

**Supplementary material for:
“Inference of structure in subdivided populations at low
levels of genetic differentiation. The correlated allele
frequencies model revisited”.**

Gilles Guillot

Centre for Ecological and Evolutionary Synthesis Department of Biology, University of Oslo,
P.O Box 1066 Blindern, 0316 Oslo Norway.

Received on april 11 2008; revised on June 17 2008; accepted on August 7 2008

Associate Editor:Dr Alex Bateman

DERIVATION OF ACROSS-POPULATION CORRELATION OF ALLELE FREQUENCIES

I derive the expression of $Cor(f_{klj}, f_{k'lj})$ under the correlated model. I make use of the moments of a random vector x with a Dirichlet distribution $D(\lambda_1, \dots, \lambda_n)$: $E[x_i] = \lambda_i/\lambda_0$, $Var[x] = \frac{\lambda_i(\lambda_0 - \lambda_i)}{\lambda_0^2(\lambda_0 + 1)}$ and $E[x_i^2] = \frac{\lambda_i + \lambda_i^2}{\lambda_0^2 + \lambda_0}$.

$$E[f_{klj}] = E[E[f_{klj}|f_A, d]] = E[f_{Alj}] \quad (1)$$

and

$$\begin{aligned} E[f_{klj}f_{k'lj}] &= E[E[f_{klj}f_{k'lj}|f_A, d]] \\ &= E[E[f_{klj}|f_A, d]E[f_{k'lj}|f_A, d]] \\ &= E[f_{Alj}^2] \end{aligned} \quad (2)$$

(3)

hence

$$Cov(f_{klj}, f_{k'lj}) = E[f_{Alj}^2] - E[f_{Alj}]^2 = Var[f_{Alj}]. \quad (4)$$

The variance of f_{klj} involves its second order moment $E[f_{klj}^2]$.

Since

$$\begin{aligned} E[f_{klj}^2|f_A, d] &= \frac{f_{Alj}q_k + f_{Alj}^2q_k^2}{q_k + q_k^2} \quad \text{where } q_k = (1 - d_k)/d_k \\ &= E[f_{Alj}d_k + f_{Alj}^2(1 - d_k)] \end{aligned} \quad (5)$$

we get

$$E[f_{klj}^2] = E[f_{Alj}]E[d_k] + E[f_{Alj}^2]E[1 - d_k]. \quad (6)$$

Hence

$$\begin{aligned} Var[f_{klj}] &= E[f_{Alj}]E[d_k] + E[f_{Alj}^2]E[1 - d_k] - E[f_{Alj}]^2 \\ &= E[d_k](E[f_{Alj}] - E[f_{Alj}^2]) + Var[f_{Alj}] \end{aligned} \quad (7)$$

(8)

and

$$\begin{aligned} Cor(f_{klj}, f_{k'lj}) &= Cov(f_{klj}, f_{k'lj})/Var[f_{klj}] \\ &= \frac{Var[f_{Alj}]}{E[d_k](E[f_{Alj}] - E[f_{Alj}^2]) + Var[f_{Alj}]} \\ &= \frac{1}{1 + E[d_k] \frac{E[f_{Alj}] - E[f_{Alj}^2]}{E[f_{Alj}^2] - E[f_{Alj}]^2}} \end{aligned} \quad (9)$$

DETAIL OF MCMC COMPUTATIONS

Joint update of population memberships and allele frequencies

Attempting to make a move from $\theta = (K, p, d, f_A, f)$ to $\theta^* = (K, p^*, d, f_A, f^*)$, I propose a new state p^* from a distribution $q(p^*|p)$ (which I leave undefined at this step), and new frequencies f^* sampled from the full conditional $\pi(f|z, K, d, f_A, p^*)$ in the spirit of a Gibbs sampler. The Metropolis-Hastings ratio writes

$$\begin{aligned} R &= \frac{\pi(z|\theta^*)}{\pi(z|\theta)} \frac{\pi(p^*|K)}{\pi(p|K)} \frac{\pi(f^*|K, d, f_A)}{\pi(f|K, d, f_A)} \times \\ &\quad \frac{q(p|p^*)}{q(p^*|p)} \frac{q(f|f^*, p, K, d, f_A)}{q(f^*|f, p^*, K, d, f_A)} \\ &= \frac{\pi(p^*|K)}{\pi(p|K)} \frac{q(p|p^*)}{q(p^*|p)} \prod_{k,l} \frac{B(f_{Al} q_k + n_{kl}^*)}{B(f_{Al} q_k + n_{kl})} \end{aligned} \quad (10)$$

where B is the multinomial Beta function.

In particular, the frequencies f and f^* cancel out, thus the acceptance ratio does not depend on the state f^* proposed. Further simplification occur for symmetric proposal and/or particular choices of prior for p .

Split-merge of populations

Considering the case of the split of a population, a move from $\theta = (K, p, d, f_A, f)$ to $\theta^* = (K^* = K + 1, p^*, d^*, f_A, f^*)$ is proposed as follows: I propose a new state p^* from a distribution $q(p^*|p)$ in such way that the individuals of a randomly chosen population P_{k_0} are re-allocated into P_{k_0} and P_{K+1} . Drift parameters $d_{k_0}^*$ and d_{K+1}^* are proposed as $d_{k_0} - \delta_d$ and $d_{k_0} + \delta_d$ respectively, where δ_d is a small random increment. Frequencies f_{k_0} and f_{K+1} are proposed from the full conditional distribution $\pi(f|z, K^*, d^*, f_A, p^*)$.

