

## Statistical Modeling of Valley Fever Data in Kern County, California\*

Jorge Talamantes (1), Sam Behseta (2), and Charles S. Zender (3)

(1) (Corresponding Author) Department of Physics and Geology, 14 SCI, 9001 Stockdale Highway, California State University, Bakersfield, CA 93311, USA. Email: [jtalamantes@csub.edu](mailto:jtalamantes@csub.edu). Tel.: +1-661-654-2335, Fax: +1-661-654-2040

(2) Department of Mathematics, California State University, Bakersfield, USA

(3) Department of Earth System Science, University of California, Irvine, USA

**Abstract.** Coccidioidomycosis (valley fever) is a fungal infection found in the southwestern US, northern Mexico, and some places in central and South America. The fungus which causes it (*Coccidioides immitis*) is normally soil-dwelling but, if disturbed, becomes air-borne and infects the host when its spores are inhaled. It is thus natural to surmise that weather conditions which foster the growth and dispersal of the fungus must have an effect on the number of cases in the endemic areas. We present here an attempt at the modeling of valley fever incidence in Kern County, California by the implementation of a Generalized Auto Regressive Moving Average (GARMA) model. We show that the number of valley fever cases can be predicted mainly by considering only the previous history of incidence rates in the county. The inclusion of weather-related time sequences improves the model only to a relatively minor extent. This suggests that fluctuations of incidence rates (about a seasonally-varying background value) are related to biological and/or anthropogenic reasons, and not so much to weather anomalies.

Keywords: GARMA Modeling, Time Series Analysis, Valley Fever Prediction, Coccidioidomycosis, *Coccidioides immitis*.

---

\* In press in *International Journal of Biometeorology*. The original publication is available at [www.springerlink.com](http://www.springerlink.com). DOI 10.1007/s00484-006-0065-4

1. Introduction. Much is known about the biological, medical and indeed the epidemiologic aspects of *Coccidioides immitis*, the fungus which causes valley fever (coccidioidomycosis). See, e.g., Pappagianis (1988), and references therein. *C. immitis* has a complete life cycle as a soil-dwelling organism, but if it is disturbed and becomes air-borne, it is able to infect a host via the respiratory tract when it inhales the fungus spores. About 60% of infected patients report no symptoms (Lee 1944); about 25% exhibit severe flu-like symptoms such as cough, sputum, fever, and muscle aches; the remaining 15% become very ill with pneumonia-like symptoms (e.g. pleurisy and heavier sputum) requiring medication and bed rest. In a small number of these cases (about 0.5-1%), the disease disseminates beyond the lungs to e.g. the skin, bones, and/or meninges, and it can be fatal. Certain sectors of the population seem to be more susceptible to infection, such as the very young, persons newly arrived to the endemic areas (since immunity develops with infection), and those with impaired immune systems (Pappagianis 1988).

Given its wide geographic distribution, it is apparent that *C. immitis* is able to flourish within somewhat varied climatic environments. Endemic areas include the southern part of the San Joaquin Valley in California, southern California, the southern part of Arizona, New Mexico and Texas, most on northern Mexico, and some areas in Guatemala, Honduras, Venezuela, northeastern Brazil, Argentina and Paraguay (Pappagianis 1994). There are some variations in climate in these areas, for example, the southern San Joaquin Valley gets most of its precipitation in the winter, whereas the southern part of Arizona (notably Pima, Pinal, and Maricopa counties, where many coccidioidomycosis cases are reported) gets late summer southwest monsoon rains.

