	0.335	0.01
	Oct–Jan precip, 2.5 years prior	–0.237	0.03
April	Dec–Mar temp, 3.5 years prior	–0.358	0.012
	Nov temp, 2.5 years prior	–0.316	0.025
May	Feb–Apr precip	–0.372	0.010
	Aug–Dec precip, 0.5 year prior	–0.337	0.032
	Sep–Oct temp, 1.5 years prior	–0.290	0.063
	Nov–Jan temp, 1.5 years prior	0.273	0.056
June	Sep–Dec precip, 0.5 year prior	0.462	0.00
	Apr–Aug precip trend, 3 years prior	0.427	0.00
	Sep–Dec precip, 0.5 year prior $\times$ Jul–Oct temp, 0.5 year prior	–0.414	0.00
July	Jul–Dec precip, 2 years prior	0.509	0.000
	Nov–Mar temp, 1.5 years prior	0.367	0.005
	Jun precip $\times$ Jun temp	–0.271	0.031
	Apr temp, 3.5 years prior	–0.253	0.037
	Nov–Mar precip, 0.5 year prior	0.252	0.050
	Dec temp, 0.5 year prior	0.214	0.085
August	Jul–Sep temp, 1 year prior	–0.608	0.000
	Sep precip, 1 year prior	0.462	0.005
	Apr precip, 2.5 years prior $\times$ Apr–Jun temp, 2.5 years prior	0.402	0.008
	Jul temp	0.313	0.036
	Oct–Jan temp, 1.5 years prior	0.255	0.083
	Jan precip, 0.5 year prior	0.238	0.087
September	Jul–Nov temp, 1 year prior	–0.361	0.007
	Dec–Feb precip, 0.5 year prior	0.324	0.012
	Aug–Oct precip, 3 years prior	0.254	0.051
October	May–Sep temp	0.632	0.00
	May–May precip, 2.5 years prior	–0.501	0.00
	Mar–May precip, 0.5 year prior	–0.243	0.04
November	Sep precip $\times$ Sep–Oct temp	0.334	0.010
	Jan temp, 1.5 years prior	–0.333	0.005
	Jun–Sep temp trend, 3.5 years prior	–0.317	0.007
	Feb–Jun temp, 1.5 years prior	0.306	0.019
December	May temp, 1.5 years prior	–0.415	0.002
	Dec temp, 1 year prior	0.340	0.010
	Jul temp, 0.5 year prior	–0.306	0.019

models may indicate that the fungus is not able to survive at temperatures below a certain threshold, or at least that higher winter temperatures are more conducive to high incidence than cooler conditions. Given the improved ability to forecast winter temperature and precipitation in the Southwest, because of the relationship between winter climate and the Southern Oscillation Index

(Sheppard et al. 2002), it is fortuitous that our incidence models rely more on winter than on summer variables. It was expected that the interaction terms would be useful predictors given the importance of soil moisture, but only four were included in the final regression models.



## Model evaluation

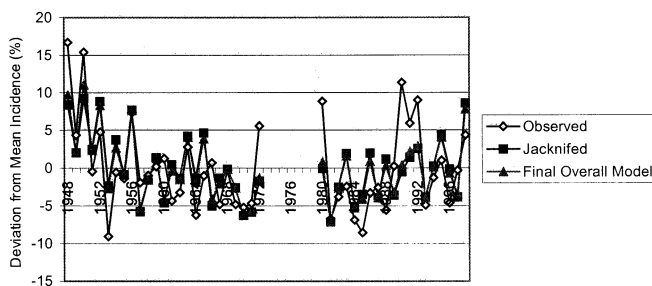
All models were evaluated using the independent data and predictions from the jack-knife process when the models were developed. The coefficient of determination ( $r^2$ , or explained variance) on independent data ranges from 0.15 (February) to 0.50 (March) (Table 5). In all cases, the  $F$  statistic associated with the model was significant ( $\alpha = 0.05$ ). The best model results in terms of explained variance were found in models for months that have the highest percentage of total annual incidence, so we are better able to predict incidence in the months that are of the highest concern. It may be that months with the highest incidence have a clearer climate signal than months with low incidence where the apparent relationship between climate and incidence may be the result of the noisiness of the data. The root-mean-squared error (RMSE) was calculated for each model (Table 5), and ranges from 27% to 50% of the mean transformed incidence values. Although RMSE values are high, those months with the highest percentage of total annual incidence have lower RMSE values than months with low percentages. The models are able to predict independent points fairly well but, as is common with regression, they fail to capture extremes in many cases. Future models could improve upon these experimental models by including population variables, such as the influx of elderly people during fall and winter months. Figure 12 illustrates the ability of the November model to predict observed values. The March (Fig. 13) and February

(Fig. 14) models, with the highest and lowest  $r^2$  values respectively, are included for comparison.

