Joint Regional Simulation of Annual Area Burned in Canadian Forest Fires

S. Magnussen*

Canadian Forest Service, Natural Resources Canada, 506 West Burnside Road, Victoria, British Columbia V8Z 1M5, Canada

Abstract: Forest fires have a major impact on Canada's forest carbon balance. National level simulations are used to gauge the potential impact of forest management, climate and other important factors on the carbon balance. In this process a joint simulation of the area burned in 29 Canadian forest fire regions is required. This study presents seven simulation algorithms for this complex task. Simulated areas burned are compared to historic data for 1959-1999 with respect to: 1) the average, 2) the variance, 3) the cumulative distribution function, 4) the first and last quartile, 5) skewness, and 6) kurtosis. Comparisons occur at three levels: regional, supra-regional, and combined regional level. Simulations based on a resampling of historic records (bootstrap) confirmed that the data were too sparse for this technique. Treating regions as independent (zero covariance) leads to a significant underestimation of the interannual variance of area burned at the supra regional and combined regional level. Thus an irregular and patchy interregional covariance structure must be taken into account. Only five regions could be viewed as independent from all other regions. A model-based simulation was challenged by the highly skewed and irregular regional distributions of area burned to which no known multivariate distribution would fit well. At the grouped regional and combined regional levels the best fit to historic data was achieved by first generating correlated probits with a multivariate-t distribution with four degrees of freedom, followed by a plug-in of these probits into empirical quantile interpolation functions. Approaches with time-varying or random covariance structures were also promising.

Keywords: Correlation, stratified bootstrap, multivariate normal, multivariate-t, correlation curves, skewness, kurtosis, forest carbon.

INTRODUCTION

Forest fires contribute significantly to global greenhouse gas emissions [1, 2] and predictions are that we can expect an increase in the frequency and severity of forest fires as the earth's atmosphere gets warmer [3-5]. Models for predicting the effect of forests on regional and national carbon budgets [6, 7] must be able to quantify the impact of forest fires at multiple scales (regional, supra-regional, and combined regional). The prediction of the area burned in a given region over a fixed time period is made difficult by extreme regional interannual variation in size of areas burned in a fire [8-14]. Additional difficulties stem from the high sensitivity of fire propensity to minor changes in various climate drivers [15]. In regional models it is the annual aggregate effect of fire that is most often predicted. The annual area burned in forest fire is a key variable in this context. A good model takes into account any spatial or temporal correlation structure relevant at the level of resolution [16-18]. For a multiregional model the presence of a significant interregional, temporal, or spatial correlation structure must be verified and taken into account in the simulation process. If spatial and temporal associations between regions are ignored the simulated distribution of the area burned in a group of more than two regions could be severely biased in terms of variance and higher moments [19-21]. Even seemingly weak regional correlations can exert a disproportionate influence on the variance of a sum of random variables [22].

Multiregional and multiyear simulation of area burned in forest fires in the boreal forest is a complex challenge when the data suggest a non-trivial correlation structure. The distribution of annual areas burned is often highly skewed with extremes occurring in years with either a warm and dry summer or in years with a late spring followed by a cold and wet summer [23]. It can be argued on physical grounds that in northern climates the fire-regime in one year is independent of the fire-regimes in the immediate past [24]. In Canada the winter snow-pack acts as an annual reset of the fire-danger rating to near zero during the last weeks of winter. Global cycles like El Niño [25] may generate apparent temporal correlations as serious conflagrations of forest fires in the boreal forest tend to co-occur during periods of El Niño [8, 26]. Overall, however, temporal correlations tend to be localized and short-lived [27].

In this study we assess the performance of seven algorithms for the joint simulation of area burned in 29 Canadian forest fire regions. Fire regions are those portions of the intersection of ecoclimatic zones and provinces and territories where fire is actively managed to protect forest resource values or communities. The 29 regions account for about a third of the total forested area burned annually in Canada. All simulations attempt to reproduce the regional and joint re-

*Address correspondence to this author at the Canadian Forest Service, Natural Resources Canada, 506 West Burnside Road, Victoria, British Columbia V8Z 1M5, Canada. Email: steen.magnussen@nrcan.gc.ca

1874-3986/08

gional distribution of the area burned in the 29 forest fire regions between 1959 and 1999. The ability to reproduce the distribution for groups of regions and for all regions combined is a key performance criterion. We demonstrate a serious underestimation of the interannual variance of area burned in groups of regions and all regions combined if regions are treated as independent by virtue of ignoring an irregular and patchy inter-regional correlation structure.

Five of the seven simulation algorithms are variants of the empirical percentile method of [28]. We acknowledge the shortcomings of this approach but failed to identify a more promising approach. No single parametric multivariate distribution or spatio-temporal model [29] was suitable for this simulation problem [20] nor do copula or related techniques for joining marginal distributions into a multivariate joint distribution apply due to the large number of correlated regions which would make the computations infeasible [30-32].

MATERIAL AND METHODS

Data

Estimates of annual forested area burned (*BA*) in fires with a minimum size of 200 ha ($BA_{\geq 200}$) between 1959 and 1999 in 29 of Canada's forest fire regions were used as data for simulation of 'future' regional *BA*-values. Fig. (**1**) indicates the spatial domain of the 29 fire regions. Stocks also reported on the methods and procedures behind these estimates. The reliance on remotely sensed data since 1975 also means that the error structure changed at that time [33-35]. Estimates of *BA* in forest fires consuming less than 200 ha of forest ($BA_{<200}$) were obtained and added to $BA_{\geq 200}$. To obtain $BA_{<200}$ estimates the provincial/territorial records of area burned in seven size categories from the National Forest Database Program were used (Canadian Council of Forest Ministers, Compendium of Canadian Forestry Statistics, http://nfdp.ccfm.org as accessed on Jan. 22nd 2007). Empirical distribution functions of area burned by size class were created to obtain an estimate of $BA_{<200}$ and to establish the relationship between $BA_{<200}$ and $BA_{\geq 200}$ *via* nonlinear regressions. Provincial/territorial-level regression models were then used to predict regional $BA_{<200}$ values for each year between 1959 and 1999. Specifically:

$$\widehat{BA}_{<200} = \text{Max}\left\{0, Exp\left[\hat{a} + \hat{b} \times Log\left(BA_{\geq 200} + 1\right) + 0.5 \times \widehat{Var}\left(resid.\right)\right] - 1\right\} \quad (1)$$

Regional sums $\widehat{BA} = \widehat{BA}_{<200} + \widehat{BA}_{\geq 200}$ are used as data; in most regions the $BA_{<200}$ estimates account for less than 1% of the annual area burned.

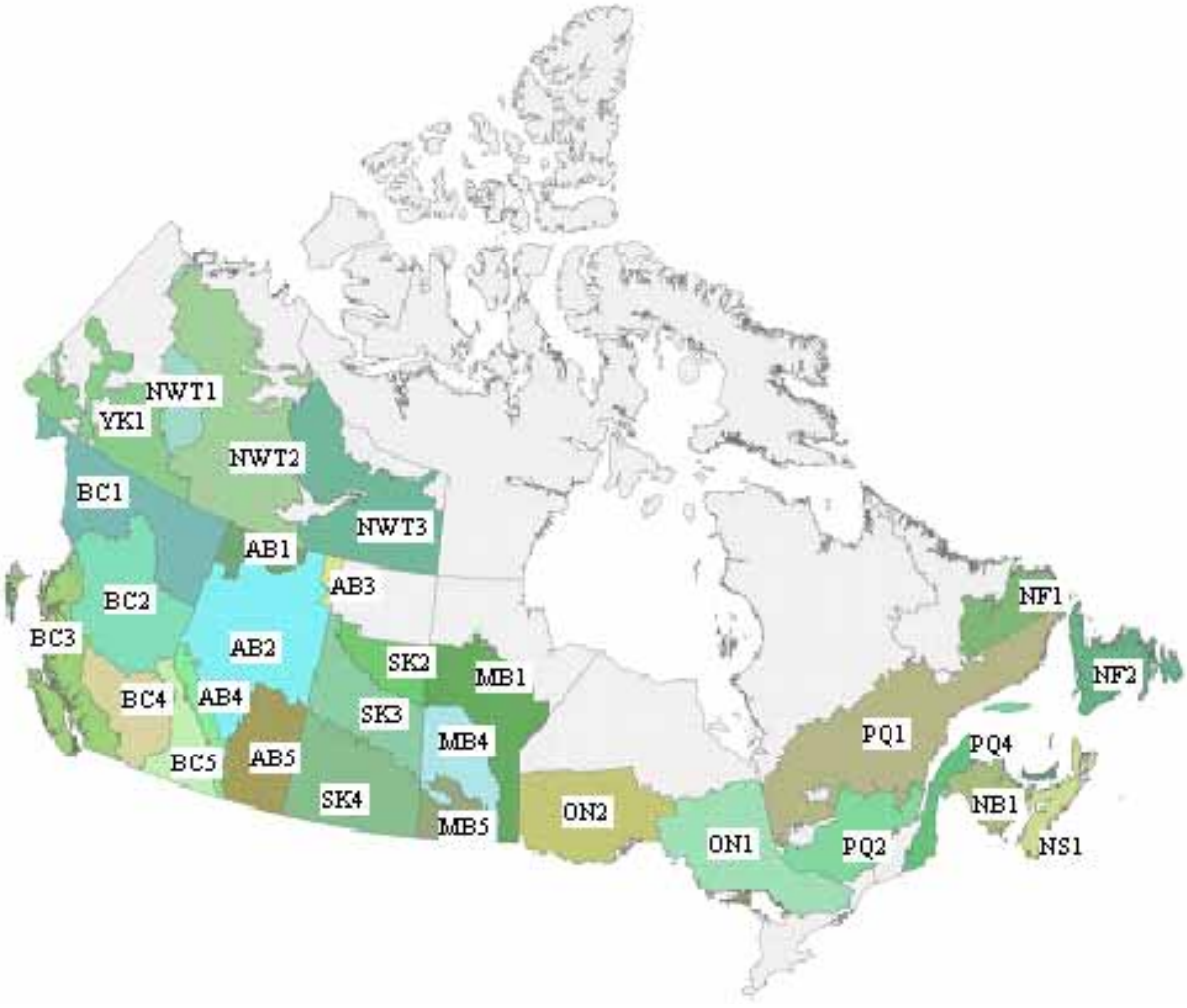

Fig. (1). Map of 29 forest fire regions in Canada.

Simulation Objectives

The objective of the simulation is to generate a joint sequence of annual *BA*-values for 29 Canadian forest fire regions. For each region the distribution of simulated values of *BA* should, in principle, match the empirical distributions for 1959-1999. Simulations should strive to conserve both interregional and temporal correlation where they exist in the data. A conservation of correlation patterns assures that the distribution of simulated sums of regional *BA*-values also matches the empirical distribution. We apply this requirement to four grouped regions, each composed of four to ten individual fire regions. The grouped regions are: Maritime Canada (MC = {NB1,NF1,NF2,NS1}), Eastern Canada (EC = {PQ1,PQ2,PQ4,ON1,ON2,MB1,MB4,MB5}), Central Canada (CC = {SK2-SK4,AB1-AB5,NWT1-NWT3}), and Western Canada (WC = {BC1-BC5,YK1}). The requirement was also extended to the total of annual burned area formed by summing 29 regional *BA*-values.

We propose seven simulation algorithms and evaluate their ability to reproduce six key statistics (mean, variance, cumulative distribution function, first and last quartile, skewness, and kurtosis) of the historic data at the regional, the four supra regions (MC, EC, CC, WC), and all regions combined [36]. Using each of the seven algorithms, we generated 200 replicates of 41 years of simulated regional *BA*-values and compared the six statistics of each replicate to the statistics of the historic data and performed a statistical test of the null hypothesis of equality.