In Kern County, California, where the present study was conducted, the seasonal climate was described in detail in Zender and Talamantes (2006). The area receives in the neighborhood of 16.5 cm of rain a year, largely in the winter. Summers are hot and dry – usually reaching 43 C in July, with virtually no precipitation. May and June experience the largest wind speeds – and average of the order of 3.5 m/s, with maxima usually less than 15 m/s, and very rarely exceeding 20 m/s. The incidence rate (number of cases per 100, 000 population) also has a corresponding yearly cycle, with the number generally increasing towards the late fall (with an incidence of the order of 17 per month per 100, 000 population), decreasing in the winter, and reaching a minimum in the spring and summer (with an incidence of the order of 4.7 per month per 100, 000 population). It is thus natural to expect that environmental fluctuations might affect the rate at which humans become infected (see, e.g. Pappagianis 1988). For example, a wetter-than-normal rainy season should help *C. immitis* bloom; windy spells might facilitate the dispersal of the fungus; hot summers can be anticipated to suppress competing organisms, thus enhancing the survival of *C. immitis* (Kolivras *et. al.* 2001). Indeed, the medical community, and the community of Kern County at large have long held that windy days, in particular dust storms, increase the rate of infection. Such anecdotal evidence is well-documented in the literature [See, e.g., Smith *et. al.* (1946), Hugenholtz (1957), Maddy (1957), Pappagianis and Einstein (1978), Centers for Disease Control and Prevention (1993, 1994, 1996, and 2003), Kirkland and Fierer (1996), Schneider *et. al.* (1997)]. To our knowledge, however, there have been only two mathematical attempts at quantifying and demonstrating this connection. The first such work was presented by Kolivras and Comrie (2003), who found a significant connection between temperature and

precipitation, and incidence in Pima County, Arizona. In contrast, the second work, Zender and Talamantes (2006), found only a weak correlation between anomalies in weather variables and incidence in Kern County. The discrepancy might have been due to the fact that climate exhibits important differences in Pima and Kern Counties – for example, Pima experiences monsoon rains, but Kern does not. The mathematical treatments were also different. The differences in the two approaches are mainly in the processing of raw incidence and weather data – although both papers analyzed correlations in a linear model.

Understanding what gives rise to fluctuations in the number of valley fever infections has obvious practical implications. The 1991-1995 epidemic in Kern County had important economic repercussions (Jinadu 1995), as well as human suffering. This is presumably also true of the 2001-present surge in valley fever incidence in Kern County (Larwood, Emery 2005). According to Jinadu (1995), the cost associated with coccidioidomycosis in Kern County in the period 1991-1994 was of the order of \$66.6 million, with hospitalization requiring 63% of that sum, doctor's visits 18%, lost wages 12% and the rest was spent on drugs.

In this paper, we continue to try to assess the effect of weather on valley fever incidence in Kern County, and present a model which might be helpful in predicting outbreaks. We look at nine years of weekly-mean atmospheric conditions and try to ascertain if these records help us predict the weekly incidence some time later. We do not examine the role of longer-term climate fluctuations – the available incidence record is insufficient for that purpose. We examine weekly incidence data starting in January 1995 and ending in December 2003. During this period, weekly incidence varied from zero to about 8 cases per week per 100, 000 population. The objectives here are different from Zender and Talamantes (2006) – that work looked for linear correlations between incidence and monthly climate fluctuations; here, we strive to find a statistical model which allows a degree of prediction. In section 2 we describe the sources and nature of the data we used for the work presented here. In section 3 we specify some mathematical details of the model used in this paper. In section 4 we present the results obtained from that regression, and in section 5 we discuss the effectiveness of the regression and offer some conclusions.

**2. Data description.** We obtained the weekly number of reported cases (Jan 1995 – Dec 2003) from the Kern County Department of Public Health. These data were then normalized to 100, 000 population, and constitute the time series we refer to as incidence  $N$ . We note that, since many valley fever cases go unreported, incidence is only an approximation to the actual situation (Pappagianis 1988, 1994, Larwood and Emery 2005). Clearly, one can only estimate the level of under-reporting – an infected person may never be tested; however, according to Emery (2006), the Kern County Health Department Laboratory performs about 95% of all the coccidioidomycosis tests in the county. Reporting then takes place directly through laboratory surveillance, as opposed to individual health care providers doing the reporting; furthermore, the Kern County Health Department has been using the same standardized case definition (developed in collaboration with the US Centers for Disease Control and Prevention) through the period of study. Reporting is therefore as reliable and thorough as one might hope. Most

importantly, the level of reliability has not changed during the period of interest – the analysis we present in this paper is not sensitive to an overall constant scaling factor, and so our conclusions are unaffected by a time-independent level of under-reporting.

