Asymptotic results for the MPL estimators of the Contact Process

Xavier Guyon[1] and Besnik Pumo[2]

[1] SAMOS, Université Paris 1,
90 rue de Tolbiac, 75634 Paris 13e, France
(e-mail: `Xavier.Guyon@univ-paris1.fr`)

[2] Institut National d'Horticulture - UMR SAGAH,
42 rue G. Morel, BP 60057, 49071 Beaucouzé, France
(e-mail: `Besnik.Pumo@inh.fr`)

Abstract. Let X be a discrete time contact process (CP) on $\mathbf{Z}^2$ as defined by Durrett and Levin (1994). We study the estimation of the model based on space-time evolution of X, that is, $T+1$ successive observations of X on a finite subset S of sites. We consider the maximum marginal pseudo-likelihood (MPL) estimator and show that, when $T \to \infty$, this estimator is consistent and asymptotically normal for a non vanishing supercritical CP. Numerical studies confirm the theoretical results and compare the MPL estimators with coding method estimators. Finally we present some results on CP of order d.

Keywords: Contact process, supercritical process, marginal pseudo-likelihood, identifiability of a model, consistency, asymptotic normality.

1 Introduction and description of the model

Consider a simple model of spread of a single species population evolving in $\mathbf{Z}^2$. Depending on some biological parameters, the dynamics is determined by specifying, for each site $s \in \mathbf{Z}^2$, the conditional probability that site s will be in state $X_{t+1}(s) = y \in \{0,1\}$ at time $t+1$ given X_t, the configuration at time t. State 1 (respectively 0) means that there is a (respectively no) plant in s. In this paper we propose an estimator for the parameters of the model, based on observations of X at instants $t = 0, \dots, T$ on a finite and fixed subset S of $\mathbf{Z}^2$ and study the asymptotic properties of the estimator when the process is non vanishing on S. Fiocco and Zwet considered the estimation problem based on one observation at time t, when t is sufficiently large ([Fiocco and Zwet, 2003]).

We consider the *discrete time version* of the *Contact Process* (CP) as defined by Durrett & Levin [Durrett and Levin, 1994]. Suppose that the transition probability at a site s and at time t is stationary in space and time and depends locally on $x_{t-1}(\mathcal{N}_1(s))$, the first order neighbourhood of the site s at time $t-1$, where $\mathcal{N}_d(s) = \{u \in \mathbf{Z}^2 : \|s-u\|_1 \leq d\}$.

The system evolves as follows:

a. Each plant alive at time t dies with a probability γ at time $t+1$,

b. If the plant in s survives, then it produces an offspring that is dispersed to $u \in \partial s$, where $\partial s = \mathcal{N}_1(s) \setminus \{s\}$, with probability λ; the reproduction events for different values of s and different $u \in \partial s$ are independent,
c. If one or more plants are dispersed to s, or if there is a plant at s that survives between t and $t+1$, then $X_{t+1}(s) = 1$; otherwise $X_{t+1}(s) = 0$.

Furthermore, events defined on (a) and (b) are independent in time.

This model depends on the parameter $\theta = (\gamma, \lambda)$ and we suppose that $\theta \in (0,1)^2$. Other models are possible by defining different rules of evolution (cf. [Mollison, 1977] for example). Finally, some of the methods developed in our paper can be generalized for non stationary processes in space and/or in time.

The 'all 0' in $\mathbf{Z}^2$ state is an absorbing state. So, to make sense, for the asymptotic study, we need a condition **(I)**, verified with probability 1 conditionally to the non-extinction of X on S, the fixed domain of observation. Note that a CP survives with positive probability for a supercritical process that is CP such that $P(\tau = +\infty) > 0$ where τ gives the extinction time of the process ([Durrett and Levin, 1994]).

The paper is organized as follows. In section 2 we define the *marginal pseudo-likelihood* (MPL) estimator of θ. The identifiability of MPL is presented in section 3 and asymptotic results of MPL estimator in section 4. In section 5 we consider some simulations studies and compare numerically MLP estimators with coding method estimators proposed by Besag ([Besag, 1972]). A brief discussion on CP of order d is given in section 6.

Proofs of results are to be found in [Guyon and Pumo, 2004].

