A statistical model linking Siberian forest fire scars with early summer rainfall anomalies

Tim E. Jupp,[1,2] Christopher M. Taylor,[2] Heiko Balzter,[1] and Charles T. George[1]

Received 24 April 2006; revised 13 June 2006; accepted 15 June 2006; published 20 July 2006.

[1] Forest fires in Siberia have a significant effect on the global carbon balance. It is therefore of interest to study the environmental factors that may be responsible for their variability. Here we examine variability in the annual number of forest fire scars at a spatial scale of 2.5°. This is decomposed statistically into a spatio–temporal component correlated with low summer rainfall, a spatial component correlated with population density and a temporal component correlated with the Arctic Oscillation. Data come from ten years of satellite–derived data, incorporating both the number of forest fire scars and monthly rainfall. The expected number of fire scars halves for each additional 0.35 mm per day of rainfall in the period April–July. Our findings may prove useful in parameterising both fire models within climate simulations and fire warning systems based on numerical weather predictions of regional dry anomalies. **Citation:** Jupp, T. E., C. M. Taylor, H. Balzter, and C. T. George (2006), A statistical model linking Siberian forest fire scars with early summer rainfall anomalies, *Geophys. Res. Lett.*, *33*, L14701, doi:10.1029/2006GL026679.

1. Introduction

[2] Forest fires in Siberia are subject to substantial interannual variability [*Zhang et al.*, 2003; *Balzter et al.*, 2005; *Sukhinin et al.*, 2004]. Total direct carbon emissions in boreal Siberia ranged from 116 Tg C in 1999 to 520 Tg C in 2002 [*Soja et al.*, 2004]. This is equivalent to 5% and 20%, respectively, of the total global carbon emissions from forest and grassland burning. It has been estimated that Russian boreal forest fires in 1998 constituted 14%–20% of average annual global carbon emissions from forest fires [*Conard et al.*, 2002] and that in extreme fire years total direct carbon emissions can be 37%–41% greater than in normal fire years [*Soja et al.*, 2004], due mainly to increased soil organic matter consumption.

[3] Extreme Siberian forest fires can lead to atmospheric pollution in North America. Satellite data and global aerosol transport models suggest that Siberian fire emissions were the primary source of three air pollution events off the coast of Washington State in 2003 [*Bertschi and Jaffe*, 2005]. On 2 June 2003 O_3 concentrations exceeded the U.S. Environmental Protection Agency 8-hour standard (>84 ppbv) in Enumclaw, Washington, and on 5 August 2003 in Seattle, Washington, daily average fine particle concentrations exceeded 18 $\mu g\ m^{-3}$ [*Bertschi and Jaffe*, 2005]. These effects have been attributed to air masses originating in a region including Central Siberia (35°N–70°N and 70°E–170°E) [*Jaffe et al.*, 2004; *Bertschi and Jaffe*, 2005].

[4] There is a trend for numerical vegetation models to include mechanistic models of forest fire [e.g., *Lenihan et al.*, 1998; *Venevsky et al.*, 2002; *Arora and Boer*, 2005]. It follows that quantitative descriptions of spatio–temporal variability are useful in the inter–comparison, validation and parameterisation of these models. Ultimately, improvements in vegetation models can affect numerical climate simulations when the two are run interactively.

[5] In this study we concentrate on 'Central Siberia', an area of about 3 million km^2 that can be defined by the administrative regions of Irkutsk Oblast, Krasnoyarsk Kray, Taimyr, Khakass Republic, Buryat Republic and Evenksky Autonomous Oblast (approx. 51°N–78°N, 79°E–119°E).

[6] Previous research found that interannual variability in burned forest area in Central Siberia for the period 1992–2003 could largely be reproduced by a linear combination of two factors: large-scale climate represented by the Arctic Oscillation index and regional conditions represented by the summer temperature [*Balzter et al.*, 2005]. In this study we investigate spatial patterns in the number of fire scars within Central Siberia. The number of fire scars is modelled statistically with parameters estimated using Bayesian techniques. We examine possible influences on the parameters in the model, and relate them to factors likely to influence fire frequency.