Temporal and Regional Correlations

Regional *BA*-values for any two distinct years from 1959 to 1999 were moderately and positively correlated (mean = 0.25, median = 0.18) with about a third of the correlation coefficients statistically significant at the 5% level. Fig. (**2**) shows a histogram of the relative frequencies of correlation coefficients binned to 0.1-wide classes. The correlation was largely driven by a co-occurrence of relatively large *BA*-values in one, two, or sometimes three regions. However, within regions there was generally a lack of significant time trends and temporal autocorrelation of *BA*-values [37] for years $t+\Delta$ and $t, t=1,...,41-\Delta, \Delta=1,\ldots,8$. A significant time trend (cubic polynomial) was identified in just one region (ON1) where an analysis of variance suggested (P = 0.04) that the area burned during the last third of the observed 41 years has been significantly larger than in first two blocks of 13(14) years. Yet, correlations involving *BA*-values from ON1 did not seem to be confounded by this time trend. As a consequence no temporal detrending of the data was attempted. In terms of year-to-year autocorrelations, only three lag one ($\Delta=1$) correlations for AB1, AB2, and NWT2 reached significance at the 5% level. The three autocorrelations varied from 0.42 to 0.5. Higher order $(\Delta>1)$ autocorrelations were all close to zero and none reached a 20% level of significance. Consequently we decided to ignore a possible significant time trend in the ON1 data and we shall henceforth treat the regional data as a series of 41 independent observations.

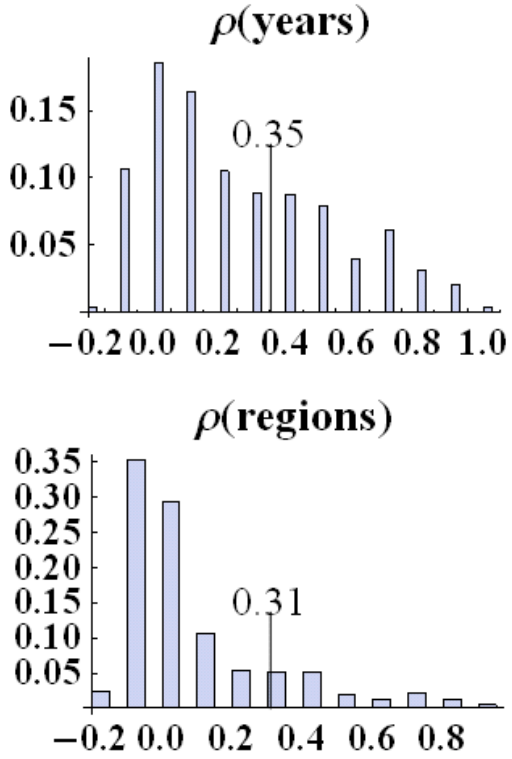

Fig. (2). Distribution of year-to-year and region-to-region correlation coefficients (ρ, rounded to nearest 0.1) of annual forest area burned.

Regional Pearson's product moment correlation coefficients (ρ_{OBS}) of 1959-1999 *BA*-values were, for the most part, close to zero (mean = 0.06) with 56% slightly below zero (Fig. **2**, bottom). However, a total of 61 coefficients (15%) were statistically significant at the 5% level $(\hat{\rho}>0.308)$ and the average of 0.50 for significant coefficients suggests they should be recognized in a simulation procedure. Under the null hypothesis of zero-valued interregional correlation coefficients tested at the 5% level of significance we would expect 5% of the correlations to be significantly larger/smaller than zero. The 29×29 matrix of empirical estimates of correlation coefficients $(\hat{\mathbf{R}})$ is listed in Table **1**.

Empirical Distribution and Quantile Functions of BA

Joint stochastic simulation of regional *BA*-values requires a model for the joint and marginal distributions of *BA* [20]. A parametric approach requires models that describe the marginal distributions well. Yet a series of preliminary analyses quickly made it clear that the historic regional distributions of *BA*-values could not be captured with any desired accuracy by any of the common distribution functions for positively valued random variables. We tried the following distributions: Gamma, Cauchy, Weibull, Logistic, Inverse Gaussian, Generalized Lambda Distribution, Logistic, and Beta with the latter two applied to *BA*-values divided by the regional forest area supporting a combustible biomass in excess of one metric tonne per ha. A truncated exponential

Table 1 Pearson's Product Moment Correlation Coefficients of Regional Annual Area Burned (*BA*). All Entries have been Multiplied by 100. Coefficients Larger than 0.31 (Displayed as 31) are Significant at the 5% Level

	AB1	AB2	AB3	AB4	AB5	BC1	BC2	BC3	BC4	BC5	MB1	MB4	MB5	NB1
AB1	100	68	2	-5	-9	50	11	5	-7	0	12	-7	-1	-1
AB2	68	100	32	9	-5	19	8	7	-4	-7	8	-3	15	-4
AB3	2	32	100	1	-7	-5	-6	-10	-3	-7	3	6	50	-2
AB4	-5	9	1	100	35	-5	4	4	24	13	22	22	1	-12
AB5	-9	-5	-7	35	100	-13	-6	2	4	-5	-5	-11	-10	-7
BC1	50	19	-5	-5	-13	100	44	34	25	50	-11	8	3	0
BC2	11	8	-6	4	-6	44	100	72	87	43	1	49	36	-9
BC3	5	7	-10	4	2	34	72	100	68	39	-5	32	13	1
BC4	-7	-4	-3	24	4	25	87	68	100	43	-3	43	29	-10
BC5	0	-7	-7	13	-5	50	43	39	43	100	-10	3	-4	-6
MB1	12	8	3	22	-5	-11	1	-5	-3	-10	100	75	8	8
MB4	-7	-3	6	22	-11	8	49	32	43	3	75	100	39	-7
MB5	-1	15	50	1	-10	3	36	13	29	-4	8	39	100	-6
NB1	-1	-4	-2	-12	-7	0	-9	1	-10	-6	-8	-7	-6	100
NF1	-13	-8	-7	-9	-8	-1	-9	-3	-10	4	-12	-10	-13	-5
NF2	-5	-3	-1	3	-8	21	87	66	82	16	2	56	41	19
NS1	-9	-7	-2	35	-9	-4	29	20	38	26	26	33	6	-6
NWT1	39	7	-6	1	-3	45	-4	-9	-5	-8	-5	-11	-14	2
NWT2	46	31	34	-12	-4	-8	-8	-11	-13	-14	29	-4	2	-3
NWT3	13	15	43	-3	-11	2	0	-4	6	18	2	-1	-4	-11
ON1	17	5	-4	10	12	-17	-13	-19	-15	-15	17	-11	-2	-7
ON2	1	15	45	0	-13	8	53	22	40	-3	10	39	65	2
PQ1	-6	4	-13	6	-6	-8	-3	-21	-16	-3	-9	-11	-13	23
PQ2	-11	-13	-10	-16	-4	-6	-1	-12	-10	-15	-11	-3	-7	-6
PQ4	21	-1	-8	-4	-2	-1	4	-8	7	-8	17	-7	-11	-8
SK2	66	77	11	-6	5	-3	2	4	-9	-12	39	5	-2	-10
SK3	49	43	26	-10	-3	-7	0	-7	-8	-13	34	0	19	-7
SK4	-3	-13	-8	-6	-4	-12	-4	-8	-3	-7	-5	1	6	-4
YK1	23	12	11	10	-6	33	-1	-10	0	-2	39	29	0	0

	NF1	NF2	NS1	NWT1	NWT2	NWT3	ON1	ON2	PQ1	PQ2	PQ4	SK2	SK3	SK4	YK1
AB1	-13	-5	-9	39	46	13	17	1	-6	-11	21	66	49	-3	23
AB2	-8	-3	-7	7	31	15	5	15	4	-13	-1	77	43	-13	12
AB3	-7	-1	-2	-6	34	43	-4	45	-13	-10	-8	11	26	-8	11
AB4	-9	3	35	1	-12	-3	10	0	6	-16	-4	-6	-10	-6	10
AB5	-8	-8	-9	-3	-4	-11	12	-13	-6	-4	-2	5	-3	-4	-6
BC1	-1	21	-4	45	-8	2	-17	8	-8	-6	-1	-3	-7	-12	33
BC2	-9	87	29	-4	-8	0	-13	53	-3	-1	4	2	0	-4	-1
BC3	-3	66	20	-9	-11	-4	-19	22	-21	-12	-8	4	-7	-8	-10
BC4	-10	82	38	-5	-13	6	-15	40	-16	-10	7	-9	-8	-3	0
BC5	4	16	26	-8	-14	18	-15	-3	-3	-15	-8	-12	-13	-7	-2
MB1	-12	2	26	-5	29	2	17	10	-9	-11	17	39	34	-5	39
MB4	-10	56	33	-11	-4	-1	-11	39	-11	-3	-7	5	0	1	29
MB5	-13	41	6	-14	2	-4	-2	65	-13	-7	-11	-2	19	6	0
NB1	-5	19	-6	2	-3	-11	-7	2	23	-6	-8	-10	-7	-4	0
NF1	100	-4	-9	-9	0	-1	9	-4	0	-8	-9	-12	-12	-12	-12
NF2	-4	100	22	-8	-5	3	-10	56	-4	-6	-2	-8	-6	-1	-3
NS1	-9	22	100	-8	-12	-10	-5	11	-13	-10	1	0	-6	-6	6
NWT1	-9	-8	-8	100	21	30	-15	-7	-11	7	-9	-2	-11	-11	72
NWT2	0	-5	-12	21	100	44	52	17	-2	-16	49	61	79	-19	30
NWT3	-1	3	-10	30	44	100	-10	1	-5	-15	-16	10	-1	-11	24
ON1	9	-10	-5	-15	52	-10	100	6	4	-5	47	31	60	-6	5
ON2	-4	56	11	-7	17	1	6	100	13	9	-1	3	24	-11	4
PQ1	0	-4	-13	-11	-2	-5	4	13	100	18	5	-9	-2	11	-11
PQ2	-8	-6	-10	7	-16	-15	-5	9	18	100	-8	-16	-15	-3	-10
PQ4	-9	-2	1	-9	49	-16	47	-1	5	-8	100	29	56	-7	-3
SK2	-12	-8	0	-2	61	10	31	3	-9	-16	29	100	76	-10	15
SK3	-12	-6	-6	-11	79	-1	60	24	-2	-15	56	76	100	-10	10
SK4	-12	-1	-6	-11	-19	-11	-6	-11	11	-3	-7	-10	-10	100	-11
YK1	-12	-3	6	72	30	24	5	4	-11	-10	-3	15	10	-11	100

distribution of log-transformed *BA*-values - as proposed by Cumming [14] - was included for completeness. A Gamma distribution was the most promising. Data from nine regions could be reasonably described by this model (Kolmogorov-Smirnov test statistics of fit were non-significant, $P > 0.10$) but for 15 regions a Gamma distribution would have been a poor choice ($P \leq 0.01$).

A suitable distribution function could not be found in Johnson's primer on multivariate statistical simulation [20]. Fleishman's method [38], for generating multivariate data with pre-specified marginal means, variances, skewness, and kurtosis was also tried. It has been used successfully by Vale and Maurelli [39] but with our data a valid transformation rule could only be found for five of the 29 regions.

Attempts to improve these results by a data-transformation were unsuccessful. The following transformations were tried: logarithmic, square-root, inverse hyperbolic sine, Box-Cox power transforms [40], and $pBA \times \left(1 - pBA\right)^{-1}$ where *pBA* is the proportion of exposed forest burned. Our conclusion from these preliminary analyses was that empirical regional distribution functions would accommodate the simulation objectives better than any attempt to fit the data to a parametric model.

Empirical distribution functions $\left(F\right)$were linear interpolation functions. For region *i* we have

$$\hat{F}_i\left(BA_{i,r(j)}\right) = \frac{r(j)}{41+1}, r(j) = 1,..,41 \quad (2)$$

where $r(j)$ is the rank (smallest = 1, largest =41) of the regional *BA*-value in year *j* [41]. To cap $\hat{F}_i$ we fixed the extremes of regional *BA*-values at $0.88 \times \underset{j}{\text{Min}}\left(BA_{ij}\right)$, and $1.08 \times \underset{j}{\text{Max}}\left(BA_{ij}\right)$, respectively. At these extremes the distribution function takes values of zero and one, respectively.

The scaling factors of 0.88 and 1.08 for capping the extremes were gained from a simulation study on the effect of sample size (number of years) on the extremes of a gamma-distributed random variable. We took 50 to be the target number of years in the simulations and then estimated the effect of increasing the sample size from 41 to 50 years on the extremes of a gamma distributed random variable. Parameters in the gamma distribution were regional maximum likelihood estimates. The above scaling factors are the estimated means taken across all regions and 800 replicate samples of size 41 and 50. Sarhan and Greenberg [42] investigated the effect of sample size on the expected values of extremes of random variables generated from a symmetric distribution. According to their results, one should expect the range of a random variable to increase by approximately 10% when the sample size is increased from 41 to 50. For a right-skewed distribution an expected increase in the range due to an increase in the length of the observational period is likely to inflate the mean. In other words the expected mean annual area burned in a finite sequence of years depends on the length of the sequence.