2 Marginal pseudo-likelihood (MPL)

Let $\mathrm{x(T)} = (\mathrm{x_0}, \mathrm{x_1}, \cdots, \mathrm{x_T})$ be $(T+1)$ successive configurations of X, S a finite subset of $\mathbf{Z}^2$ and $S_1 = \{u \in \mathbf{Z}^2 : \exists v \in S \text{ such that } \|u - v\|_1 \leq 1\}$. The estimator of θ we choose is a value which maximize a MPL of $\mathrm{x}(T)$ observed on S_1. The idea of pseudo-likelihood is classic in statistic: gaussian pseudo-likelihood for stationary field on $\mathbf{Z}^d$ ([Whittle, 1963]), conditional pseudo-likelihood for a Markov field on a lattice ([Besag, 1974]).

For a subset $A \subset S$, let denote $P_A(x_t, x_{t+1}; \theta)$ the transition-probability $P(X_{t+1}(A) = x_{t+1}(A) \mid X_t(S_1) = x_t(S_1))$. As the transition-probability for $A = S$ is analytically intractable, as $\#(S)$, the number of sites of S, is important, we will use the following *marginal pseudo-transition probability* $M_S(x_t, x_{t+1}; \theta)$ on S, in order to estimate θ. $M_S(x_t, x_{t+1}; \theta)$ is the product of $P_{\{s\}}(x_t, x_{t+1}; \theta)$ for $s \in I(x_t)$, where:

$$I(x_t, S) = \{s \in S : \ \exists x_{t+1} \ \text{ s.t. } P_{\{s\}}(x_t, x_{t+1}; \theta) > 0\}$$

The product of these marginal pseudo-transitions at consecutive instants define the MPL. For $s \in S$ and A a finite subset of $\mathbf{Z}^2$, denote $m(x_t, A) =$

$\sum_{s\in A} x_t(s)$, the number of sites of A occupied by x_t. As the model is isotropic in space, the law of $X_{t+1}(s)$ given x_t depends only on $c(x_t, s)$:

$$c(x_t, s) = (x_t(s), m(x_t, \partial s)) \in \mathcal{C}_1 = \{0, 1\} \times \{0, 1, 2, 3, 4\}. \quad (1)$$

More precisely, $X_{t+1}(s)$ conditionally on x_t is a Bernoulli random variable:

$$P_{\{s\}}(x_t, x_{t+1}; \theta) = p(x_t, s; \theta)^{1-x_{t+1}(s)}(1 - p(x_t, s; \theta))^{x_{t+1}(s)},$$

where $p(x_t, s; \theta) = \gamma^{x_t(s)}\delta^{m(x_t,\partial s)}$ and $\delta = \gamma + (1 - \gamma)(1 - \lambda)$ controls non-proliferation at time $(t + 1)$ in a site $s' \in \partial s$ of a plant present in s at time t. Since $X_{t+1}(s) = 0$ if $c(x_t, s) = (0, 0)$, only sites $s \in I(x_t)$ are informative in the transition $t \mapsto t + 1$. So:

$$M_S(x_t, x_{t+1}; \theta) = \prod_{s\in I(x_t)} p(x_t, s; \theta)^{1-x_{t+1}(s)}(1 - p(x_t, s; \theta))^{x_{t+1}(s)} \quad (2)$$

with convention $M(0, 0; \theta) = 1$ if $I(x_t) = \emptyset$. Denote $\eta = \gamma + (1 - \gamma)(1 - \lambda)^2$: η controls non-proliferation at time $(t+1)$ in the set $\{s, s'\}$ of a plant present in $u \in \partial s \cap \partial s'$ at time t.

By a direct calculation it follows that:

$$Cov(X_{t+1}(s), X_{t+1}(s') \mid x_t) = p(x_t, s; \theta)\ p(x_t, s'; \theta)\ [b(x_t, s, s'; \theta) - 1]$$

where

$$b(x_t, s, s'; \theta) = \begin{cases} \delta^{-m(x_t,\{s,s'\})} & if\ s' \in \mathcal{N}_1(s) \setminus \{s\} \\ \delta^{-2m(x_t,\partial s\cap\partial s')}\eta^{m(x_t,\partial s\cap\partial s')} & if\ s' \in \mathcal{N}_2(s) \setminus \mathcal{N}_1(s) \\ 1 & if\ s' \notin \mathcal{N}_2(s). \end{cases} \quad (3)$$

In particular if $s' \notin \mathcal{N}_2(s)$, $(X_{t+1}(s) \mid x_t)$ and $(X_{t+1}(s') \mid x_t)$ are independent.