2. A Statistical Model

[7] In this study, forest fire scar information was derived from satellite data from the AVHRR, ATSR–2 and MODIS instruments [*Balzter et al.*, 2005] while rainfall data were taken from the Global Precipitation Climatology Project (GPCP) [*Adler et al.*, 2003]. The GPCP rainfall data are provided at a spatial resolution of 2.5° by 2.5° and a temporal resolution of one month. We therefore work within an 8 × 14 spatial array of gridcells covering the region 50°N–70°N, 85°E–120°E at this resolution (Figure 1). The original fire scar data consist of the month, area, latitude and longitude of 2575 forest fire scars in the period 1992–2003, with the exception of 1994 and 1995 for which data are unavailable. We therefore define masked spatial and temporal arrays by removing the years 1994 and 1995 as well as any gridcells in which no fire scars were recorded in any year. This leaves a final spatial array of $M = 69$ gridcells (indexed by $i = 1, \ldots, M$) and a final temporal array of $N = 10$ years (indexed by $j = 1, \ldots, N$) in which we can calculate the total number of fire scars n_{ij} and the total area of fire

[1]Climate and Land-Surface Systems Interaction Centre, Centre for Ecology and Hydrology, Monks Wood, UK.
[2]Climate and Land-Surface Systems Interaction Centre, Centre for Ecology and Hydrology, Wallingford, UK.

0094-8276/06/2006GL026679$05.00

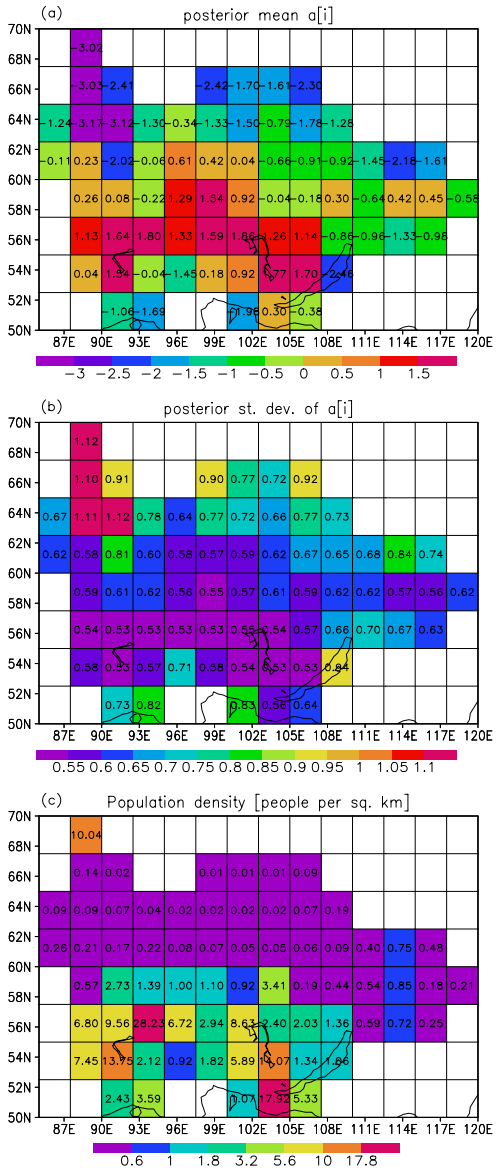

Figure 1. (a) Posterior mean estimates for a_i. (b) Posterior standard deviation of a_i. (c) Population density for the same region. Data from http://sedac.ciesin.columbia.edu/gpw, here aggregated to 2.5°. Note logarithmic colour scale. Lake Baikal is the feature around 54°N, 108°E.

scars a_{ij} in gridcell i and year j. The mean area per fire scar over the whole data set is $\mu = 44.76$ km^2 with standard deviation $\sigma = 110$ km^2. From sampling theory, one would expect $a_{ij} \approx n_{ij}\mu \pm \sqrt{n_{ij}}\sigma$ if the number of fire scars were a reliable indicator of total burned area. This is indeed the case, and so for simplicity we model the number of fire scars rather than the burned area.