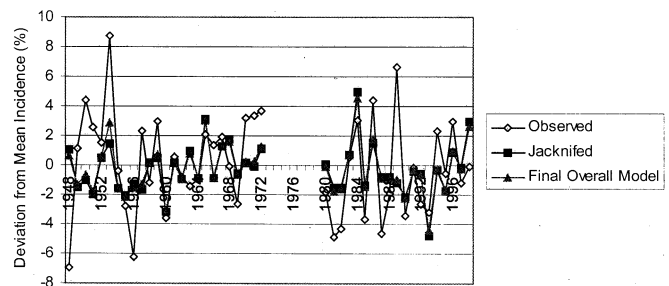
Residuals were examined in an attempt to explain the portion of the variance unaccounted for by the model variables. No clear consistent pattern is apparent in the residuals that can be explained by a variable that was not included. However, it is likely that a portion of the unexplained variance in incidence is due to climatic events that occur on shorter time scales. Individual wind and dust events occurring on a daily or weekly basis affect incidence; however, they are not captured in the model

**Table 5** Evaluation statistics for the 12 monthly models. Root-mean-squared error (RMSE) percentages of average deviation from mean incidence are moderate for months with high percentage of annual incidence

Month	Model $r^2$ (Adj. $r^2$ )	Independent $r^2$	RMSE (%)
Jan	0.32 (0.24)	0.17	32
Feb	0.26 (0.21)	0.15	44
Mar	0.67 (0.61)	0.50	33
Apr	0.40 (0.33)	0.29	47
May	0.28 (0.21)	0.14	50
Jun	0.44 (0.40)	0.34	34
Jul	0.53 (0.45)	0.31	37
Aug	0.42 (0.32)	0.21	42
Sep	0.42 (0.38)	0.32	52
Oct	0.49 (0.45)	0.44	27
Nov	0.53 (0.48)	0.40	38
Dec	0.39 (0.34)	0.41	40

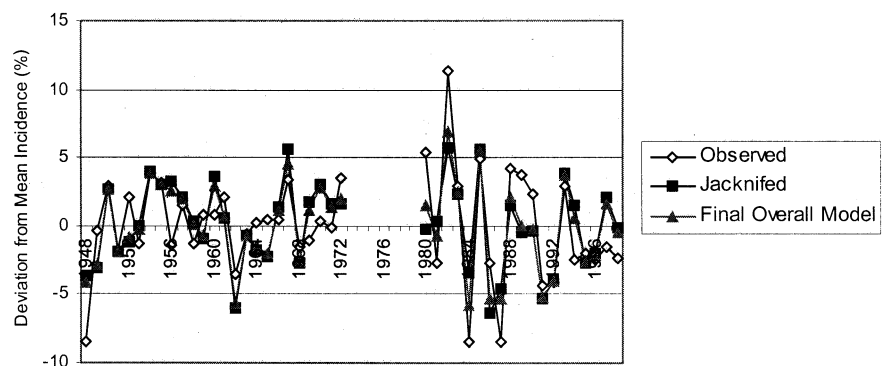


**Fig. 12** The November model is able to predict the variation in deviation from mean valley fever incidence (data unavailable 1973–1979); however, it fails to capture extreme values in most cases



**Fig. 14** The February model, with the lowest  $r^2$  value (0.15 on independent data), fails to predict many of the observed values accurately

**Fig. 13** Although extreme values are still missed in the March model, it is able to capture much of the year-to-year variability and had the highest  $r^2$  value (0.50 on independent data) of all monthly models



because of the limitations of monthly data. Also, soil moisture is likely an important factor in the lifecycle and dispersal of *C. immitis*, and therefore some way should be found to include PDSI as an indicator of antecedent conditions in future modeling work. Finally, anthropogenic factors including changes in land use and construction activity may also account for a portion of the unexplained variance.