The inverse of *F* is called a quantile (viz. percentile) function and is used to generate a random *BA*-value $\left(BA^*\right)$ from a random draw of a variable (probit, *u*) uniformly distributed on the unit interval $\left(BA^* = F^{-1}\left(u\right), 0 \leq u \leq 1\right)$. For a given *u* the empirical quantile function yields the $u \times 100$ percentile of the regional distribution of *BA*. For the observed data we have

$$\hat{F}_i^{-1}\left(\frac{r}{42}\right) = BA_{i,r(j)}, r = 1,...,41 \quad (3)$$

Examples of the empirical quantile functions are displayed in Fig. (**3**) alongside maximum likelihood estimates of the gamma distribution functions.

Simulation Procedures

Bootstrap (BOOT)

A simple way of achieving the simulation objective(s) would be to generate bootstrap samples of the data [43]. A single bootstrap sample would then consist of 41 *BA*-values for each region representing a randomly selected (with replacement) sequence of years (1959-1999). The sequence of years would be the same for all regions in a single bootstrap sample. By repeatedly drawing a new sequence of 41 years and their associated *BA*-values one obtains a bootstrap sample of the joint regional distribution of *BA*-values. However, a bootstrap sample generated from just 41 records of extremely variable numbers will tend to exaggerate the temporal variability considerably [44]. A large number of bootstrap samples will appear as unrealistic when, by chance, years with either a high or a low *BA*-value are drawn more than once in a sequence. Putting a restriction on the number of times that a *BA*-value for a given year can appear in a sample is an attempt to rectify this problem. We therefore propose a stratified bootstrapping procedure whereby the regional *BA*-values are stratified by years of observation. Eight strata of *BA*-values corresponding to years 1-5, 6-10, ..., 30-35, and 36-41 were formed. A bootstrap sample was then generated by first drawing five years at random (with replacement) from each of the first seven strata and six years from the last strata, and then collecting the corresponding regional *BA*-values. Thus a *BA*-value for a specific year between 1959 and 1992 can appear at most five times in a bootstrap sample while a *BA* -value from a year between 1993 and 1999 can appear at most six times.

Call the vector of years drawn for a single bootstrap sample $\mathbf{Y}_{41}^* = (\mathbf{Y}_{1:5}^*, \mathbf{Y}_{6:10}^*, ..., \mathbf{Y}_{35:41}^*)$. A bootstrap sample of 41 years of *BA*-values in 29 fire regions forms a 29 × 41 matrix

$$\mathbf{BA}_{BOOT}^* = \{\mathbf{BA}_1(\mathbf{Y}_{41}^*), ..., \mathbf{BA}_{29}(\mathbf{Y}_{41}^*)\} \quad (4)$$

where $\mathbf{BA}_i\left(\mathbf{Y}_{41}^*\right)$ is the *i*th regional vector of 41 *BA*-values specific to the stratified bootstrapped sequence of years. Bootstrap results applicable to MC, EC, CC, WC and all regions combined (Canada) were obtained by a simple summation of appropriate bootstrap records.

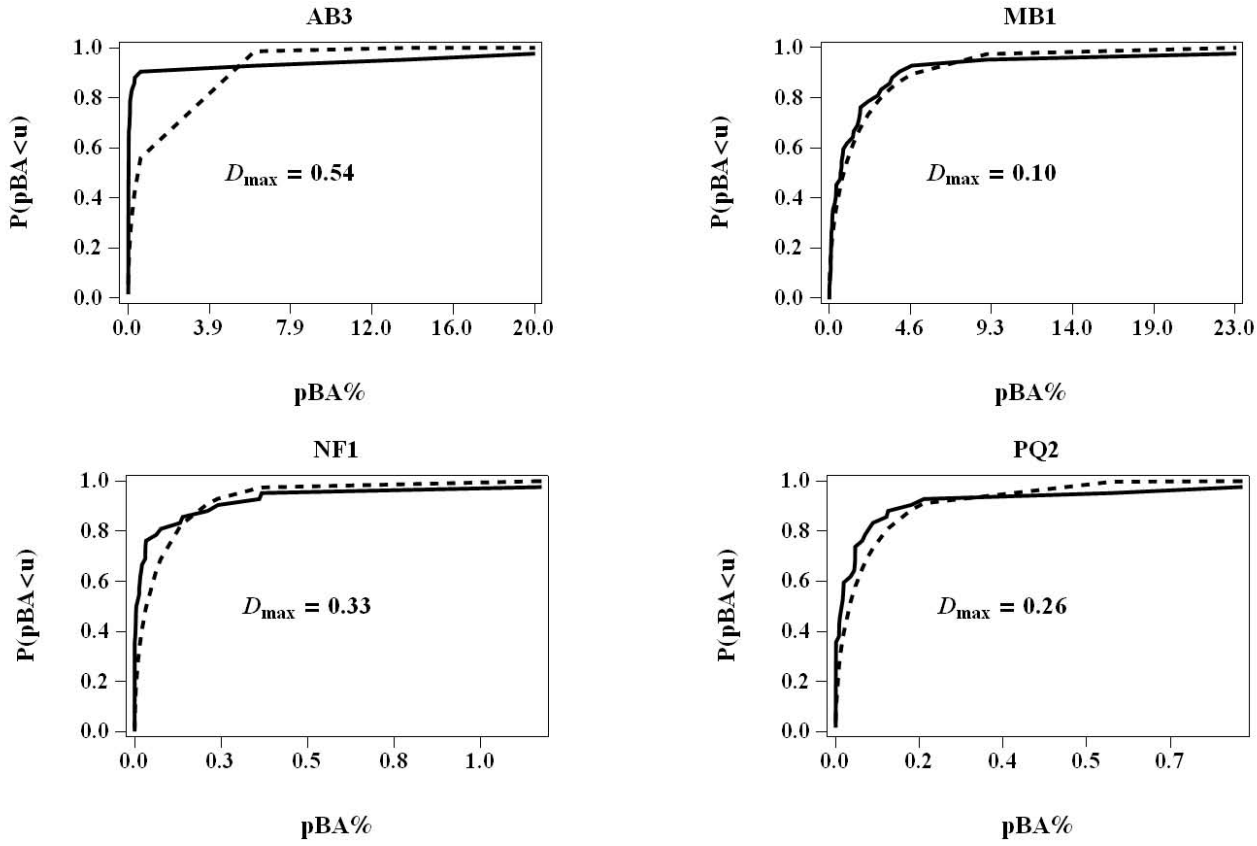

Fig. (3). Four (random) examples of empirical quantile functions (full lines) and the maximum likelihood estimate of the gamma distribution function (dashed line). Function values for u = $j\times\left(41+1\right)^{-1}$, $j=1,...,41$ have been connected by straight lines. $D_{\max}$ is the Kolmogorov-Smirnov test statistics of maximum absolute difference in distribution functions (evaluated at *u*).

Multivariate Normal (MVN)

The multivariate normal simulation approach has been advocated by Bradley [28] and Johnson [20]. A random vector of 29 correlated regional annual *BA*-values $\left(BA_{it}^{*}, i=1,...,29\right)$ was obtained by the following algorithm

$$\widehat{\mathbf{BA}}_{MVN}^{*}=\left(\widehat{BA}_{1}^{*},\ldots,\widehat{BA}_{29}^{*}\right)=\hat{\mathbf{F}}^{-1}\left(\Phi\left(\mathbf{z}^{*}\right)\right),\ \mathbf{z}^{*}\sim\text{MVN}\left(\mathbf{0},\hat{\mathbf{R}}\right) \quad (5)$$

where $\hat{\mathbf{F}}^{-1}$ is a vector of the 29 regional empirical quantile functions defined in (3), Φ is the multivariate normal distribution function (MVN), $\mathbf{z}^{*}$ is a random draw from a standard (mean **0** and variance **1**) multivariate normal distribution of dimension 29, and $\hat{\mathbf{R}}$ is the 29×29 matrix of Pearson's product moment correlation coefficients of regional *BA*-values $\left(\hat{\rho}_{OBS}\left(BA_{ij},BA_{i'j}\right), j=1,....,41,\ i,i'=1,...,29\right)$. We have $\mathbf{u}^{*}=\Phi\left(\mathbf{z}^{*}\right)$ where $\mathbf{u}^{*}$ is a vector of regional correlated probits $\left(u_1,...,u_{29}\right)$ on the unit interval. The predominantly right-skewed distributions of *BA* predispose an attenuation of the correlation coefficients in the simulated data.

Multivariate-t (MVT)

To mitigate the anticipated attenuation problem we propose a multivariate-*t* distribution function (MVT) **T** for generating a vector of correlated probits $\left(\mathbf{u}^{*}\right)$

$$\widehat{\mathbf{BA}}_{MVT}^{*}=\left(\widehat{BA}_{T1}^{*},\ldots,\widehat{BA}_{T29}^{*}\right)=\hat{\mathbf{F}}^{-1}\left(\mathbf{T}\left(\mathbf{t}_{\hat{\nu}}^{*}\right)\right),\ \mathbf{t}_{\hat{\nu}}^{*}\sim\text{MVT}\left(\hat{\nu},\hat{\mathbf{R}}\right) \quad (6)$$

where $\hat{\nu}$ is the degree of freedom for **T** (to be estimated). We chose $\hat{\nu}$ to minimize the sum of squared differences between the empirical and simulated correlation coefficients. Thus,

$$\hat{\nu}=\underset{\nu}{\text{argmin}}\left(\sum_{i=1}^{29}\sum_{i'\neq i}\left|\hat{\rho}_{obs}\left(BA_{ij},BA_{i'j}\right)-\hat{\rho}_{MVT}\left(\nu\right)\left(\widehat{BA}_{ij}^{*},\widehat{BA}_{i'j}^{*}\right)\right|^{2}\right) \quad (7)$$

A MVT $\left(4,\hat{\mathbf{R}}\right)$ achieved the best results, i.e. $\hat{\nu}=4$.

Fig. (**4**) illustrates that the attenuations of the correlation coefficients in $\hat{\mathbf{R}}$ are less in MVT than in MVN simulations; at least for correlation coefficients larger than 0.2. Yet a MVT creates a negative bias for weaker correlation coefficients. In an attempt to strike a balance between the two opposite attenuation problems a simulation with a mixture of draws from a MVN and a MVT distribution was explored.

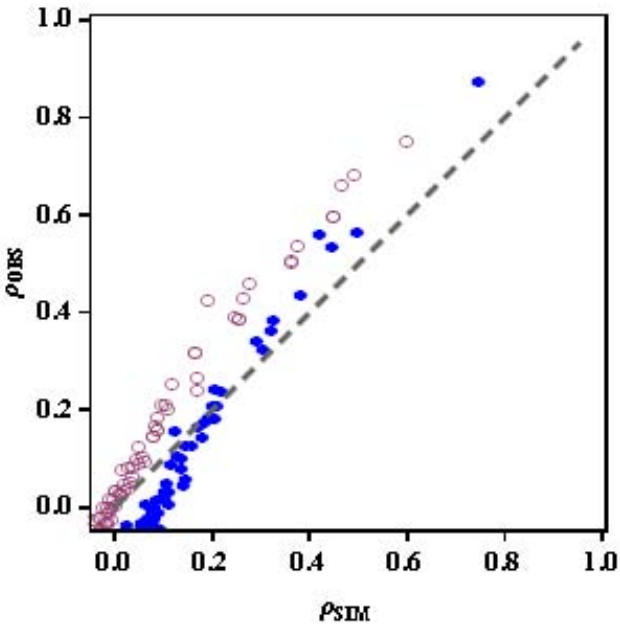

Fig. (4). Regional correlation coefficients of annual area burned in forest fires. A random sample of 100 observed correlation coefficients ρ_{OBS} are plotted against the correlation coefficient in simulated values of area burned $\left(\rho_{SIM}\right)$. Full circles: MVT simulations. Open circles: MVN simulations. A one to one line has been added for reference..