[8] Of the 2575 fire scars, the majority (2306) occur in the months April–July and so we define the early summer rainfall R_{ij} (for gridcell i and year j) to be the mean rainfall in mm day^{-1} for this period.

[9] With these conventions, the data consist of $MN = 690$ observations of the number of fire scars n_{ij} and the early summer rainfall R_{ij} across 69 gridcells and 10 years. For later convenience it is helpful to define 'rainfall anomalies' r_{ij} for each gridcell by

$$r_{ij} = R_{ij} - \frac{1}{N}\sum_{j=1}^{N} R_{ij} \tag{1}$$

[10] The observed number of fire scars n_{ij} can be regarded as a realization of a discrete random variable N_{ij}. Since this represents an integer number of events occurring in fixed intervals of space and time, one might expect it to follow a Poisson distribution. For generality, however, we model N_{ij} with the negative binomial distribution, of which the Poisson and geometric distributions are special cases. The negative binomial distribution has probability density function $P(N_{ij} = n) = P_n$ where

$$P_n = \begin{cases} \dfrac{\Gamma(n+1/k)}{\Gamma(n+1)\Gamma(1/k)} \dfrac{(k\lambda_{ij})^n}{(1+k\lambda_{ij})^{n+1/k}}, & k \neq 0 \\ \dfrac{\lambda_{ij}^n e^{-\lambda_{ij}}}{n!}, & k = 0 \end{cases} \tag{2}$$

Here $\Gamma(.)$ denotes a gamma function, so that $\Gamma(m) = (m-1)!$ when m is an integer. The mean of the distribution is λ_{ij} and the variance is $\lambda_{ij} + k\lambda_{ij}^2$. The dispersion parameter k is a measure of the relative magnitude of the variance and the mean. The negative binomial distribution reduces to a Poisson distribution when $k = 0$ and to a geometric distribution when $k = 1$. The mean λ_{ij} represents the expected number of fire scars in gridcell i in year j. The model is completed by assuming the following log–linear dependence on early summer rainfall anomalies r_{ij}:

$$\log(\lambda_{ij}) = a_i + b \cdot r_{ij} + c_j \tag{3}$$

Since the rainfall anomalies r_{ij} are known from GPCP data (equation (1)), the analysis involves using the data to find best–fitting values for the parameters a_i, c_j and k (which are dimensionless) and b (which has units of day mm^{-1}).

[11] It is important to stress that the statistical model (equations (2) and (3)) is invariant under transformations of the form

$$a_i \to a_i + \delta, \; c_j \to c_j - \delta \tag{4}$$

where δ is a constant. To remove this degree of freedom we require that $\langle c_j \rangle$ (the average value of the c_j's over all j) be zero. If $\langle c_j \rangle \neq 0$, this constraint can be imposed by performing the transformations in equation (4) with $\delta = \langle c_j \rangle$. The values of a_i and c_j reported here have all been transformed in this way.