Weather data were obtained from the National Climatic Data Center (NCDC). The raw data consisted of series of hourly measurements of (among other parameters) temperature ( $T$ ), precipitation ( $P$ ), and wind speed ( $U$ ). All the measurements were taken at Meadows Field, the county airport in Bakersfield, California. Thus, these time series too must be considered an approximate proxy to the actual parameters of interest, *i.e.*, we could not possibly account for the many micro-climates that might give rise to enhanced valley fever infections in different parts of the county. Furthermore, direct measurements of *C. immitis* air-borne, or even soil abundance do not exist. Instead, we use the wind speed at Meadows Field – which is removed several steps from the desired data. We admit that the use of these parameters can be the cause of some debate as to the interpretation of our results, but these data sets are probably as good as any other one might be able to obtain. We computed weekly-average time series of the weather parameters from the original NCDC raw time sequences. Our subsequent analysis only used those weekly averages. (This part of our procedure is similar to Zender and Talamantes (2006), except that monthly averages were used there.)

3. **Methods.** We aim to study the effect of three possible environmental predictors namely, precipitation  $P_t$ , wind speed  $U_t$ , and near-surface air temperature  $T_t$  on the response variable  $N_t$  – the reported incidence of valley fever cases in Kern county. ( $t$  is the time in weeks.) We attempt to build a statistical model (linear or non-linear) to determine two components: (i) the degree or pattern of each predictor's effect on the response, and (ii) the statistical significance of these predictors.

Developing a statistical model in the context of this problem is a rather non-trivial task due to two reasons. First, the valley fever data form a time series. This violates a fundamental assumption in the usually practiced generalized linear modeling in which the data are assumed to be independent. Second, it is quite reasonable to assume that valley fever counts at any given time  $t$  follow a Poisson distribution with the intensity rate  $\lambda$ , but this assumption complicates the model even more because the normality assumption does not hold. Nonetheless, both features of the serial nature of the response variable and the Poisson assumption are typical identifiers of modeling with Generalized Auto-Regressive Moving-Average (GARMA) modeling.

GARMA modeling was introduced by Benjamin *et. al.* (2003), who modeled poliomyelitis cases reported to the US Centers for Disease Control and Prevention for the years 1970 to 1983. To our knowledge, this is the only time that GARMA has been used to model disease incidence data. There, the authors specifically treat the case of Poisson regression under a setting which they refer to as GARMA ( $p, q$ ) models, short for generalized autoregressive of order  $p$  and moving average of order  $q$ . The fundamental idea is to be able to apply the widely used class of Auto-Regressive Moving Average (ARMA) models to non-Gaussian data. The ARMA model assumes that the time-series is strictly generated by a Gaussian model. As argued in the above, in cases such as in

modeling the valley fever data, the process of interest consists of counts; hence, the normality assumption does not hold. The Poisson assumption here makes the modeling scheme more complex. Benjamin *et. al.* (2003) demonstrated the statistical theory needed to estimate the parameters of the model using the Maximum Likelihood Estimation (MLE). When applying generalized linear models whose response is counts, one may need to take into account the extra variation that stems from the implication of Poisson distribution at time  $t$ . This important issue was presented in a paper by Breslow (1984), which lead to the idea of quasi-Poisson likelihood regression modeling of the count time series. [Kedem and Fokanis (2002) have written a text which gives a comprehensive review of regression modeling for time series.]