Using (2) for $t = 0, \cdots, T - 1$, let us give the explicit expression of MPL based on observation of x(T) on S_1. Denote $n(x_t)$ (respectively $n(x_t, c)$) the number of informative sites of the configuration x_t on S (respectively with configuration $c \in \mathcal{C}_1$) and:

$$n(T) = \sum_{t=0}^{T-1} n(x_t), \quad n(T, c) = \sum_{t=0}^{T-1} n(x_t, c).$$

Clearly $n(T) = \sum_{c\neq(0,0)} n(T, c)$. The normalized log-marginal pseudo-likelihood of x(T) observed on S_1 is:

$$l_T(\theta) = \frac{1}{n(T)} \sum_{t=0}^{T-1} \sum_{s\in I(x_t)} \{\log[p(x_t, s; \theta)]^{\bar{x}_{t+1}(s)} + \log[\bar{p}(x_t, s; \theta)]^{x_{t+1}(s)}\} \quad (4)$$

where $\bar{x}_{t+1}(s) = 1 - x_{t+1}(s), \bar{p}(x_t, s; \theta) = 1 - p(x_t, s; \theta)$. The maximum MPL estimator of θ (or MPLE) is a value which maximize the MPL,

$$\hat{\theta}_T = \arg_\theta \max\ l_T(\theta).$$

3 MPL allows identification of θ

In order to prove that MPL allows identification of θ, we need to show that π_c is strictly positive for two linearly independent configurations, where:

$$\pi_c = \underline{\lim}_{T\to\infty} \frac{n(T,c)}{n(T)}.$$

The positivity of $\pi_c > 0$ for $c \in \mathcal{C}_1^*$, the set of configurations on $\mathcal{N}_1(0)$ such that $x(0) = 1$, is obtained by the following Lemma under the condition **(I)** of non-extinction of X on S :

$$(\mathbf{I}) : I_\infty = \{\mathrm{x} = (x_t, t \geq 0) \ \textit{such} \ n(\mathrm{x(T)}) \to \infty \text{ as } \mathrm{T} \to \infty\}.$$

Lemma 1 *Let $\mathcal{C}_1^*$ be the set of configurations on $\mathcal{N}_1(0)$ such that $x(0) = 1$. Then there exists $\alpha > 0$ such that, $\forall c \in C_1^*$, and $\forall x \in I_\infty$, we have $\pi_c \geq \alpha$.*

From the positivity of π_c, it follows that under **(I)** and for large T, $\theta \to l_T(\theta)$ allows identification of θ. Indeed:

- if $\mathrm{x}(T)$ realizes two linearly independent configurations $c_a = (u_a, v_a)$ and $c_b = (u_b, v_b)$, then $\theta \mapsto l_T(\theta)$ is an injective function;
- under **(I)**, the probability that each configuration c of $\mathcal{C}_1^*$ appears on S converges to 1 when $T \to \infty$.

In conclusion let as make two important remarks:

i) As X_∞ is spatially translation-invariant and ergodic, [Durrett, 1995], it follows that $\lim_{T\to\infty} \frac{n(T,c)}{n(T)}$ exists and is strictly positive for $c \in \mathcal{C}_1$.

ii) Space and/or time invariance of the model is not crucial on the proof of the subergodicity result: a similar result can be proved for non translation invariant models under the supplementary condition that transition probabilities are uniformly positive.

4 Consistency and normality of the MPL estimator

Let $f : U \to R$ be a real function twice continuously differentiable on an open subset U of $\mathbf{R}^d$ and $f^{(1)}(\theta)$ the vector of first derivatives. The following result sets up the consistency and asymptotic normality of the maximum MPLE $\hat{\theta}_T$ associated to (4). The proofs are based on Theorem 3.4.3 and 3.4.5 of Guyon ([Guyon, 1995]). In order to prove the positivity of $J_T(\theta_o)$ we used an idea of Jensen and Künsch ([Jensen and Künsch, 1994]) and a subergodicity result which generalize Lemma 1.