[12] A physical interpretation of the parameters can be obtained by rewriting equation (3) in exponential form:

$$\lambda_{ij} = \exp(a_i) \cdot \exp(b \cdot r_{ij}) \cdot \exp(c_j) \tag{5}$$

This shows that the expected number of fire scars λ_{ij} is modelled as a product of three factors: a time–independent spatial factor $\exp(a_i)$, a space–independent temporal factor $\exp(c_j)$ and a spatio–temporal rainfall factor $\exp(b \cdot r_{ij})$. It should be stressed that this separable functional form was

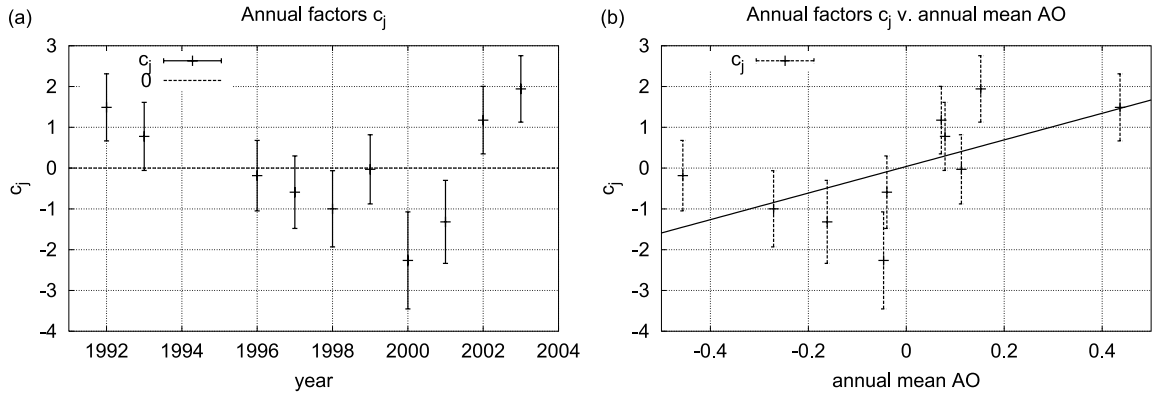

Figure 2. (a) Estimates of the space-independent temporal factors c_j (posterior mean $\pm 2\times$ standard deviation). (b) Comparison of the annual factors c_j with the annual mean value of the arctic oscillation index. A least–squares linear regression line ($c_j = 3.26 \cdot AO + 0.04$, $R^2 = 0.36$, $p = 0.07$) is shown for comparison. (AO data from http://www.cpc.ncep.noaa.gov/).

chosen for simplicity, and allows the three factors to be treated separately. Similarly, the spatio–temporal factor is restricted to rainfall for simplicity, but could be extended to include factors like soil moisture and relative humidity.

[13] The 69 parameters a_i represent the component of spatial variability in the number of fire scars that remains constant from year to year. For example, factors such as forested area, forest type, latitude and proximity to human settlement (amongst other things) are all likely contributors to the value of a_i derived for a particular gridcell i. On the other hand, the 10 parameters c_j quantify any year-to-year variability that applies across all locations in Central Siberia independently of the regional rainfall R_{ij}. (It is possible that the temporal factors c_j apply in a region larger than Central Siberia itself, but the finite spatial range of the data means that we cannot say for sure.) The dispersion parameter k quantifies how closely, on average, the observed number of fire scars n_{ij} matches the expected number λ_{ij}. For smaller values of k the observed number tends to be closer to the expected number and the statistical model makes relatively precise predictions. As k increases, however, the predictions become less precise, although they are still correct 'on average'. Finally, the parameter b quantifies the extent to which spatial variability in the number of fire scars correlates with regional early summer rainfall. Mathematically this can take any value, but on physical grounds, one might expect b to be negative, indicating that a *decrease* in regional rainfall was associated with an *increase* in the number of fire scars. (A lack of correlation between the number of fire scars and regional rainfall anomalies would be indicated by $b = 0$.)

[14] We used the computer code WinBUGS to estimate the parameters a_i, b, c_j and k in the model of equations (2) and (3). WinBUGS uses a Monte Carlo Markov Chain method to estimate Bayesian posterior distributions for the parameters [*Spiegelhalter et al.*, 2003]. This has the advantage of yielding a confidence interval (the posterior standard deviation) as well as a point estimate (the posterior mean) for each parameter.