In essence, we examine here three different GARMA models. We begin our analysis by regarding  $N_t$  (the incidence at time  $t$ ) as the one response variable. We start by considering a GARMA model whose predictors (or independent, or explanatory variables) consist of the autoregressive terms of various lags (*i.e.* the history of  $N$ ) *as well as* the weather variables ( $P$ ,  $T$ , and  $U$ ). We also include a set of predictors whose role is to account for the seasonality of the incidence. The collection of predictors consists of the following set of 24 independent variables: the three weather parameters  $U_t$ ,  $T_t$ , and  $P_t$ , along with six predictors to account for the periodic nature of the series, namely,  $\cos(2\pi t / 52)$  and  $\sin(2\pi t / 52)$  to account for the annual seasonality,  $\cos(2\pi t / 26)$  and  $\sin(2\pi t / 26)$  to represent any semi-annual seasonality, and  $\cos(2\pi t / 13)$  and  $\sin(2\pi t / 13)$  to cover potential quarter-annual fluctuations.

Additionally, we consider 15 autoregressive terms  $N_{t-k}$  of lags  $k$  equal to 1 to 13, 26, and 52 weeks. We note that at this point, we introduce autoregressive terms  $N_{t-k}$  purely for mathematical reasons: our intent is to predict the value of  $N_t$  – we thus put into the model any variables we suspect might even remotely help us in this prediction task, while at the same time keeping the number of such explanatory variables to a minimum to keep computational expense down to a reasonable level. We then estimate the coefficients of the model with the method of MLE (see, *e.g.*, Box, Jenkins and Reinsel, 1994, and Benjamin *et. al.*, 2003). We call this the full model.

Clearly, *if* the full model contains sufficient *relevant* information, one should be able to predict the time series with such a large pool of predictors. However, pragmatically, it is preferable to work with parsimonious models while with more or less the same prediction capabilities. These reduced models also shed light on the relative importance of the various predictors. This reduction necessitates the application of variable search and selection techniques.

We next refine the full model by searching through the space of all possible sub-models. To implement the search we employ a likelihood-based measurement known as the Akaike Information Criterion or the AIC (Akaike, 1974). A better model has a smaller AIC. It should be mentioned that searching for the best model in this context is computationally expensive due to the fact that there exist  $2^n$  sub-models in the presence of  $n$  explanatory variables. Eventually, we find an optimal model whose AIC is as small as possible. We name this the final model. We should point out that the minimization procedure is a standard way to evaluate a statistical model. Our modeling approach is

quite general; however, model parameters and their significance will change as one attempts to model a different time segment of the incidence time series. We would certainly expect those parameters to be different if we considered different data sets.

With the aid of graphical presentation and the calculation of the sum of squares of errors or SSE, we show that the final model may be used to predict the nuances in the incidence time-series. We prefer the use of our final model due to its parsimony, its easier interpretation, and most importantly, the statistical significance of its parameters. Finally, to evaluate the sole effect of the weather-related variables we implement a model with only these effects while excluding the autoregressive terms. We show that this model falls short in predicting the stochastic shocks to the incidence time series. We term this the environmental model.

4. Results. Our AIC-based stepwise search algorithm yields the following estimated model, which we refer to as the final model:

$$\ln \lambda_t = 0.21 - 0.23 \sin(2\pi t / 52) + 0.12 N_{t-1} + 0.07 N_{t-2} + 0.05 N_{t-4} + 0.06 N_{t-26} \quad ,$$

in which  $\lambda_t$  represents the rate of the Poisson function of incidences, and the natural logarithm is the link function of the generalized linear model. We note that 0.21 is the constant that corresponds to the background incidence, and the sinusoidal term with the periodicity of 52 weeks confirms the significance of the annual seasonality of the time series. The most important finding here is the absence of weather effects in the final model. The AIC for the final model is 1324, with a residual deviance of 419.8 which follows a chi-squared distribution with 409 degrees of freedom. One may take this as a confirmation for the goodness of fit since the residual deviance and the degrees of freedom are close (Venables and Ripley, 1997). In figure 1, the top and middle panels demonstrate the predicted time series resulting from the full and the final models, respectively. In each panel, the prediction time series is overlaid on the incidence data. The SSE for the full model is 368, and 466 for the final model. The statistical significance for the terms included in the final model is shown in table 1.