Let I_2 be the 2×2 identity matrix, $A_T(\theta_o)$, $B_T(\theta_o)$ the 2×2 matrices:

$$A_T(\theta_o) = \frac{1}{n(T)} \sum_{t=0}^{T-1} \sum_{s \in I(x_t)} \frac{p^{(1)t}[p^{(1)}]}{p(1-p)}(x_t, s; \theta_o) \tag{5}$$

$$B_T(\theta_o) = \frac{1}{n(T)} \sum_{t=0}^{T-1} \sum_{s,s' \in I(x_t)} [b(x_t, s, s'; \theta_o) - 1] \frac{p^{(1)}(x_t, s; \theta_o)\, {}^t[p^{(1)}(x_t, s'; \theta_o)]}{[\bar{p}(x_t, s; \theta_o)]\, [\bar{p}(x_t, s'; \theta_o)]} \tag{6}$$

with $b(x_t, s, s'; \theta_o)$ given by (3).

Theorem 1 *Let us suppose that $\theta_o = (\gamma_o, \lambda_o)$, the true unknown value of the parameter, is an interior point of a compact $\Theta \subset]0,1[^2$. Then, under condition **(I)** the maximum MPL estimator is consistent:*

$$\lim_{T \to \infty} \hat{\theta}_T \overset{a.s.}{=} \theta_o.$$

and asymptotically normal:

$$\sqrt{n(T)}\, [A_T(\theta_o) + B_T(\theta_o)]^{-1/2} A_T(\theta_o)(\hat{\theta}_T - \theta_o) \xrightarrow{d} \mathbf{G}_2(0, I_2).$$

5 Numerical studies

In this section we give some empirical results with S the 64×64 square lattice and initial configuration 'all sites occupied'. To avoid boundary effects we have used periodic boundary conditions. In Fig. 1 we present the evolution

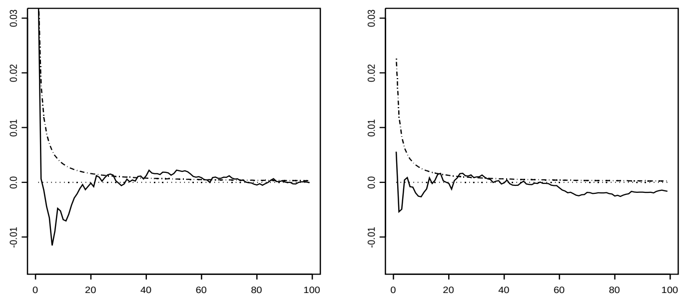

Fig. 1. Evolution of the bias (solid lines) and standard deviation (multiplied by 100, dotted lines) for the estimators of γ_o (left) and λ_o (right) for the supercritical CP with parameters $\gamma_o = 0.35, \lambda_o = 0.25$.

of the bias and the standard deviation of $\hat{\gamma}_T$ and $\hat{\lambda}_T$ for $T = 1, \ldots, 99$ for the supercritical CP with parameters $\gamma_o = 0.35, \lambda_o = 0.25$.

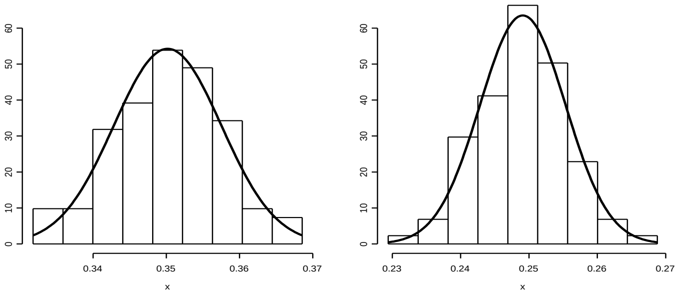

Fig. 2. Histograms of 100 estimations of γ_o (left) and λ_o (right) for the supercritical CP with parameters $\gamma_o = 0.35, \lambda_o = 0.25$.

Empirical study of asymptotic normality of estimators for a supercritical CP is based in 100 simulations with $T = 99$. Histograms are presented in Fig. 2. Asymptotic normality is checked by using a chi-squared test at level 5% and defining 9 equiprobable classes. Normality is accepted for $\hat{\gamma}$ (respectively $\hat{\lambda}$) since $\chi^2 = 1.7$ (respectively $\chi^2 = 4.4$) and $\chi^2_{0.95}(6) = 12.59$.