A Mixture of MVN and MVT Simulations (MIX)

In this algorithm a proportion $\left(\lambda\right)$ of the simulations were done according to (4) and the remainder $\left(1-\lambda\right)$ according to (5). We denote a simulation from this mixture distribution as $\widehat{BA}^{*}_{MIX}$. The best mixing ratio $\left(\hat{\lambda}\right)$ was estimated at 0.37 by minimizing a least-squares criterion like the one in (6). In every 100 MIX simulations 37 will come from a MVN distribution and 63 from a MVT distribution.

MVN with a Stochastic Correlation Matrix (RND)

Regional correlation coefficients of *BA* have so far been interpreted as if they were deterministic. In a simulation context the question is whether observed (historic) correlations apply. Significant interregional correlation coefficients were, in most cases, determined by a co-occurrence of years with relatively large *BA* values. A censoring of these events would - in general - sharply reduce the magnitude of historic correlations. In other words, the observed significant correlation coefficients depend critically on a few years with concomitant relatively large *BA*-values. Fig. (**5**) purveys an impression of this dependency by showing the wide distribution of correlation coefficients based on 10 randomly selected years (without replacement). More than 40% of the random 10-year correlation coefficients deviate by more than 0.05 from their historic 41-year value (Fig. **6**). Our results suggest that MVN (or MVT) simulations with a stochastic correlation-matrix instead of $\hat{\mathbf{R}}$ may make the results less dependent on the historic correlation structure embodied in $\hat{\mathbf{R}}$. We propose two procedures (RND and TS) to lessen this dependency.

In RND a random subset of 10 years $\left(\mathbf{J}^{*}_{RND}\right)$ is selected (without replacement) from the 41 years of observation. For each vector $\mathbf{J}^{*}_{RND}$ a 29 × 29 matrix $\hat{\mathbf{R}}^{*}_{RND}$ of product moment correlation coefficients was computed in the obvious way from the randomly selected subset of 10 years of observations $\left(BA_{ij'},\ j' \in \mathbf{J}^{*}_{RND}\right)$. A random vector of BA-values was then simulated as in MVN but with $\hat{\mathbf{R}}^{*}_{RND}$ replacing $\hat{\mathbf{R}}$

$$\widehat{\mathbf{BA}}^{*}_{RND} = \left(\widehat{BA}^{*}_{RND1},\ldots,\widehat{BA}^{*}_{RND29}\right) = \hat{\mathbf{F}}^{-1}\left(\Phi\left(\mathbf{z}^{*}_{RND}\right)\right),\ \mathbf{z}^{*}_{RND} \sim \mathrm{MVN}\left(\mathbf{0},\hat{\mathbf{R}}^{*}_{RND}\right) \qquad (8)$$

MVN with Time-Varying Correlation Coefficients (TS)

In TS the correlation coefficients of regional *BA*-values are functions of time [45]. Let $\rho_{ii'}\left(j\right)$ denote a correlation coefficient of linear association between *BA* -values in regions i and i' at time j (j = 1, ..., 41). For each year of observation (j) an in-time localized estimate of the expected regional correlation is obtained by methods outlined in Doksum [45]. We used a linear moving-average filter with a 10-year bandwidth to obtain smoothed time-trends of regional *BA*-values, and estimates of the standard deviation and temporal covariance structure of the smoothed *BA*-values. The smoothed *BA*-values were considered as a continuous periodic correlation curve in time with a return period of 41 years. Thus every correlation coefficient $\rho_{ii'}\left(j\right)$ is estimated as the weighted average in a symmetric 10-year neighbourhood extending 5 years into the past and 5 years into the future. Our choice of 10 years for the moving average filter was grounded on consideration of accuracy. A shorter bandwidth gave unacceptable standard errors of the estimated local correlation coefficients while a larger bandwidth apparently over-smoothed jumps and spikes in the correlation curve. Estimated correlation-curves are displayed in Fig. (**7**) for the same six examples used in Fig. (**6**). Estimated confidence bands of $\pm 1.\times \widehat{\mathrm{se}}\left(\hat{\rho}_{ii'}\left(j\right)\right)$ conveys an impression of the precision of the localized correlation coefficient. Spikes and valleys interspersed by periods of near zero correlation coefficients are characteristic features of the correlation curves. The average of a correlation curve was, as a rule, 20%-70% lower for region pairs with an observed correlation in excess of 0.4. As expected, for weakly correlated regions the average of a correlation curve matched closely $\hat{\rho}_{OBS}$. To simulate the joint distribution of regional *BA*-values with the *TS* procedure a random vector of regional *BA*-values was generated from a randomly selected year (j^{*} = 1,..., 41).

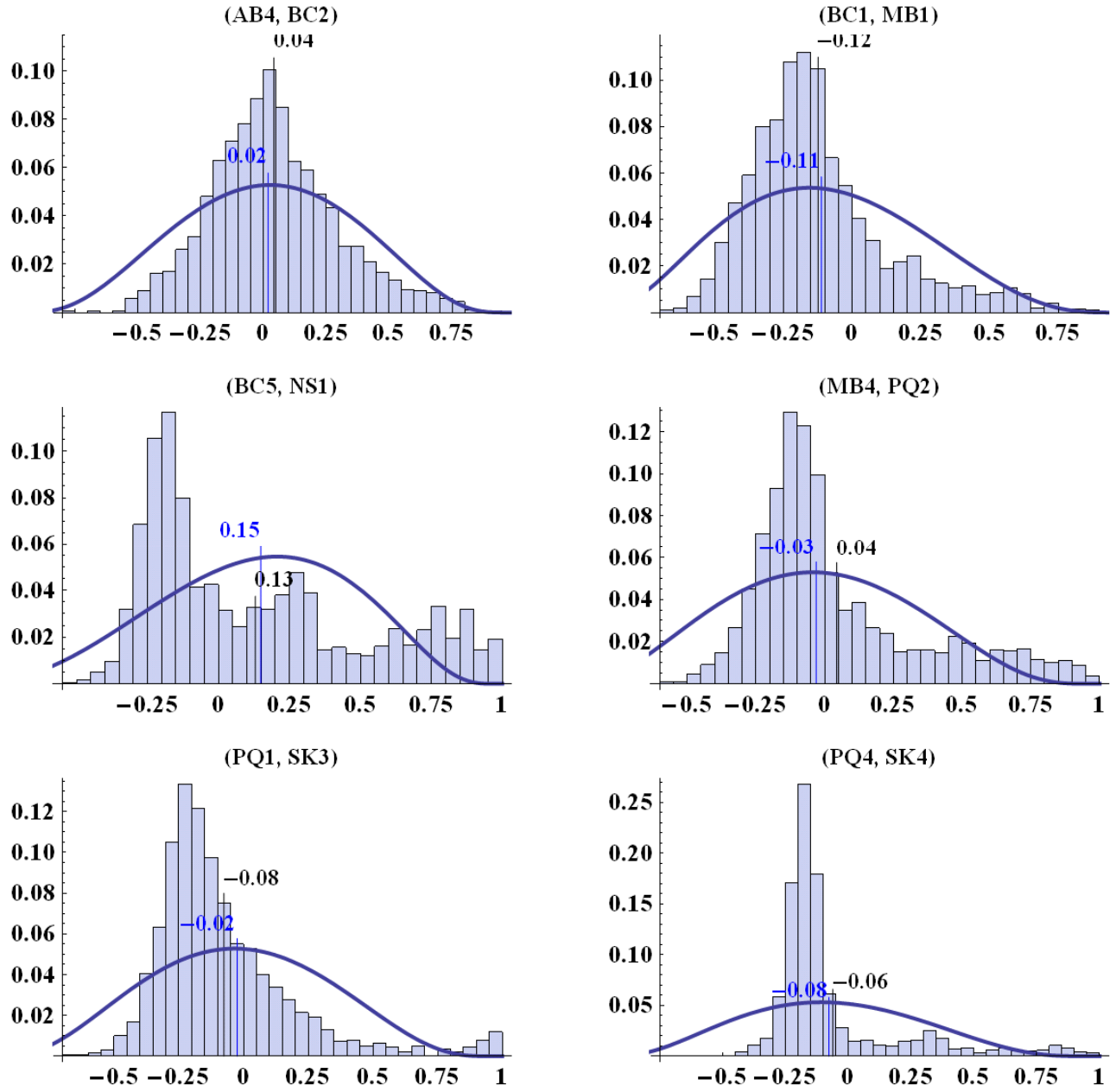

Fig. (5). Six (random) examples of the distribution of regional correlation coefficients (rounded to nearest 0.05) based on 10 randomly selected years (1959-1999). The expected distribution based on the assumption of an underlying linear relationship with constant variance is shown by the full line. The means of the two distributions are indicated by vertical lines.

$$\widehat{\mathbf{BA}}^{*}_{TS} = \left(\widehat{BA}^{*}_{TS1}, \ldots, \widehat{BA}^{*}_{TS29} \right) = \hat{\mathbf{F}}^{-1}\left(\Phi\left(\mathbf{z}^{*}_{TS} \right) \right),\ \mathbf{z}^{*}_{TS} \sim \text{MVN}\left(\mathbf{0}, \hat{\mathbf{R}}_{TS}\left(j^{*} \right) \right) \quad (9)$$

where $\hat{\mathbf{R}}_{TS}\left(j^{*} \right)$ is a 29×29 matrix of estimates $\hat{\rho}_{ii'}\left(j^{*} \right)$.

Regional Independence (NULL)

A region-by-region simulation of *BA*-values *via* the empirical regional quantile (percentile) functions ignores any correlation structure in the data by assuming that the *BA*-values of a region are independent of the BA-values in all other regions. Simulations with the NULL model serve as a benchmark for assessing the impact of interregional correlations on the distribution of sums of *BA*-values for groups of regions (e.g. MC, EC, CC, and WC) or the total of all regions. To facilitate a direct comparison of NULL and MVN results, the generating algorithm for the NULL procedure was identical to that of MVN in (4) except for the replacement of the correlation matrix $\hat{\mathbf{R}}$ with the identity matrix (**I**)

$$\widehat{\mathbf{BA}}^{*}_{NULL} = \left(\widehat{BA}^{*}_{1}, \ldots, \widehat{BA}^{*}_{29} \right) = \hat{\mathbf{F}}^{-1}\left(\Phi\left(\mathbf{z}^{*}_{0} \right) \right),\ \mathbf{z}^{*}_{0} \sim \text{MVN}\left(\mathbf{0}, \mathbf{I} \right) \quad (10)$$

Assessment of Performance

Simulated data will not mirror historical data. We require, however, that important statistics pertaining to the marginal and joint distribution of simulated regional annual

area burned in forest fires match the same statistics in the historic data. To this end, six statistics were computed for each simulation period of 41 years and compared to the corresponding statistics for the historic data [46]. A test of the null hypothesis of equality was performed for each statistic. The six statistics were: 1) the mean, 2) the variance, 3) the cumulative distribution function (CDF), 4) the first and last quartile (25 and 75 percentiles), 5) the skewness, and 6) the kurtosis. Comparisons were made at the regional, grouped-regional, and combined regional levels.

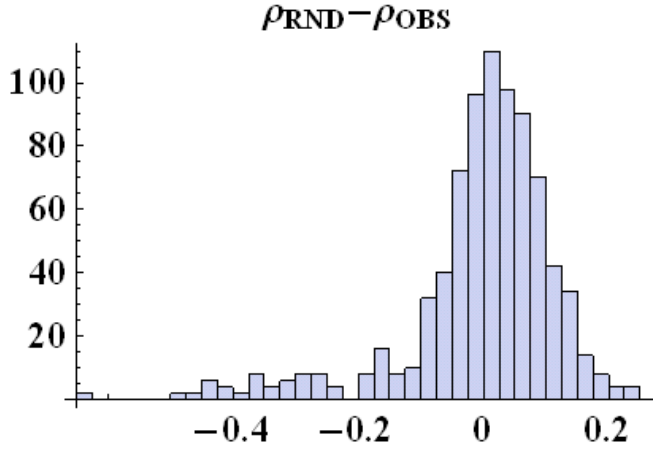

Fig. (6). Frequency distribution of differences between regional correlation coefficients (rounded to nearest 0.05) of *BA* computed from *BA* of 10 randomly selected years (1959-1999) and the correlation coefficients based on observed 1959-1999 data.