The inclusion of  $N_{t-k}$  terms but not  $P_{t-k}$ ,  $U_{t-k}$  and  $T_{t-k}$  in our full model clearly puts incidence history on a different footing *a priori*. To ascertain the effect of the asymmetry between incidence and weather histories, we therefore performed two additional and separate analyses: (i) we included  $P_{t-k}$ ,  $U_{t-k}$  and  $T_{t-k}$  terms in our full model with lags  $k = 1, \dots, 13$  weeks, and (ii) we included those weather variables with lags  $k = 1, \dots, 26$  as well as terms  $P_{t-k}$  with  $k = 30, \dots, 34$  (which was suggested by the weak correlation between incidence and precipitation eight months prior found in Zender and Talamantes 2006). We found that the inclusion (in the initial choice of full model) of historical terms associated with weather effects (*i.e.*  $P_{t-k}$ ,  $U_{t-k}$  and  $T_{t-k}$ ) does not affect our final model (after our AIC-based minimization search). An important feature of the AIC-based model selection is that most often, the terms included in the final model are statistically significant. (But there is no guarantee that this is the case.) As a matter of fact, we observed this to be true in our application. This outcome is reassuring, since it resulted

automatically from the AIC minimization procedure, as opposed to manually forcing the deletion of any terms. We note, however, that the inclusion of so many parameters in the initial (full) model makes it difficult to interpret the results. This is because we are trying to model a valley fever data set consisting of only  $9 \text{ years} \times 52 = 468$  data points. In analysis (i), the number of parameters is  $24 + 3 \times 13 = 63$  -- of the order of 14% of the number of data points, but in analysis (ii) the number of parameters is  $24 + 3 \times 26 + 5 = 107$ , which is not much less than the number of data points. Thus, ultimately, our main justification for not explicitly including the histories of precipitation, wind speed and temperature in our full model comes *a posteriori*: the resulting predictions of the final model are quite close to the observed valley fever incidence data set.

To examine the use of the weather effects in predicting incidence, we ran a model whose predictors consist only of the three weather variables (the environmental model). The last panel in fig. 1 reveals the weakness of the three variables of speed, temperature and precipitation in the context of prediction. The SSE equals 943 for this model.

**5. Discussion and Conclusions.** We have presented a model that allows us to predict the incidence  $N_t$  of valley fever at some  $t$ , given the incidence history for earlier times.

We presented in section 3 various models which include different amounts of input data. When the time series corresponding to  $N$ ,  $U$ ,  $T$ , and  $P$  are all accounted for, the fit to the observed incidence of valley fever is very good as measured by the SSE. When only the previous incidence history is included in the model, however, the SSE does not increase by much. When we include only the  $U$ ,  $T$ , and  $P$  histories in the model, the predicted incidence values are significantly worse, again, as measured by the SSE. Our main conclusion is therefore that the dependence of incidence rate fluctuations on weather parameters in Kern County is rather weak. Evidently, Kern has the right environment for *C. immitis* to thrive, but given that the fungus is well-established, and that this causes a certain seasonally-dependent incidence background, the *fluctuations* about that background tend to exhibit only a weak dependence on weather events. This result is entirely consistent with those of Zender and Talamantes (2006), where only a weak linear (lag) correlation between monthly-mean climate and valley fever incidence was found. We surmise that the reasons behind the observed fluctuations in valley fever incidence in Kern County (*vis á vis* the 1991-1995 epidemic, and the 2001-present surge) are biological and/or anthropogenic in nature – most likely new construction on previously-undisturbed soil. We suggest that weather fluctuations, at least on the weekly scale investigated in this work and the monthly scale used in Zender and Talamantes (2006), are too small to explain fluctuations in incidence. This is especially clear with temperature; it is most likely also true for precipitation, which exhibits the largest fluctuations of the weather variables investigated.