#### Concluding remarks

The first portion of this study consisted of an exploratory data analysis that sought to identify the basic relationships between climate conditions and valley fever incidence. The bivariate and composite analyses provided insight into the conditions up to 4 years prior to a month with high or low incidence. This process also aided in the selection of candidate variables for the multivariate models. Predictive models were developed using a backward stepwise regression, and incorporated temperature and precipitation variables at varying periods prior to the month being predicted. The resulting models included variables that were mainly from periods of more than 1 year prior to the month being predicted. Also, winter climate conditions appear to be important incidence predictors, as winter temperature and precipitation variables frequently appear in the models. Months with the highest percentage of total annual incidence have the best-performing models, according to  $r^2$  and RMSE. Therefore, we are best able to predict incidence in the months that experience the greatest number of cases.

Several hypotheses in the literature were supported by our findings, while evidence was not apparent for others. The results of the compositing analyses were consistent with the hypothesis regarding the timing of soil moisture conditions on the growth and dispersal of *C. immitis*, whereby moisture is required for the fungus to grow in large amounts within the soil, but a dry period is required for airborne dispersal. This pattern was found in the composite graphs for months experiencing above-average incidence. Also, the soil sterilization hypothesis was indirectly supported by findings in the bivariate analysis linking temperature and incidence. The positive relationship between incidence and summer temperatures indicates that high summer temperatures may lead to a high number of cases in the months that follow, perhaps because of decreased soil moisture or soil sterilization. During the study, interesting relationships were found that had not previously been documented that may be real, or may be artifacts of the data processing and analyses. They are statistically significant, however, and should be investigated further. The importance of winter precipitation and temperature variables in the models points to the winter season as having more of an impact on year-round incidence than other seasons. This result was unexpected given previous studies and the exploratory analysis. In future studies, more attention should be given to the role of winter climate in influencing incidence.

Valley fever incidence is increasing within the endemic zone in Arizona as the general population grows, as well as in the population of susceptible groups (Ampel et al. 1998; Galgiani 1999). The rate of valley fever more than doubled between 1997 and 2001 (ADHS 2002). Previous research has linked valley fever incidence with climate conditions. This study adds to that literature by improving our understanding of the complex relationship between incidence and climate, specifically temperature and precipitation, and by developing monthly predictive models that will be used experimentally for future model improvement and development. We are working with state health officials as well as researchers within the Valley Fever Center for Excellence at the University of Arizona to improve the monthly models, so that in the future they may serve as a guide to the likelihood of above or below-average incidence in the coming months. Pima County is currently experiencing an upward trend in cases (Ampel et al. 1998; Galgiani 1999), and model results can be used to identify whether a portion of the increase is related to climate. Given past, current or forecast temperature and precipitation conditions, the user can determine if incidence will be high in future months as a type of early warning system. We envisage that, as understanding and model accuracy improve with continuing study, the information can be passed on to health-care providers who can prepare for increased cases by ensuring that the proper treatment is available. Also, doctors in other regions may recommend that susceptible people do not travel to or through the endemic zone if conditions are right for increased cases, and people in occupations such as archeology or geology might use special precautions when working in the field or plan work during a time when climate conditions indicate that exposure is less likely. Model runs using forecast climate conditions are sensitive to the quality of those forecasts, which must be considered by the user.

Future modeling could be combined with spatial variables including soil type, disturbance regime, and proximity to a riparian zone (Kolivras et al. 2001). An analysis of wind gust data and an appropriate measure of soil moisture would be very useful to aid our understanding of the relationship between climate conditions and valley fever incidence. The noisiness of the incidence data may make them a good candidate for non-linear approaches such as artificial neural networks. Finally, better time-series data for valley fever or *C. immitis* could greatly improve analyses and models. More consistent incidence statistics may be available for particular populations (e.g., military bases, campus health services), and fungal spore count data may become available as laboratory techniques are developed for accurate testing of soil samples.