Means were compared with a conventional *t*-test statistic with 39 degrees of freedom. Equality of empirical and simulated CDF's was tested *via* a Kolmogorov-Smirnov test of the maximum absolute difference [47]. Equality of variances was tested with an *F*-ratio test modified to account for skewness in the data [48]. Quartiles (0.25, 0.75) were compared by counting the number of simulated *BA*-values below (above) the historic quartiles and assuming that this number is a random variable drawn from a binomial distribution with 41 trials and probability of success equal to 0.25 and 0.75, respectively. Differences in skewness and kurtosis were assessed with a permutation test [49] with 1000 realizations from the null distribution generated by a random re-assignment of 41 historic and 41 simulated *BA*-values to either historic or simulated data.

The 5% level of significance was used as the threshold for rejection of the null hypothesis of equality. Simulation methods (MVN, MVT,...,NULL) were ranked from worst to best by the number of rejections of the six null hypotheses. The ranking was done for results at the regional, the grouped regional, and combined regional levels. Rejections of the hypotheses of equality of means, variances, and CDFs were considered as more serious than rejections for any other hypothesis. Equality of the 25 and 75 percentiles was, however, deemed more important than equality of skewness and kurtosis.

RESULTS

A summary of the forested areas burned annually between 1959 and 1999 in the 29 forest fire regions is in Table **2**. Area burned is expressed as a percentage (*BA*%) of the effective forest area with a minimum of one metric tonne of combustible forest biomass per ha [50]. The regional averages confirm a wide range from 0.01% in NS1 to 1.8% in MB1. Medians were much lower and had a narrower range (0.00%-0.60%). The minimum was close to 0% in every region while the regional maxima (mean = 4.7%) almost reached a catastrophic level in some regions. In all but a few regions the distribution of *BA%* was heavily right-skewed with a large standard deviation – typically two to three times the mean. (To qualify the large numbers from western Canada it should be emphasized that they cover areas affected by a forest fire (perimeter), not necessarily equal to the area of forest destroyed by fire).

Table 2. Summary Statistics of Percent Forest Area Burned Annually Between 1959 and 1999 in 29 Forest Fire Regions. Percentages are Computed as Area Burned Divided by the Forest Area with a Minimum of One Metric Tonne of Combustible Biomass Per Ha

Region	Mean	Median	Min	Max	Std. Dev.	Skewness
AB1	0.75	0.10	0.00	7.70	1.80	3.00
AB2	0.40	0.10	0.00	5.30	0.97	3.66
AB3	1.00	0.00	0.00	19.60	3.74	4.07
AB4	0.03	0.00	0.00	0.10	0.05	0.91
AB5	0.02	0.00	0.00	0.40	0.07	4.03
BC1	0.22	0.10	0.00	1.80	0.38	2.50
BC2	0.09	0.00	0.00	1.00	0.17	4.20
BC3	0.01	0.00	0.00	0.10	0.03	2.31
BC4	0.08	0.00	0.00	1.00	0.17	3.91
BC5	0.13	0.00	0.00	1.40	0.30	2.97
MB1	1.80	0.60	0.00	23.20	3.84	4.51
MB4	1.01	0.10	0.00	15.7	3.05	4.08
MB5	0.40	0.00	0.00	4.50	0.96	3.10
NB1	0.02	0.00	0.00	0.70	0.11	5.67
NF1	0.07	0.00	0.00	1.20	0.21	4.28
NF2	0.30	0.00	0.00	8.10	1.29	5.64
NS1	0.01	0.00	0.00	0.30	0.05	4.31
NWT1	0.06	0.00	0.00	0.70	0.18	2.96
NWT2	0.79	0.40	0.00	6.70	1.29	2.90
NWT3	0.68	0.30	0.00	5.40	1.20	2.84
ON1	0.05	0.00	0.00	0.50	0.11	2.97
ON2	0.32	0.10	0.00	2.60	0.64	2.62
PQ1	0.27	0.10	0.00	1.90	0.49	2.09
PQ2	0.06	0.00	0.00	0.90	0.17	3.84
PQ4	0.03	0.00	0.00	0.50	0.10	3.99
SK2	0.91	0.10	0.00	10.00	1.92	3.36
SK3	0.86	0.10	0.00	10.50	1.91	3.64
SK4	0.19	0.00	0.00	2.80	0.56	3.51
YK1	0.51	0.10	0.00	4.10	0.87	2.43

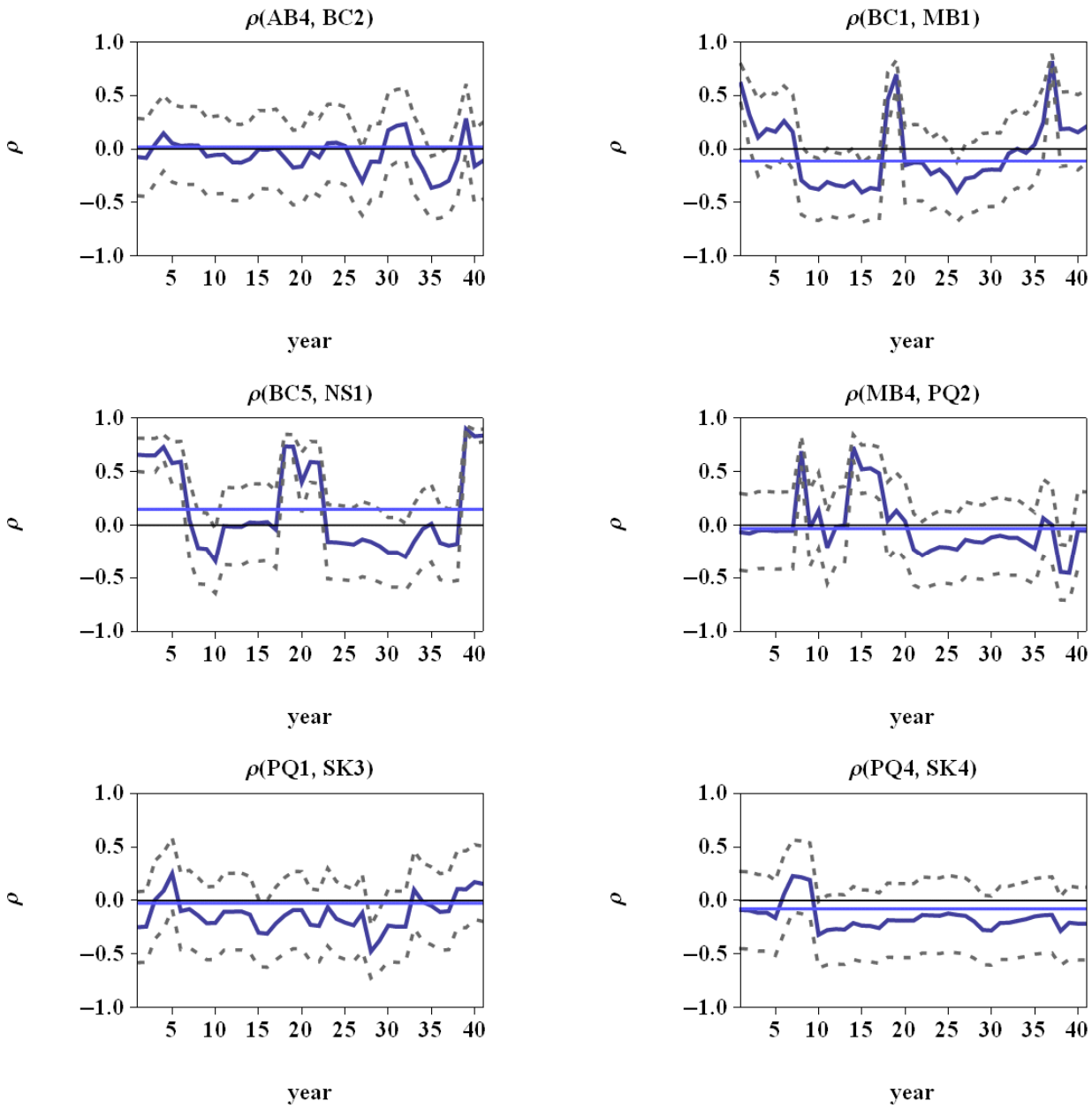

Fig. (7). Six examples of temporal regional correlation curves (full line) with a (dashed) confidence band of $\pm 1.\times \widehat{\text{se}}\left(\hat{\rho}\left(\text{year}\right)\right)$. The correlation coefficient estimated from the 1959-1999 data is indicated by the horizontal line.

All methods were, as expected, about equally good at reproducing the regional, grouped regional, and the combined regional average of annual area burned during a period of 41 years. The rejection rate of the null hypothesis of equality between a simulated and historic mean was consistently below 5%.

The variance of simulated regional *BA*-values was more likely to deviate significantly from the historic variance than the mean (Table **3**). The maximum (regional) rejection rate of the null hypothesis was well above 5% for all simulation methods. The median rejection rate, however, was fairly close to the nominal 5% which suggests that the variance of simulated *BA*-values in about half the regions was not significantly different from the observed variance. MVN, MVT, MIX, and NULL performed slightly better than BOOT, RND, and TS. In twelve regions the rejection rate for BOOT was over three times the expected rate of 5%. A large percentage (> 30%) of the BOOT simulated 41-year sequences had a variance that is either much larger or much smaller than the historic variance (Fig. **8**) - a typical problem encountered in resampling from a heavily skewed and short sequence of data. Our stratified bootstrap procedure mitigated this problem but did not eliminate it.

At the grouped-regional and combined regional level the ability to reproduce the inter-annual variance in *BA*-values varied considerably among the methods. BOOT, MVN,

Table 3. The Relative Frequency (%) of Rejecting the Null Hypothesis (H_0) of Equality Between Simulated and Observed Variance of Annual Area Burned During a 41 Year Period. Level of Significance = 5%

	BOOT	MVN	MVT	MIX	RND	TS	NULL
Regional (n = 29) **Mean**	9.4	6.7	6.7	6.3	10.6	9.5	6.7
Median	5.5	5.0	5.5	4.0	7.5	6.5	5.5
Min	0.0	0.5	0.5	0.0	1.0	1.5	0.0
Max	31.5	17.5	16.5	18.5	33.5	16.5	14.5
Supra-regional (n = 4) **Mean**	2.5	4.9	6.5	5.3	8.8	11.4	18.8
Median	2.5	5.0	6.3	5.5	8.3	11.5	14.3
Min	1.0	2.0	4.0	2.5	7.0	6.0	6.5
Max	4.0	7.5	9.5	7.5	11.5	16.5	40.0
All regions (n = 1)	0.5	5.0	4.5	4.5	7.5	12.0	61.2

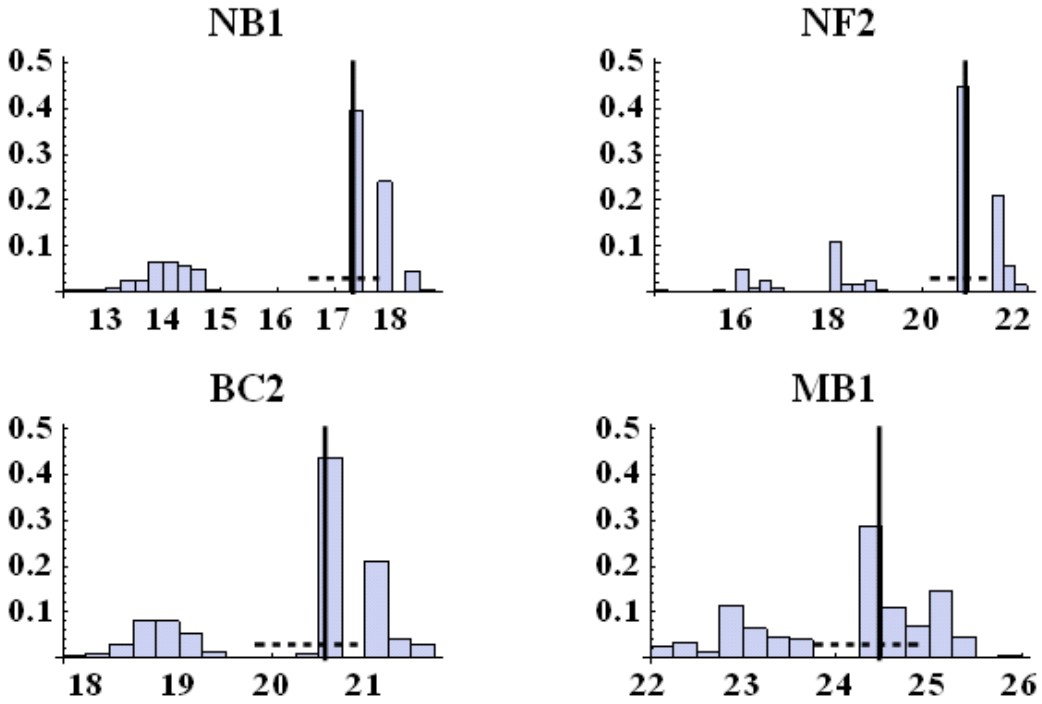

Fig. (8). Relative frequency distribution of the logarithm of the variance of bootstrap simulated area burned in four regions during a period of 41 years. The four regions are regions in which the variance of bootstrap simulations deviates significantly from the observed value. The logarithm of the observed variance and its approximate 95% confidence interval is indicated by a vertical and a dashed line, respectively.