Our model allows for the prediction of the value of the incidence  $N_t$  at time  $t$  given certain required values  $N_{t-k}$ . This is a prediction in the following sense. Given the *measured* values of  $N_{t-k}$ , the model “returns” a prediction  $N_t$ . Prediction values are obtained by plugging in the maximum likelihood estimations of the coefficients into our final model. We note that attempting to use this model to *forecast* values  $N_t$  (where  $t$  is in

the future) would be unwise because one would need to rely on values  $N_{t-k}$  which themselves were predictions (not *measurements*). The uncertainty in such forecasted values of  $N_t$  would thus grow very rapidly.

It is undeniable that the available data are not of optimal quality – we use weather observations at the county airport and take them as representative of the whole county; the reported incidence numbers are most likely less than actual infection rates (Pappagianis 1988, Larwood and Emery 2005); we use precipitation and wind speed as proxy for *C. immitis* abundance in the soil and the atmosphere; it takes some variable number of weeks (perhaps 5 to 6) between infection and reporting (Larwood 2003). These problems add to the difficulty in interpreting our results; nevertheless, it is quite surprising and reassuring that the modeling results come out so close to the observations. This is probably due to the strong dependence of incidence on its own history, and so modeling the problem is (fortuitously) only weakly dependent on the details of the fluctuations of the environmental data. The fact that our model so closely resembles reported incidence data is an indication that our approach in minimizing the model error has been successful.

At this point, the most likely and feasible direction our work might take is to investigate the universality of our results. That is, it would be interesting to see whether modeling incidence data in other areas of the southern San Joaquin Valley might give qualitatively similar results to Kern County. Perhaps more interestingly, we would want to see if endemic areas with different climates (such as Pima, Pinal and Maricopa counties in Arizona) can be modeled similarly to the results in this work, or if weather and climate anomalies explain more incidence there than in Kern County. Furthermore, we might try to examine the effect of using a different proxy for spore abundance in the atmosphere; for example, we might use  $PM_{10}$  measurements. (However, one must be mindful of the fact that these numbers may be increased due to factors completely unrelated to *C. immitis* spores, e.g. industrial processes and increased vehicular traffic.) We would like to emphasize that our findings are only interpretable within the framework of the variables that we originally included in our full model. In other words, this modeling scheme should be viewed considering that in a study of this nature there exists a number of other potentially interesting variables that may affect valley fever incidence in Kern County. The inclusion of such auxiliary sets of variables may change the overall profile of our final model as presented in this paper.

**Acknowledgements.** Thanks are due to Emma Chaput of the Kern County Department of Public Health for providing us with the weekly incidence data, and to Francisco Beltran of the CSUB Mathematics Department for assisting us with some of the computer programming. We are immensely thankful to Thomas Larwood, M.D. and to Kirt Emery, M.P.H., of the Kern County Department of Public Health for important discussions regarding this work. Finally, we thank the anonymous referees for their careful suggestions. Their comments helped us improved the manuscript considerably both in readability as well as content.