**Acknowledgements** We would like to thank Diana Liverman, Stephen Yool, Neil Ampel, Mark Bultman, John Galgiani, Ken Komatsu, Mary Kay O'Rourke, and Lisa Shubitz for their helpful input to this paper. We would also like to thank Marc Orbach, and Seumas Rogan for useful discussions and the anonymous review-

ers for their valuable suggestions. Support for this research was provided by the National Oceanic and Atmospheric Administration's Office of Global Programs as part of the Climate Impact Assessment for the Southwest (CLIMAS). This study conforms with the current laws of the United States.

## References

- ADHS (Arizona Department of Health Services) (2002) Reported rates of communicable disease in Arizona, 1996–2001. Available: <http://www.hs.state.az.us/phs/oids/epi/9601rates.htm>
- Ampel NM, Mosley DG, England B, Vertz PD, Komatsu K, Hajjeh RA (1998) Coccidioidomycosis in Arizona: increase in incidence from 1990 to 1995. *Clin Infect Dis* 27:1528–1530
- Einstein H, Johnson R (1992) Coccidioidomycosis: new aspects of epidemiology and therapy. *Clin Infect Dis* 16:349–356
- Fiese M (1958) Coccidioidomycosis. Thomas, Springfield, Ill, pp 24–32
- Galgiani JN (1999) Coccidioidomycosis: a regional disease of national importance. *Ann Intern Med* 130:293–300
- Guttman N (1991) A sensitivity analysis of the Palmer Hydrologic Drought Index. *Water Resour Bull* 27:797–807
- Hughenoltz P (1957) Climate and coccidioidomycosis. In: Proceedings of the Symposium on Coccidioidomycosis, Phoenix, Az. Was US Public Health Service Publication 575. US Public Health Services, Wash, DC, pp 136–143
- Johnson WM (1981) Occupational factors in coccidioidomycosis. *J Occup Med* 23:367–374
- Kirkland T, Fierer J (1996) Coccidioidomycosis: a reemerging infectious disease. *Emerging Infect Dis* 2:192–199
- Kolivas KN, Johnson P, Comrie AC, Yool SR (2001) Environmental variability and coccidioidomycosis (valley fever). *Aerobiologia* 17:31–42
- Maddy K (1957) Ecological factors of the geographic distribution of *Coccidioides immitis*. *J Am Vet Med* 130:475–476
- Maddy K (1958) The geographic distribution of *Coccidioides immitis* and possible ecological implications. *Ariz Med* 15:178–188
- Maddy K (1965) Observations on *Coccidioides immitis* found growing naturally in soil. *Ariz Med* 22:281–288
- Maddy K, Coccozza J (1964) The probable geographic distribution of *Coccidioides immitis* in Mexico. *Bol Oficina Sanit Panam* July: 44–54
- Mardo D, Christensen RA, Nielson N, Hutt S, Hyun R, Shaffer J, et al (2001) Coccidioidomycosis in workers at an archeologic site – Dinosaur National Monument, Utah, June–July 2001. *Morb. Mortal Wkly Rep* 50:1005–1008
- Pappagianis D (1980) Epidemiology of coccidioidomycosis. In: Stevens D (ed) *Coccidioidomycosis, a text*. Plenum, New York, pp 63–85
- Pappagianis D (1988) Epidemiology of coccidioidomycosis. *Curr Top Med Mycol* 2:199–238
- Reed RE (1960) Ecology and epizootiology of coccidioidomycosis. *Rocky Mt Vet* 8(10):10–16
- Sheppard PR, Comrie AC, Packin GD, Angersbach K, Hughes MK (2002) The climate of the Southwest. *Clim Res* 21:219–238
- Smith CB, Beard R, Whiting E, Rosenberger H (1946) Varieties of coccidioid infection in relation to the epidemiology and control of the disease. *Am J Public Health* 36:1394–1402
- Stevens D (1995) Coccidioidomycosis. *N Engl J Med* 332:1077
- United States Census Bureau (2001) Available: <http://www.census.gov> (cited 5 May)
- Valley Fever Center for Excellence (2001) Available: <http://vfce.arl.arizona.edu> (cited 7 June)
- Werner SB, Pappagianis D (1973) Coccidioidomycosis in northern California – an outbreak among archeology students near Red Bluff. *Calif Med* 119:16–20