MVT, and MIX performed best and NULL the worst (Table **3**). The poor performance of the NULL model is a clear indication of the significance of interregional correlations. Low variance in the NULL simulations caused the rejections (Figs. **9**,**10**). For example, the observed variance in CC was 2.3 times larger than in the NULL simulations - a result that agrees with an expectation of a ratio of 2.4 estimated from the average interregional correlation of 0.14 in CC [51]. NULL simulations for all regions combined were equally poor. Although the average regional correlation of *BA*-values is only 0.06 the impact on the variance of the regional total is an expected inflation of 280% over the expected variance with independent regions. The actual inflation (DATA / NULL) was 240%.

Cumulative distribution functions of simulated *BA*-values matched the historical *CDF*s quite well. The rate of rejections was consistently below 5% for all methods. A lack of statistical power [52] due to the sample size of 41 is considered to be the main cause of the poor discrimination between the methods.

First and last quartiles in 41 years of simulated regional *BA*-values were in general not significantly different from the historic quartiles. Methods like RND and TS performed poorly in regions AB5 and NB1 with unusually high rejec-

tion rates of the null hypothesis. For groups of regions or all regions combined, the NULL model is, however, again unacceptable with rejection rates well above 25%.

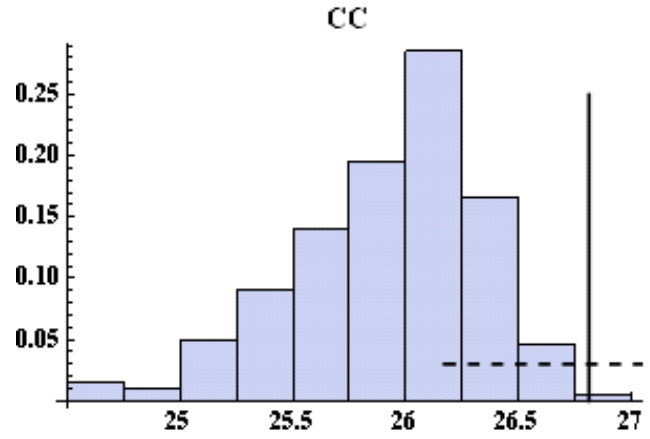

Fig. (9). Relative frequency distribution of the logarithm of the variance of simulated annual forest area burned in grouped region "CC" during a period of 41 years. Model: regional independence (NULL). The logarithm of the observed variance and its approximate 95% confidence interval is indicated by a vertical and a dashed line, respectively.

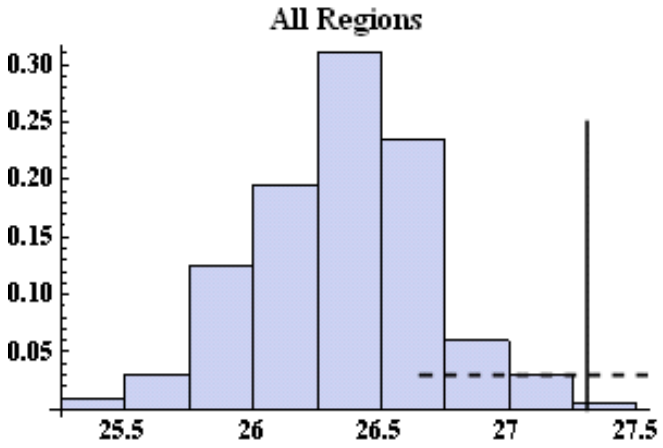

Fig. (10). Relative frequency distribution of the logarithm of the variance of simulated annual forest area burned in all regions combined during a period of 41 years. Model: regional independence (NULL). The logarithm of the observed variance and its approximate 95% confidence interval is indicated by a vertical and a dashed line, respectively.

All methods were about equally good at reproducing the coefficient of skewness and kurtosis in the historic data. Rejection rates under the null hypotheses were - with two exceptions - well below the nominal 5% level. Exceptions were for skewness in BOOT (overestimation) and MVN (underestimation) simulations. Yet it is clear that a sample size of 41 years is insufficient to detect even marked departures from historic values.

Overall MVT showed the most consistent performance at the grouped-regional and combined regional level. The erratic variances of the BOOT simulations led us to reject this method. MVN and MIX were close contenders but MVT is slightly better at reproducing skewness and quartiles. RND and TS came up short against historic data. The implicit (RND) and explicit (TS) smoothing of interregional correlations is the culprit. Yet a change in the simulation objectives or simulation period may improve their performance relative to the other methods.

A summary of MVT simulations for two randomly selected regions, grouped region CC (the supra region with the strongest average correlation between member regions), and the regional total, are in Figs. (**11-15**). At the regional level, the MVT simulations generated 41-year sequences of *BA*-values that resembled the historic sequence reasonably well, at least in terms of the six statistics considered. There was a small surplus of large 41-year averages (Fig. **11**) due to the asymmetric scaling of the empirical quantile functions that generated a much larger extension of the maximum than reduction of the minimum. There is also a slight surplus of variances smaller than expected in the MVT simulations (Fig. **12**), which challenge our assumption of a zero within-regional temporal autocorrelation. If years are not independent, our simulations will underestimate the variance within a 41-year sequence and overestimate the variance between periods of 41 years.

Shortcomings of the MVT simulations are of course most pronounced at the grouped regional and combined regional levels. For example, the number of simulated *BA*-values below the first historic quartile matched the expected number although the distribution of this number was distinctly different from the expected binomial distribution (Fig. **15**), a mismatch that became more pronounced for the third quartile. A small surplus of sequences with inflated skewness and kurtosis coefficients is also apparent.

DISCUSSION AND CONCLUSIONS

The joint simulation of area burned in Canada's forest fire regions remains a complex and difficult task. Multivariate simulations of highly skewed positive valued vectors with a dimension above three remains a challenge in the absence of a suitable transformation to a family of known parametric distributions. Candidate models are essentially limited to gamma-type or generalized lambda distributions [20, 31, 53-55]. Our data did not fit any of the popular parametric distribution functions, which severely curtailed our options. Linking 29 different marginal distributions into a joint multivariate distribution using copula or related techniques [30] would be computationally infeasible. Even if we treated five of the regions as independent, the dimension of the problem would still be unmanageable. A hierarchical Bayesian Markov Chain Monte Carlo (MCMC) approach to the simulation of complex multivariate latent processes has become popular [56] but we failed to detect any useful or consistent covariance pattern in our data that we could exploit within such a framework. The observed interregional correlation matrix was too irregular and patchy.

A joint simulation based on a mixture model [57] of, for example, multivariate gamma or log-normal distributions of

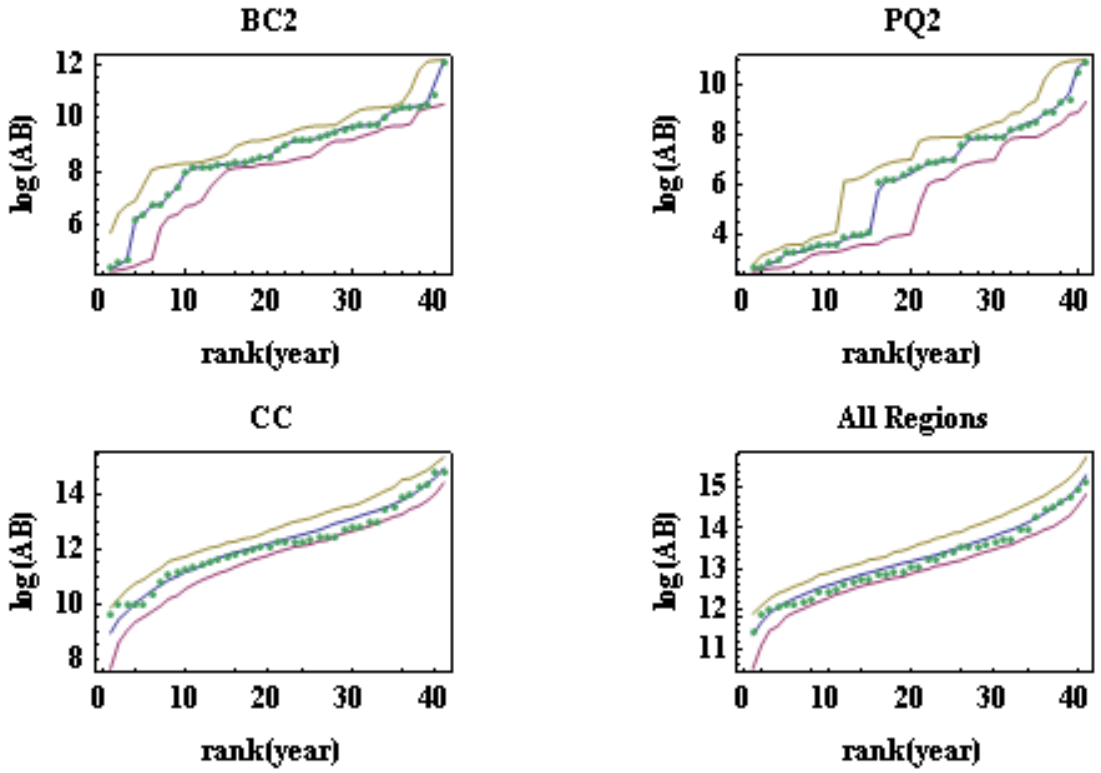

Fig. (11). Median of ranked (1,...,41) log-transformed annual area burned in MVT simulations (center line). Top and bottom lines demarcate the 95% simulation envelope in 200 replications. Observed (ranked) values are indicated by dots. Results are for two randomly selected regions, one grouped region ("CC") and all regions combined.