[15] Prior distributions for Bayesian analysis are designed to be maximally uninformative given the prior information available. For the parameters a_i and b we choose Gaussian priors with zero mean and large variance. (In practice the prior variance is limited by numerical constraints to 10. This is because large parameters in an exponential model can lead to numerical overflow). The parameter k is non-negative and so we use a Jeffreys prior of the form $\pi(k) \sim 1/k$ [*Jaynes and Bretthorst*, 2003].

3. Results

[16] The posterior estimate for the dispersion parameter is $\hat{k} = 1.05 \pm 0.26$ (mean $\pm 2\times$ standard deviation). This range of values is incompatible with a Poisson distribution ($k = 0$) but consistent with a geometric distribution ($k = 1$). This allows a simple interpretation to be placed on the model. When $k = 1$ the probability density function of equation (2) reduces to:

$$P(N_{ij} = n) = \frac{\lambda_{ij}^n}{(1 + \lambda_{ij})^{n+1}} \tag{6}$$

In a series of independent Bernoulli trials – each of which is either a success or a failure – the geometric distribution describes the number of failures before the first success. This can be seen by rewriting equation (6) in the alternative form

$$P(N_{ij} = n) = p_{ij}(1 - p_{ij})^n \tag{7}$$

where $p_{ij} = 1/(1 + \lambda_{ij})$ is the probability of success on each trial. The geometric distribution is therefore consistent with a toy model in which each fire scar has probability p_{ij} of being the final fire scar to be recorded in gridcell i and year j. Thus, fire scars are treated as events which are statistically independent but whose associated probabilities vary over space and time.

[17] Posterior estimates for the spatial factors a_i are shown in Figure 1a with associated uncertainties in Figure 1b. Fire scars are most prevalent between 52°N and 60°N in the region west of Lake Baikal. Since most of the fire scar data come from this area the uncertainties in a_i (Figure 1b) are correspondingly smaller here. The accompanying population map shows that this coincides with an area of relatively high population density (Figure 1c).

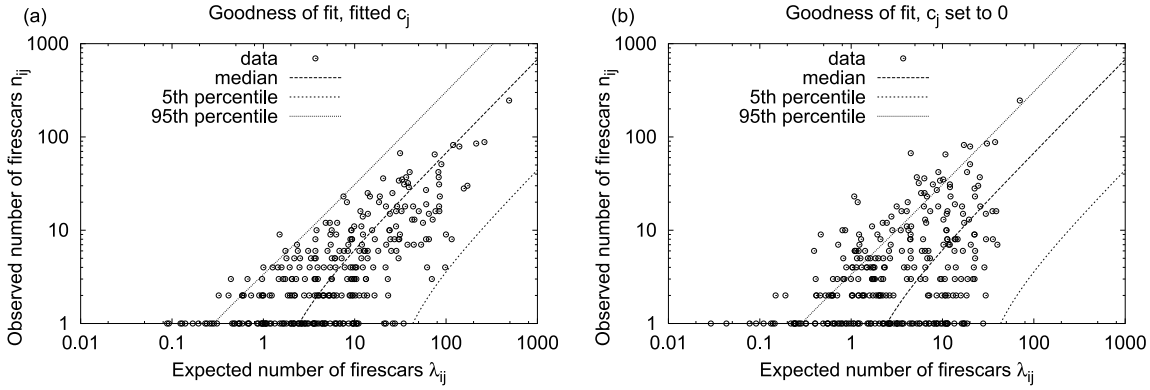

Figure 3. (a) The goodness-of-fit of the full model, with posterior mean values used as parameter estimates for a_i, b, c_j and k. (b) The predictive model, with all c_j set to zero, but other parameters as in Figure 3a. Each figure contains 302 of the 690 data points (This is because 388 data points have $n_{ij} = 0$ and do not appear on logarithmic axes).