## References

- Akaike, Hirotugu (December 1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6): 716–723.
- Benjamin MA, Rigby RA and Stasinopoulos DM (2003) Generalized Autoregressive Moving Average Models, *Journal of the American Statistical Association*, 98:214-223
- Box GP, Jenkins GM, and Reinsel GC (1994) *Time Series Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 3rd edition
- Breslow NE (1984) Extra-Poisson Variation in Log-Linear Models, *Appli. Statist.* 33:38-44
- Centers for Disease Control and Prevention (1993) *Coccidioidomycosis – United States, 1991-1992*. *Morb Mortal Wkly Rep* 42: 21-24
- Centers for Disease Control and Prevention (1994) Update: coccidioidomycosis – California 1991-1993. *Morb Mortal Wkly Rep* 43: 421-423
- Centers for Disease Control and Prevention (1996) *Coccidioidomycosis – Arizona 1990-1995*. *Morb Mortal Wkly Rep* 45: 1069-1073
- Centers for Disease Control and Prevention (2003) Increase in coccidioidomycosis – Arizona, 1998-2001. *Morb Mortal Wkly Rep* 52: 109-112
- Emery K (2006) private communication.
- Hughenoltz PG (1957) Climate and Coccidioidomycosis. In: *Proceedings of the Symposium on Coccidioidomycosis*, Phoenix, Arizona. Public Health Service, Washington, D.C., pp 136-143
- Jinadu BA (1995) Valley Fever Task Force report on the control of *Coccidioides immitis*. Technical Report, Kern County Health Department, Bakersfield, California
- Kedem B and Fokianos K (2002) *Regression Models for Time Series Analysis*, Wiley, New York
- Kirkland TN, Fierer J (1996) Coccidioidomycosis: A reemerging infectious disease. *Emerg Infect Dis* 3:192-199
- Kolivras KM, Johnson PS, Comrie AC, Yool SR (2001) Environmental variability and coccidioidomycosis (valley fever). *Aerobiologia* 17:31-42
- Kolivras KM, Comrie AC (2003) Modeling valley fever (coccidioidomycosis) incidence on the basis of climate conditions. *Int J Biometeorol* 47:87-101
- Larwood T (2003) private communication.

Larwood T and Emery K (2005) private communication.

Lee RV (1944) Coccidioidomycosis in the Western Flying Training Command. *CA & Western Med*, 61:133-134.

Maddy KT (1957) Ecological Factors possibly relating to the geographic distribution of *Coccidioides immitis*. In: Proceedings of the Symposium on Coccidioidomycosis, Phoenix, Arizona. Public Health Service, Washington, D.C., pp 144-157

Nelder JA and Wedderburn RWM (1972) *Generalized Linear Models*, Wiley, New York

Pappagianis D (1988) Epidemiology of coccidioidomycosis. In: McGinnins M (ed) *Current topics in mycology*, vol 2. Springer, New York Berlin Heidelberg, pp 199-238

Pappagianis D, Einstein H (1978) Tempest from Tehachapi takes toll or *Coccidioides* conveyed aloft and afar. *West J Med* 129:527-530

Pappagianis D (1994) Marked Increase in Cases of Coccidioidomycosis in California:1991, 1992, and 1993. *Clinical Infectious Diseases* 19 (Suppl 1):S14-8

Schneider E, Hajjeh RA, Spiegel RA, Jibson, RW, Harp EL, Marshall GA, Gunn RA, McNeil MN, Pinner RW, Baron RC, Burger RC, Hutwagner LC, Crump C, Kaufman L, Reef SE, Feldman GM, Pappagianis D, Werner SB (1997) A coccidioidomycosis outbreak following the Northridge, Calif, Earthquake *JAMA* 277:904-908

Smith CE, Beard RR, Rosenberger HG, Whiting EG (1946) Effect of season and dust control on coccidioidomycosis *JAMA* 132:833-838

Venables WN, Ripley BD (1997) *Modern Applied Statistics with S-Plus*, second edition, Springer, New York.