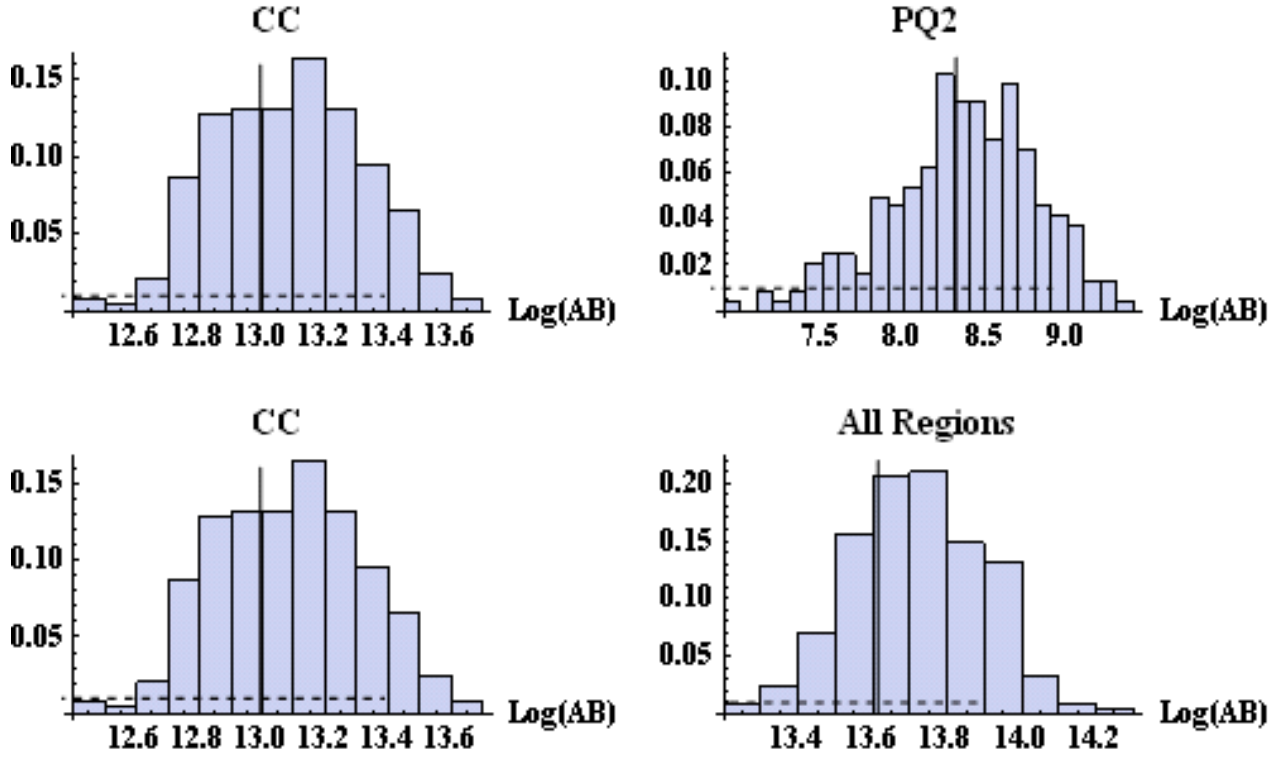

Fig. (12). Histogram of log-transformed 41-year averages of area burned in 200 MVT simulations. The observed value is indicated by the full vertical line. An approximate 95% confidence interval of this mean is indicated by the dashed line. Results are for two randomly selected regions, one grouped region ("CC") and all regions combined.

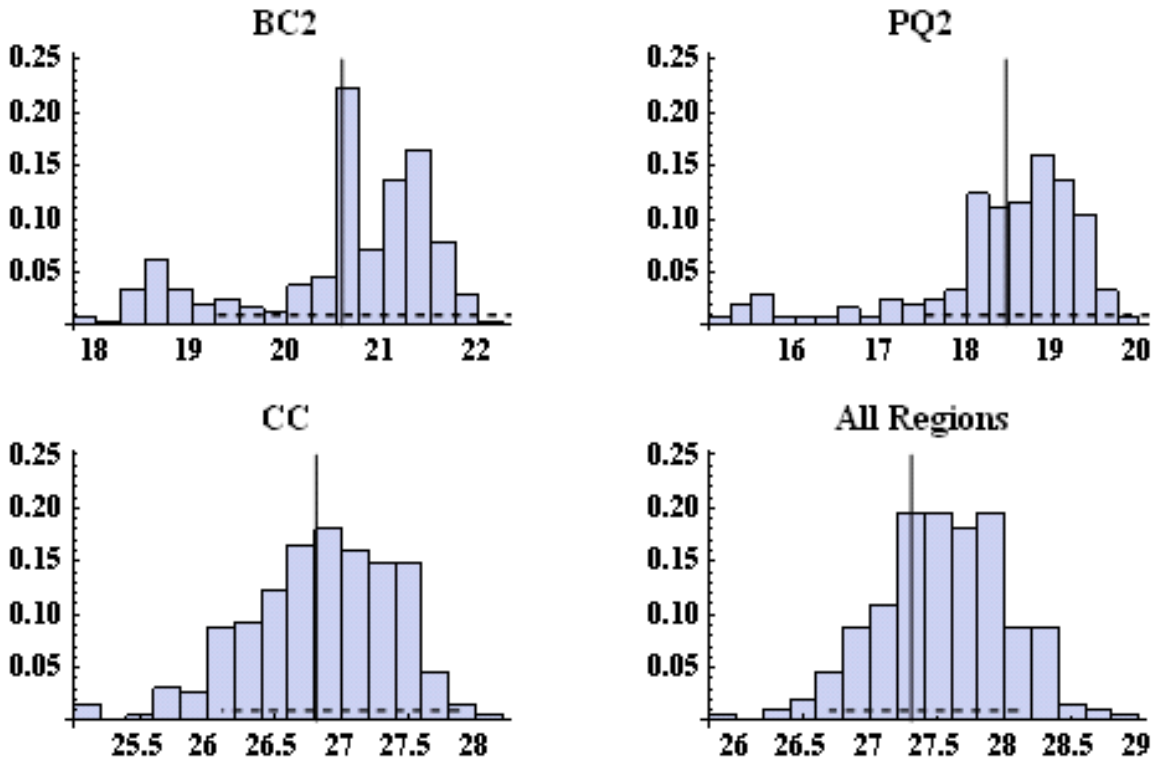

Fig. (13). Histogram of log-transformed 41-year variances of annual area burned in 200 MVT simulations. The observed log-transformed variance is indicated by the full vertical line. An approximate 95% confidence interval of this mean is indicated by a dashed line. Results are for two randomly selected regions, one grouped region ("CC") and all regions combined.

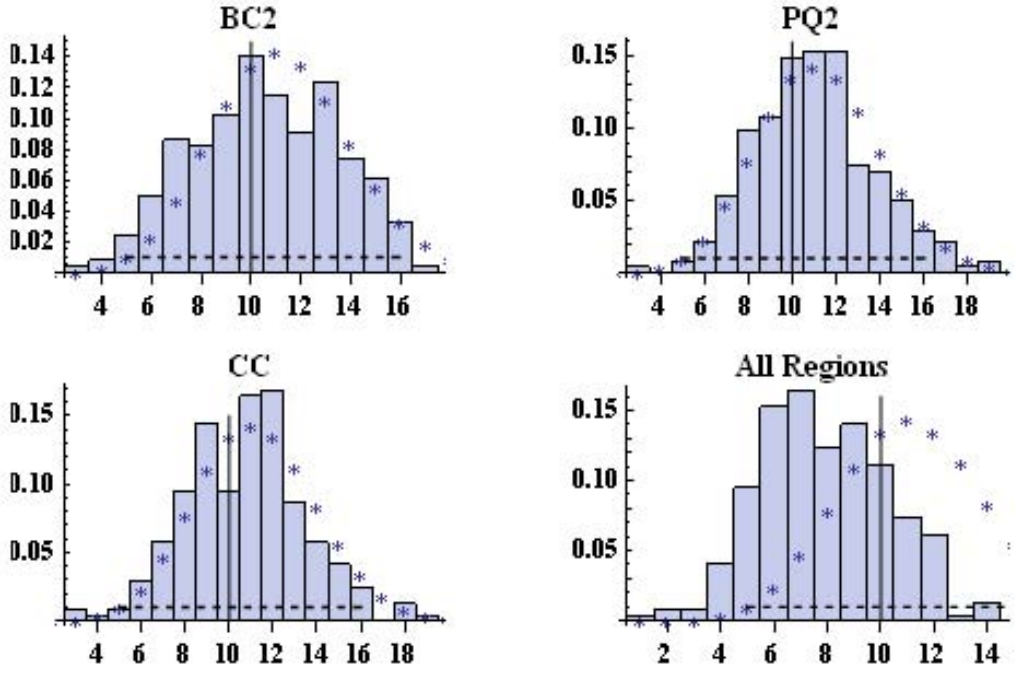

Fig. (14). Histogram of the number of MVT simulated values of annual area burned in a 41-year period that are smaller than the observed 25% percentile. Expected value is indicated by the vertical line (full) and a 95% confidence interval of this expectation under the assumption of a binomial distribution (stars) is indicated by the horizontal (dashed) line. Results are for two randomly selected regions, one grouped region ("CC") and all regions combined.

large and small *BA*-values might be an option for longer time-series of data. An observation period of 41 years is too short to allow a reliable estimation of mixture proportions and mixture components. Furthermore, combining marginal mixture models into a single joint simulation model would also lead to a MCMC approach.

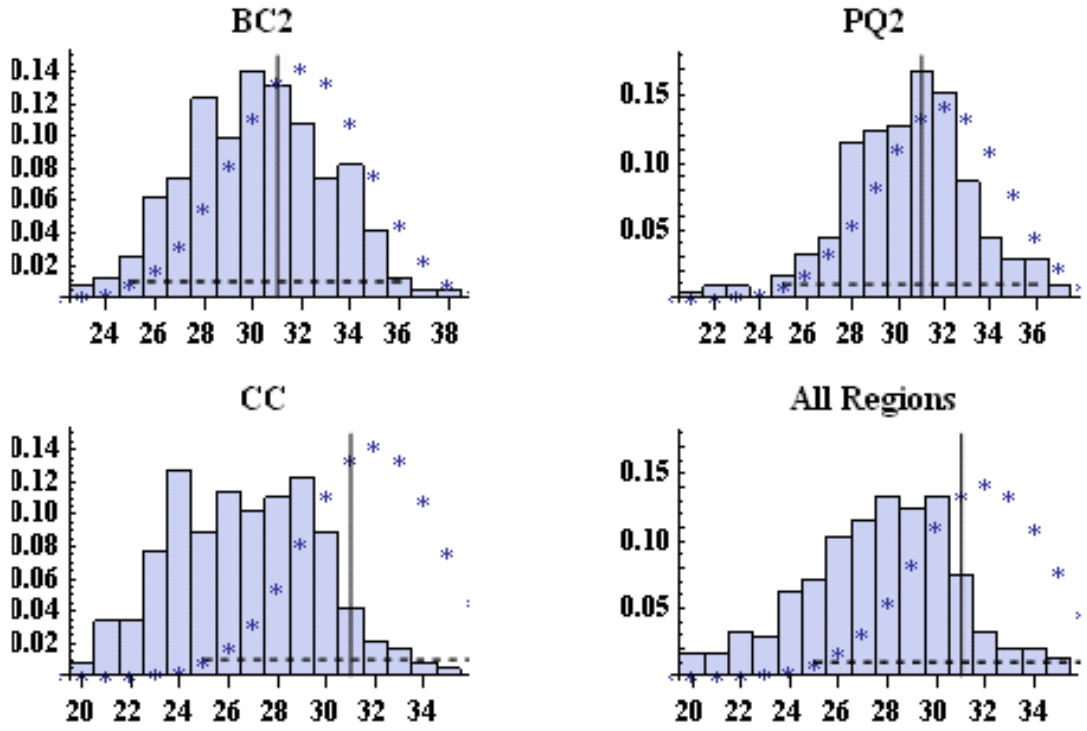

Fig. (15). Histogram of the number of MVT simulated values of annual area burned in a 41-year period that are under the observed 75% percentile. The expected value is indicated by the full vertical line and a 95% confidence interval of this expectation under the assumption of a binomial distribution (stars) is indicated by a dashed line. Results are for two randomly selected regions, one grouped region ("CC") and all regions combined.

The bootstrap procedure [43] is, in principle, ideally suited for joint simulation or recasting of historic data. To avoid a very jagged and irregular outline of the marginal bootstrapped distributions the data density must be quite high. In a 29-dimensional space a sample size of 41 years can only qualify as sparse [44]. Our stratified bootstrap procedure improved the situation but many marginal distributions of bootstrapped regional *BA*-values were still much more irregular than the historic distributions.

It was argued that regional *BA*-values from two distinct years could be regarded as independent. Support for a significant temporal first-order autocorrelation was limited to three regions (AB1, AB2, and NWT2). In addition, even in these few cases a shift in the observational period could either increase or decrease the correlation coefficient which suggests that temporal autocorrelations are unstable. A longer period of observation (> 50 yrs) would increase our power to detect evidence of a temporal autocorrelation during years characterized as having an above-average or a below-average fire-danger rating. Addition of a temporal dimension to the interregional correlations of *BA*-values would, of course, greatly complicate the task of finding a suitable model. A MCMC approach to the joint simulation would, again, be a natural framework for the simulations.

To use empirical distribution functions as the generator of the marginal distributions of regional area burned will satisfy the objective of a good fit between simulated and historic data, but only at the regional level. Yet the need to cap empirical distribution functions poses a nontrivial problem. In any event, the capping must be tailored to the simulation objectives. Even for our simple objective of simulating a 50-year sequence of regional *BA*-values that - in distribution - matches 41 historic records, the best choice of capping is not entirely clear. A much longer observational period is needed before use the observed minimum and maximum as caps would be justified. For a right-skewed distribution the expected maximum will increase more rapidly with sample size than the minimum is expected to decrease. Consequently, capping an empirical distribution function derived from a relatively small sample leads to a built-in lack-of-fit to historic data.