Population density is only one of many spatial factors that influence forest fires, but this demonstrates consistency with the idea that many fires are caused by human activity (e.g., accidental spread from camp fires and controlled burning to combat insect outbreaks) [*Korovin*, 1996; *Valendik*, 1996; *Mollicone et al.*, 2006].

[18] Posterior estimates for the annual factors c_j (Figure 2a) illustrate the component of interannual variability that cannot be explained by regional rainfall anomalies r_{ij}. For example, the total expected number of fires in 2003 (when $c_j \approx 2$) is roughly $\exp(2) \approx 7.4$ times greater than in a 'typical' year when $c_j \approx 0$, irrespective of the spatial variations that correlate with regional rainfall anomalies. *Balzter et al.* [2005] suggest that interannual variability in forest fires correlates with the annual mean Arctic Oscillation index (AO). This index quantifies the difference in atmospheric pressure between the northern middle and high latitudes, and serves as a measure of large scale climatic conditions. The correlation between the annual factors c_j and AO is shown in Figure 2b for comparison with this result. There is an imperfect but significant correlation between the two, suggesting that some of the interannual variability in the number of fire scars can be attributed to changes in AO. This is presumably because the large–scale climatic conditions which are conducive to increased biomass availability and reduced water availability are associated with positive values of AO [*Balzter et al.*, 2005].

[19] The posterior estimate for the effect of rain is $\hat{b} = -1.99 \pm 0.56$ day mm^{-1} (mean $\pm 2 \times$ standard deviation). This range of uncertainty does not include zero, providing strong evidence for a relationship between the number of fire scars and rainfall anomalies. Quantitatively, for each additional 0.1 mm day^{-1} of anomalous rainfall, the expected number of fire scars per gridcell per year falls by approximately 18%. Equivalently, the expected number of fire scars doubles (halves) for an anomalous negative (positive) rainfall of $r_{ij} = \log(2)/1.99 = 0.35$ mm day^{-1}.

[20] The posterior mean estimates for a_i, b, c_j and k together yield a statistical model for the number of fire scars. The goodness-of-fit of this model is illustrated in Figure 3a. There is an acceptable fit to the data with almost all of the data points falling between the 5th and 95th percentiles. This model cannot be used predictively, however, since the annual factors c_j are derived from each year's observations. If forecasts of seasonal rainfall R_{ij} are available a predictive version of the model could be constructed by setting all of the temporal factors c_j equal to zero (consistent with the general requirement that $\langle c_j \rangle = 0$). This leads, naturally, to a decline in goodness-of-fit (Figure 3b) but does yield predictive ability.

4. Conclusions

[21] In this paper, we consider the relationship between April–July rainfall and the number of forest fire scars in Central Siberia. A negative binomial model for the number of forest fire scars (equations (2) and (3)) provides a statistically acceptable fit to the data (Figure 3a). The estimated value for the dispersion parameter ($k = 1.05 \pm 0.26$) is consistent with a geometric distribution but not with a Poisson distribution. The estimated rainfall factor $b = -1.99 \pm 0.56$ day mm^{-1} reveals a convincing, quantitative link between the number of fire scars and regional variability in early summer rainfall. The expected number of fire scars halves for each additional 0.35 mm day^{-1} of rainfall in the period April–July. The full model relies on empirical estimates of the interannual variability c_j but can be used in predictive mode by setting $c_j = 0$. This work establishes a quantitative link between regional–scale drought patches and forest fires which could be used to incorporate seasonal rainfall forecasts into predictions of the likely number of forest fires. In addition, it could be used to improve the representation of fire disturbances in global climate models.

[22] **Acknowledgments.** This work was supported by the NERC Climate and Land-Surface Systems Interaction Centre (CLASSIC). We thank Mike Bradshaw and Lisa Barber of the University of Leicester for helpful discussions, and Sergey Venevsky and an anonymous reviewer for valuable comments on an earlier draft. The fire scar mapping was supported by the SIBERIA-2 project, Multi-Sensor Concepts for Greenhouse Gas Accounting in Northern Eurasia, supported by the 5th Framework Programme of the European Commission, contract EVG1-CT-2001-00048. Project coordination: Chris Schmullius, Friedrich-Schiller University Jena.