We also compared the estimated standard errors $\hat{\sigma}_{\hat{\gamma}}, \hat{\sigma}_{\hat{\lambda}}$ and empirical standard errors $s_{\hat{\gamma}}, s_{\hat{\lambda}}$ for the supercritical CP with parameter $\gamma_o = 0.35$, $\lambda_o = 0.25$. The values $\hat{\sigma}_{\hat{\gamma}_4}, \hat{\sigma}_{\hat{\lambda}_4}$ are obtained from a single simulation with $T = 4$ by applying Theorem 1 where $A_4(\theta_o)$ (respectively $B_4(\theta_o)$) are approximated by $A_4(\hat{\theta}_4)$ (respectively $B_4(\hat{\theta}_4)$). The empirical standard errors $s_{\hat{\gamma}_4}, s_{\hat{\lambda}_4}$ are obtained from 100 estimations for the 100 simulations. The results are presented in Table 1. As expected, there are few differences between estimated standard errors and empirical standard errors. Finally, Ta-

	$\hat{\sigma}_{\hat{\gamma}_4}$	$s_{\hat{\gamma}_4}$	$\hat{\sigma}_{\hat{\lambda}_4}$	$s_{\hat{\lambda}_4}$
MPL estimations	0.0074	0.0074	0.0063	0.0058

Table 1. Comparison of estimated and empirical standard deviation

ble 2 gives the estimations of γ_o and λ_o for six CP with parameters (γ_o, λ_o) $\in (0.2, 0.4, 0.6) \times (0.1, 0.2)$. In these simulations, $T = 4$ and 40% of sites, randomly chosen, were occupied at time $t = 0$. We compare MPL estimators with coding method of estimation introduced by Besag ([Besag, 1972]). Let $K = 3 \times \mathbf{Z}^2 \cap S$, a *strong-coding subset* that is $\partial s \cap \partial s' = \emptyset$ for $s \neq s'$ of K. As variables $\{(X_{t+1}(s) \mid X_t = x), s \in K\}$ are independent, the normalized log-conditional likelihood of the CP restricted on sites s of K is given by:

$$l_{T,K}(\theta) = \frac{1}{n_K(T)} \sum_{t=0}^{T-1} \sum_{s \in I_K(x_t)} \{\log[p(x_t, s; \theta)]^{\bar{x}_{t+1}(s)} + \log[\bar{p}(x_t, s; \theta)]^{x_{t+1}(s)}\}$$

	$\lambda = 0.1$					$\lambda = 0.2$				
γ	$\hat{\gamma}_4$	$\hat{\sigma}_{\hat{\gamma}_4}$	$\hat{\lambda}_4$	$\hat{\sigma}_{\hat{\lambda}_4}$	$n(4)$	$\hat{\gamma}_4$	$\hat{\sigma}_{\hat{\gamma}_4}$	$\hat{\lambda}_4$	$\hat{\sigma}_{\hat{\lambda}_4}$	$n(4)$
0.2	0.210	0.006	0.104	0.003	14600	0.189	0.006	0.193	0.004	15285
0.4	0.391	0.008	0.106	0.004	12406	0.399	0.008	0.188	0.005	13573
0.6	0.597	0.009	0.100	0.005	9223	0.607	0.009	0.206	0.008	10457

Table 2. Estimation of the parameters and their standard deviation.

where $I_K(x_t)$ gives the set of informative sites of K, $n_K(x_t) = \sharp(I_K(x_t))$ and $n_K(T) = \sum_{t=0}^{T-1} n_K(x_t)$. The *$K$-coding estimator* of θ is a value which maximize $l_{T,K}(\theta)$. By applying this method of estimation for six CP we obtained the results presented in Table 3.