Zender CS and Talamantes J (2006) Climate controls on valley fever incidence in Kern County, California. *Int J Biometeorol* 50:174-182

<b>Term</b>	<b>Estimate</b>	<b>Std. error</b>	<b>Z-value</b>	<b>P-value</b>
Intercept	0.21	0.066	3.125	.0017
$\sin(2\pi t / 52)$	-0.23	0.060	-3.803	.0001
$N_{t-1}$	0.12	0.0025	4.75	2.54e-6
$N_{t-2}$	0.07	0.027	2.373	0.0176
$N_{t-4}$	0.05	0.023	1.977	0.0486
$N_{t-26}$	0.06	0.024	2.540	0.0110

**Table 1.** The final model. All included terms are statistically significant.

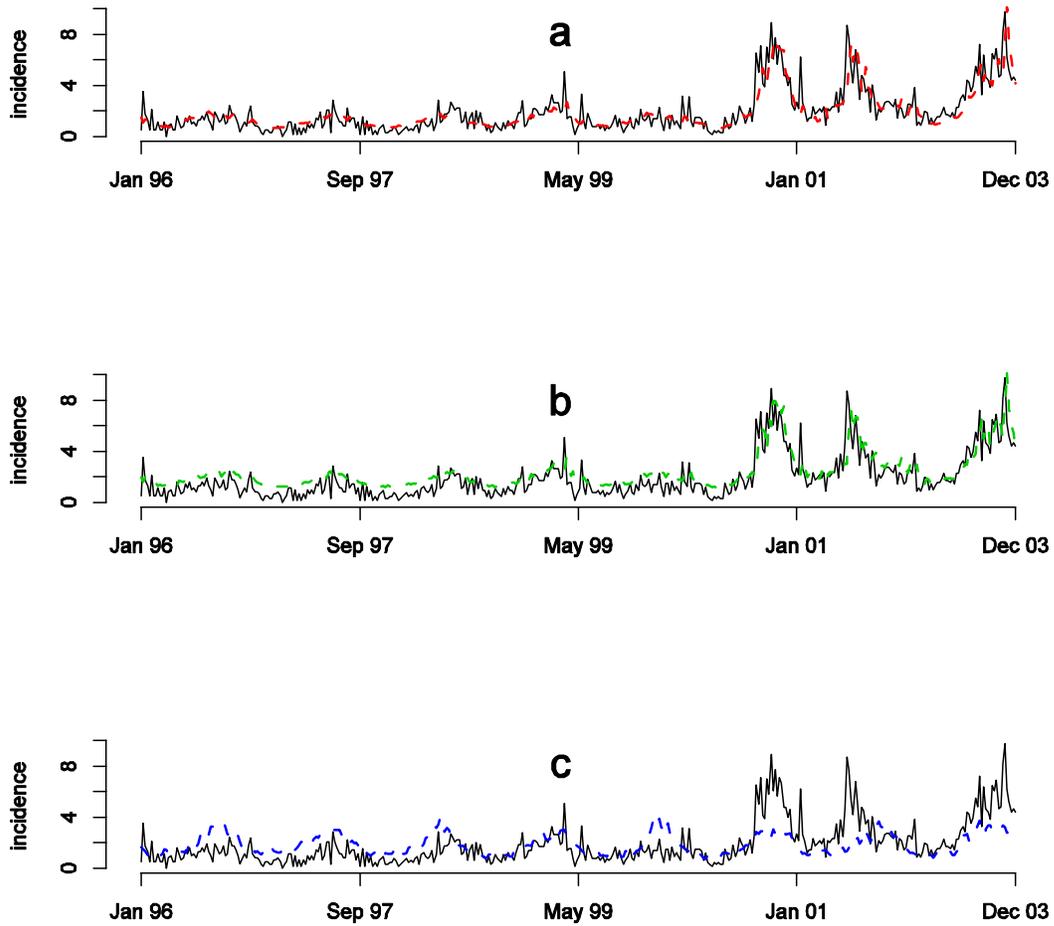


Figure Captions.

Fig. 1. Reported valley fever incidence in Kern County, California (cases per 100, 000 population) and three models for the period Jan 1996- Dec 2003. In all three panels, solid lines indicate incidence, and dashed lines correspond to model results: (a) full model, (b) final model, and (c) environmental model. The year 1995 is missing because our model requires at least one year of valley fever incidence to start predicting future values.