References

Arora, V. K., and G. J. Boer (2005), Fire as an interactive component of dynamic vegetation models, *J. Geophys. Res.*, *110*, G02008, doi:10.1029/2005JG000042.

Adler, R. F., et al. (2003), The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present), *J. Hydrometeorol.*, *4*, 1147–1167.

Balzter, H., et al. (2005), Impact of the Arctic Oscillation pattern on interannual forest fire variability in Central Siberia, *Geophys. Res. Lett.*, *32*, L14709, doi:10.1029/2005GL022526.

Bertschi, I. T., and D. A. Jaffe (2005), Long-range transport of ozone, carbon monoxide, and aerosols to the NE Pacific troposphere during the summer of 2003: Observations of smoke plumes from Asian boreal fires, *J. Geophys. Res.*, *110*, D05303, doi:10.1029/2004JD005135.

Conard, S. G., et al. (2002), Determining effects of area burned and fire severity on carbon cycling and emissions in Siberia, *Clim. Change*, *55*, 197–211.

Jaffe, D., et al. (2004), Long-range transport of Siberian biomass burning emissions and impact on surface ozone in western North America, *Geophys. Res. Lett.*, *31*, L16106, doi:10.1029/2004GL020093.

Jaynes, E. T., and G. L. Bretthorst (2003), *Probability Theory: The Logic of Science*, Cambridge Univ. Press, New York.

Korovin, G. N. (1996), Analysis of the distribution of forest fires in Russia, in *Fire in Ecosystems of Boreal Eurasia*, edited by J. G. Goldammer and V. V. Furyaev, Springer, New York.

Lenihan, J. M., et al. (1998), Simulating broad-scale fire severity in a Dynamic Global Vegetation Model, *Northwest Sci.*, *72*, 91–103.

Mollicone, D., et al. (2006), Human role in Russian wild fires, *Nature*, *440*, 436–4377.

Soja, A. J., W. R. Cofer, H. H. Shugart, A. I. Sukhinin, P. W. Stackhouse Jr., D. J. McRae, and S. G. Conard (2004), Estimating fire emissions and disparities in boreal Siberia (1998–2002), *J. Geophys. Res.*, *109*, D14S06, doi:10.1029/2004JD004570.

Spiegelhalter, D., et al. (2003), WinBUGS user manual, version 1.4, MRC Biostat. Unit, Cambridge, UK. (Available at http://www.mrc-bsu.cam.ac.uk/bugs)

Sukhinin, A. I., et al. (2004), AVHRR-based mapping of fires in Russia: New products of fire management and carbon cycle studies, *Remote Sens. Environ.*, *93*, 546–564.

Valendik, E. N. (1996), Temporal and spatial distribution of forest fies in Siberia, in *Fire in Ecosystems of Boreal Eurasia*, edited by J. G. Goldammer and V. V. Furyaev, Springer, New York.

Venevsky, S., et al. (2002), Simulating fire regimes in human-dominated ecosystems: Iberian Peninsula case study, *Global Change Biol.*, *8*(10), 984–998.

Zhang, Y.-H., et al. (2003), Monthly burned area and forest fire carbon emission estimates for the Russian Federation from SPOT VGT, *Remote Sens. Environ.*, *87*, 1–15.

H. Balzter, C. T. George, and T. E. Jupp, Climate and Land-Surface Systems Interaction Centre, CEH Monks Wood, Huntingdon, Cambridgeshire, PE28 2LS, UK. (tju@ceh.ac.uk)

C. M. Taylor, Climate and Land-Surface Systems Interaction Centre, CEH Wallingford, Crowmarsh Gifford, Oxfordshire OX10 8BB, UK.