We saw only one case (ON1) of a significant time-trend in the data. Otherwise the extreme temporal variability in *BA*-values in most fire regions in Canada may have masked any actual trend. Had there been more pronounced time trends they would have had to be incorporated into the simulation process. A deterministic time-trend, however, would force the simulations to become year-specific stochastic predictions for either the period of years defined by the data or a future with predicted regional time trends. Trends would most likely differ between parts of Canada and greatly complicate the simulation procedure [58].

To satisfy simulation objectives pertaining to the joint distribution of regional *BA*-values, the choice of model for the generation of correlated uniformly distributed random variables on the unit interval (probits) becomes all-important. Only a few choices are realistic and they are essentially limited to the multivariate-normal and multivariate

t-distributions. We confirmed Bradley's experience with attenuated interregional correlations when a MVN distribution is used to generate correlated probits of non-Gaussian data. The attenuation problem is less severe in MVT simulations, but adopting this distribution entails a stronger negative bias in weaker correlations (-0.2 to 0). On balance the MVT approach has more appeal than the MVN approach. The optimal degree of freedom for the *t*-distribution will have to be found by trial-and-error.

Our performance criteria for the simulations were limited to a comparison of statistics of the simulated and historic data. For a simulation of future *BA*-values the criterion would need to be revised. We therefore advocate a cautious interpretation of a historic correlation structure in regional *BA*-values. We saw how selecting 10 years at random could change a 10-year correlation pattern considerably. The observed correlations may very well be unique to the years covered by the data. Attempts to obtain robust correlations [59] were discouraged by the non-Gaussian nature of the data. Adding a few more years of observations or dropping the most dated observations could change the correlation coefficients by non-trivial amounts. Our RND and TS procedures were designed to make the results more robust against changes in the time support. Output from RND and TS confirmed a shrinking (smoothing) of interregional correlation coefficients. A similar shrinking effect could arise in a MVT procedure using a random draw from a scaled Wishart distribution as the correlation matrix instead of $\hat{\mathbf{R}}$. On theoretical grounds, a simulation of future *BA*-values with RND or TS has appeal. Integrating them into an empirical Bayesian approach [60] with, for example, the first half of a historic record used to formulate priors on the model parameters and the second half used to compute a likelihood of a candidate model could be an interesting avenue of future research.

ACKNOWLEDGEMENTS

My colleagues M Flannigan, W Kurz, S Taylor, and W Yonghe provided many helpful comments and suggestions of improvement on an earlier version of this manuscript. Caren Dymond is thanked for organizing the data file.

REFERENCES

[1] Fearnside PM. Global warming and tropical land-use change: greenhouse gas emissions from biomass burning, decomposition and soils in forest conversion, shifting cultivation and secondary vegetation. Clim Chg 2000; 46: 115-58.

[2] Amiro BD, Todd JB, Wotton BM, *et al.* Direct carbon emissions from Canadian forest fires, 1959-1999. Can J For Res 2001; 31: 512-25.

[3] Stocks BJ, Fosberg MA, Lynham TJ, *et al.* Climate Change and Forest Fire Potential in Russian and Canadian Boreal Forests. Clim Chg 1998; 38: 1-13.

[4] Gillett NP, Weaver AJ, Zwiers FW, Flannigan MD. Detecting the effect of climate change on Canadian forest fires. Geophys Res L 2004; 31: L18211.

[5] Weber MG, Flannigan MD. Canadian boreal forest ecosystem structure and function in a changing climate: impact on fire regimes. Environ Rev 1997; 5: 145-166.

[6] Kurz W, Stinson G, Rampley G, Dymond C, Neilson E. Risk of natural disturbances makes future contribution of Canada's forests to the global carbon cycle highly uncertain. Proc Nat Acad Sci USA 2008; 105: 1551-1555.

[7] Balshi M, McGuire A, Zhuang Q, *et al*. The role of historical fire disturbance in the carbon dynamics of the pan-boreal region: A process-based analysis. Appl J Geoph Res 2007; 112.

[8] Beverly JL, Martell DL. Characterizing historical variability in boreal forest fire frequency: Implications for fire and forest management. Dordrecht NL: Kluwer; 2003.

[9] de Groot WJ. Modeling Canadian wildland fire carbon emissions with the Boreal Fire Effects (BORFIRE) model. In: Viegas DX, Ed. 2006: Figueira da Foz, Portugal: Elsevier, Amsterdam; 2006; pp.1-10.

[10] Ferragut L, Asenio MI, Montenegro R, *et al.* A model for fire simulation in landscapes. 1996: Paris, France: Wiley, Chichester UK; 1996; pp.111-16.

[11] Flannigan M, Campbell I, Wotton B, *et al.* Future fire in Canada's boreal forest: paleoecology results and general circulation model - regional climate model simulations. Can J For Res 2001; 31: 854-864.

[12] Tymstra C, Flannigan MD, Armitage OB, Logan KA. Impact of climate change on area burned in Alberta's boreal forest. Int J Wildl Fire 2007; 16: 153-160.

[13] Syphard AD, Clarke KC, Franklin J. Simulating fire frequency and urban growth in southern California coastal shrublands, USA. Landsc Ecol 2007; 22: 431-445.

[14] Cumming SG. A parametric model of the fire-size distribution. Can J For Res 2001; 31: 1297-1303.

[15] Rupp TS, Chen X, Olson M, McGuire AD. Sensitivity of simulated boreal fire dynamics to uncertainties in climate drivers - art. no. 3. Earth Interactions 2007; 11: 3.

[16] Campbell KA, Dewhurst SM. A hierarchical simulation-through-optimization approach to forest disturbance modelling. Ecol Model 2007; 202: 281-96.

[17] Kurz WA, Apps MJ. Developing Canada's national forest carbon monitoring, accounting and reporting system to meet the reporting requirements of the Kyoto protocol. Mit Adap Strat Glob Chg 2006; 11: 33-43.

[18] Gardner RH, Romme WH, Turner MG, *et al.* Spatial modeling of forest landscape change: Approaches and Applications. Cambridge, Cambridge University Press 1999; 163-85.

[19] Springer MD. The Algebra of Random Variables. New York: Wiley; 1979.

[20] Johnson ME. Multivariate Statistical Simulation. New York: Wiley; 1987.

[21] Robert CP, Casella G. Monte Carlo Statistical Methods. New York: Springer; 1999.

[22] Magnussen S. Fast pre-survey computation of the mean spatial autocorrelation in large plots composed of a regular array of secondary sampling units. Math Mod Sci Comp 2001; 13: 204-217.

[23] Martell DL. Wildfire regime in the boreal forest. Cons Biol 2002; 16: 1177.

[24] Stocks BJ, Mason JA, Todd JB, *et al.* Large forest fires in Canada, 1959-1997. J Geophys Res 2002; 108: 1-12.

[25] Trenberth KE, Hoar TJ. The 1990-1995 El Niño southern oscillation event: Longest on record. Geophys Res Lett 1996; 23: 57-60.

[26] Podur J, Martell DL, Csillag F. Spatial point pattern analysis of lightning-caused forest fires in the boreal forest region of Ontario. Syst Anal For Res Proc 2003; 7: 61-68.

[27] Larsen CPS. Fire and climate dynamics in the boreal forest of northern Alberta, Canada, from AD 1850 to 1989. The Holocene 2007; 6: 449-456.

[28] Bradley DR. Multivariate simulation with DATASIM: The Mihal and Barrett study. Beh Res Meth, Instr & Comp 1993; 25: 148-163.

[29] Cressie NAC. Statistics for Spatial Data. New York: Wiley; 1991.

[30] Nelsen RB. An Introduction to Copulas. New York: Springer; 1999.

[31] Karian ZA, Dudewicz EJ. Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods. Boca Raton, FL.: Chapman & Hall; 2000.

[32] Arnold BC, Castillo E, Sarabia JM. Families of multivariate distributions involving the Rosenblatt construction. J Amer Stat Assoc 2006; 101: 1652-1662.

[33] Giglio L, Descloitres J, Justice CO, Kaufman YJ. An enhanced contextual fire detection algorithm for MODIS. Rem Sens Env 2003; 87: 273-282.

[34] Ranson KJ, Kovacs K, Sun G, Kharuk VI. Disturbance recognition in the boreal forest using radar and Landsat-7. Can J Rem Sens 2003; 29: 271-285.
[35] Loboda T, Csiszar I. Estimating burned area from AVHRR and MODIS: validation results and sources of error. Proceedings of the Current Aspects of Rem Sens Earth Space Conf, Moscow, Russia; 2004: pp.16-18.
[36] Rubenstein RY. Simulation and the Monte Carlo Method. New York: Wiley; 1981.
[37] Anderson OD. Time Series Analysis and Forecasting. The Box-Jenkins approach. London: Butterworths; 1976.
[38] Fleishman AIA. A method for simulating nonnormal distributions. Psychometrika 1978; 43: 521-532.
[39] Vale CD, Maurelli VA. Simulating multivariate nonnormal distributions. Pscychometrika 1983; 48: 465-471.
[40] Atkinson AC. Transforming both sides of a tree. Am Stat 1994; 48: 307-313.
[41] Serfling RJ. Approximation Theorems of Mathematical Statistics. New York: J. Wiley Sons; 1980.
[42] Sarhan AE, Greenberg BG. Contributions to Order Statistics. Wiley; 1962.
[43] Efron B, Tibshirani RJ. An Introduction to the Bootstrap. Boca Raton: Chapman & Hall; 1993.
[44] Scott DW. Multivariate density estimation: Theory, Practice and Visualization. New York: Wiley; 1992.
[45] Doksum K, Blyth S, Bradlow E, Meng XL, Zhao H. Correlation curves as local measures of variance explained by regression. J Amer Stat Assoc 1994; 89: 571-582.
[46] Waller LA, Smith D, Childs JE, Real LA. Monte Carlo assessments of goodness-of-fit for ecological simulation models. Ecol Model 2003; 164: 49-63.
[47] Barr DR, Davidson T. A Kolmogorov-Smirnov test for censored samples. Technometrics 1973; 15: 732-757.
[48] Shoemaker LH. Fixing the F test for equal variances. Am Stat 2003; 57 105-114.
[49] Good P. Permutation Tests. New York: Springer; 1993.
[50] Power K, Gillis M. Canada's Forest Inventory 2001. 2006; Can For Serv Pac For Cent Inf Rep BC-X-408.
[51] Cochran WG. Sampling Techniques. New York: Wiley; 1977.
[52] Thomas DR, Rao JNK. Small-sample comparisons of level and power for simple goodness-of-fit statistics under cluster sampling. J Amer Stat Assoc 1987; 82: 630-636.
[53] Carpenter M, Diawara N. A multivariate gamma distribution and its characterization. 2007.
[54] Aalo VA, Piboongungon T. On the multivariate generalized gamma distribution with exponential correlation. 2005: Proc IEEE Glob Telecom Conf GlobeCom'05, St Louis Missouri; 2005; pp.5.
[55] Nomoto S, Kishi Y, Nanba S. Multivariate gamma distributions and their numerical evaluations for M-branch selections diversity study. Electronics and Communications in Japan (Part I: Communications) 2004; 87: 1-12.
[56] Gilks WR, Richardson S, Spiegelhalter DJ.In: Gilks WR, Richardson S, Spiegelhalter DJ, Eds. Markov Chain Monte Carlo in Practice. Bacon Raton, Chapman & Hall/CRC 1996; 1-19.
[57] Casella G, Robert CP, Wells MT. Mixture models, latent variables and partitioned importance sampling. Statist Meth 2004; 1: 1-18.
[58] Bergeron Y, Flannigan M, Gauthier S, Leduc A, Lefort P. Past, current and future fire frequency in the canadian boreal forest: implications for sustainable forest management. Ambio 2004; 33: 356-360.
[59] Maronna R. Robust M-Estimators of multivariate location and scatter. Ann Stat 1976; 4: 51-67.
[60] Congdon P. Bayesian Statistical Modelling.2. Chichester, England: Wiley; 2006.

Received: March 28, 2008 Revised: June 5, 2008 Accepted: June 5, 2008