	$\lambda = 0.1$					$\lambda = 0.2$				
γ	$\hat{\gamma}_4$	$\hat{\sigma}_{\hat{\gamma}_4}$	$\hat{\lambda}_4$	$\hat{\sigma}_{\hat{\lambda}_4}$	$n(4)$	$\hat{\gamma}_4$	$\hat{\sigma}_{\hat{\gamma}_4}$	$\hat{\lambda}_4$	$\hat{\sigma}_{\hat{\lambda}_4}$	$n(4)$
0.2	0.217	0.014	0.100	0.008	2442	0.183	0.013	0.170	0.010	2568
0.4	0.400	0.019	0.113	0.011	2106	0.400	0.019	0.184	0.014	2256
0.6	0.590	0.022	0.084	0.013	1542	0.602	0.022	0.198	0.023	1766

Table 3. Estimation of the parameters and their standard deviation obtained by K-coding method

6 Estimation of parameters of CP of order *d*

In this section we briefly present results for the CP of order d presented also in [Pumo and Le Corff, 2001] and which generalize the standard CP defined in the introduction. Denote ∂s a general neighbourhood of s. In order to define the CP of order d we only substitute b in the definition of the standard CP with b':

b'. If the plant in s survives, then it produces an offspring that is dispersed to $u = z + s \in \partial s$ with probability $g(z)$; the reproduction events for different values of s and different $u \in \partial s$ are independent,

Denote $\lambda = (\lambda_1, \ldots, \lambda_d)'$ the vector of different values of $g(z)$, $z \in \partial 0 \backslash \{0\}$. Then we call d the order of the CP. The unknown parameter θ is defined now by $\theta = (\gamma, \lambda')$. It can be shown that similar results remains valid for the CP of order d. Furthermore, by applying Theorem 3.4.6 in [Guyon, 1995] we can do tests on parameters λ in order to determine the optimal neighbourhood for the definition of the model. In Table 4 we give estimations of six CP of order 2 with parameters $\theta_o = (\gamma_o, \lambda_{1o}, \lambda_{2o})$ where $(\gamma_o, \lambda_{1o}) \in (0.2, 0.4, 0.6) \times (0.1, 0.2)$ and $\lambda_{2o} = \lambda_{1o}/\sqrt{2}$. In these simulations we considered a 100×100 lattice and at time $t = 0$ all sites were occupied.

	$\lambda_1 = 0.1, \lambda_2 = 0.0707$			$\lambda_1 = 0.2, \lambda_2 = 0.1414$		
γ	$\hat{\gamma}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\gamma}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$
0.2	0.199	0.098	0.072	0.200	0.197	0.144
0.4	0.399	0.101	0.072	0.399	0.200	0.141
0.6	0.597	0.102	0.065	0.598	0.199	0.135

Table 4. Estimation of parameters of CP of order 2.

References

[Besag, 1972]J. Besag. On the statistical analysis of nearest-neighbour systems. In *Proc. 9th Europian Meetying of Statist., Budapest*, pages 101–105, 1972.

[Besag, 1974]J. Besag. Spatial interaction and the statistical analysis of lattice systems. *JRSS B*, pages 192–225, 1974.

[Durrett and Levin, 1994]R. Durrett and S.A. Levin. Stochastic spatial models: a user's guide to ecological applications. *Phil. Trans. R. Soc. Lond.*, pages 329–350, 1994.

[Durrett, 1995]R. Durrett. *Ten lectures on particle systems. Cours de Saint Flour (1993).* Springer, USA, 1995.

[Fiocco and Zwet, 2003]M. Fiocco and W.R. Zwet. Parameter estimation for the supercritical contact process. *Bernoulli*, pages 1071–1092, 2003.

[Guyon and Pumo, 2004]X. Guyon and B. Pumo. Estimation spatio-temporelle d'un système de particules. *Préprint SAMOS*, pages 1–21, 2004.

[Guyon, 1995]X. Guyon. *Random fields on a network: modelling, statistics and applications.* Springer, Berlin, 1995.

[Jensen and Künsch, 1994]J.L. Jensen and H.R. Künsch. On asymptotic normality of pseudo-likelihood estimate for pairwise interaction processes. *Ann. Inst. Statist. Math.*, pages 475–486, 1994.

[Mollison, 1977]D. Mollison. Spatial contact models for ecological and epidemic spread, with discussion. *JRSS B.*, pages 283–326, 1977.

[Pumo and Le Corff, 2001]B. Pumo and J. Le Corff. Parameter estimation of the contact process on a lattice. In G. Govaert, J. Janssen, and N. Limnios, editors, *ASMDA 2001 Proceddings*, pages 866–871, 2001.

[Whittle, 1954]P. Whittle. On stationary process in the plane. *Biometrika*, pages 434–449, 1954.