The acceptance ratio is then

$$\begin{aligned} R &= \frac{\pi(z|\theta^*)}{\pi(z|\theta)} \times \\ &\quad \frac{\pi(p^*|K)}{\pi(p|K)} \frac{\pi(d^*|K)}{\pi(d|K)} \frac{\pi(f^*|K^*, d^*, f_A)}{\pi(f|K, d, f_A)} \times \\ &\quad \frac{q(p|K, p^*)}{q(p^*|K^*, p)} \frac{q(d|K, d^*)}{q(d^*|K^*, d)} \frac{q(f|f^*, p, K, d, f_A)}{q(f^*|f, p^*, K^*, d^*, f_A)} \end{aligned} \quad (11)$$

Again, the terms in f cancel out and I get

$$\begin{aligned} R &= \frac{\pi(p^*|K)}{\pi(p|K)} \frac{\pi(d^*|K)}{\pi(d|K)} \frac{q(p|K, p^*)}{q(p^*|K^*, p)} \frac{q(d|K, d^*)}{q(d^*|K^*, d)} \times \\ &\quad \prod_l \frac{\Gamma(q_{k_0}^*)}{\Gamma(n_{k_0l} + q_{k_0}^*)} \prod_j \frac{\Gamma(n_{k_0lj} + f_{Alj} q_{k_0}^*)}{\Gamma(f_{Alj} q_{k_0}^*)} \times \\ &\quad \prod_l \frac{\Gamma(q_{K+1}^*)}{\Gamma(n_{K+1l} + q_{K+1}^*)} \prod_j \frac{\Gamma(n_{K+1lj} + f_{Alj} q_{K+1}^*)}{\Gamma(f_{Alj} q_{K+1}^*)} \times \\ &\quad \left(\prod_l \frac{\Gamma(q_{k_0})}{\Gamma(n_{k_0l} + q_{k_0})} \prod_j \frac{\Gamma(n_{k_0lj} + f_{Alj} q_{k_0})}{\Gamma(f_{Alj} q_{k_0})} \right)^{-1} \end{aligned} \quad (12)$$

In my implementation, the random increment δ_d is centered and normally distributed with variance σ_d^2 . This choice gives better results in terms of mixing than a uniform proposal, as the reversibility constraint in a merge often leads to the rejection of a move. I get

$$\frac{q(d|K, d^*)}{q(d^*|K^*, d)} = \frac{2\sigma_d}{\exp(-\delta_d^2/2)/\sqrt{2\pi}} \quad (13)$$

And with Beta independent prior for d with common shape parameters a and b , I get

$$\frac{\pi(d^*|K)}{\pi(d|K)} = \frac{d_{k_0}^{*a-1} (1 - d_{k_0}^*)^{b-1} d_{K+1}^{*a-1} (1 - d_{K+1}^*)^{b-1} \Gamma(a+b)}{d_{k_0}^{a-1} (1 - d_{k_0})^{b-1} \Gamma(a)\Gamma(b)} \quad (14)$$

In all the numerical computations reported here, σ_d was set to $a/(a+b)$ where a and b are the parameters of prior Beta distribution of parameters d_k .

DETAIL OF THE SOLUTION TO THE LABEL SWITCHING ISSUE

-
- (i) From the whole MCMC output with variable number of populations $(\theta^{(t)})_t$ estimate K as $\hat{K} = \text{Argmax}_K \pi(K|data)$
 - (ii) From the whole MCMC output $(\theta^{(t)})_t$ extract the subset $(\tilde{\theta}^{(t)})_t$ of states where $K = \hat{K}$
 - (iii) On this extracted subset $(\tilde{\theta}^{(t)})_t$ compute the pivot defined as $\theta_{piv} = \text{Argmax}_{\theta \in (\tilde{\theta}^{(t)})_t} \pi(\theta|z)$
 - (iv) For each state $\theta^{(t)}$ in $(\tilde{\theta}^{(t)})_t$ find permutation τ_t that maximises the scalar product $\langle f_{piv}, f_{\tau_t}^{(t)} \rangle$
 - (v) From the relabeled subset $(\tilde{\theta}_{\tau_t}^{(t)})_t$, estimate assignments of population memberships by maximum a posteriori.
-

Table 1. Algorithm proposed to relabel populations and make assignments from a vector parameters $(\theta^{(t)})_t$ resulting from a single MCMC run. Note that this algorithm can also be used on a run resulting from the concatenation of several MCMC independent runs.

ILLUSTRATION OF IMPROVEMENTS

Simulations from the prior-likelihood model

	L = 10	L = 20	L = 50	L = 100
Non spatial simulation and inference				
CFM	0.23	0.19	0.13	0.09
UFM	0.24	0.23	0.21	0.2
Spatial simulation and inference				
CFM	0.15	0.11	0.08	0.06
UFM	0.19	0.18	0.18	0.17

Table 2. Accuracy of inference on simulated data as a function of the number of loci. The numbers given are the proportions of individuals not correctly assigned to their population of origin. First line: simulation and inference were performed assuming a non spatial model. Second line: simulation and inference assuming a spatial model. Genotypes were simulated from the correlated allele frequency model. The prior assumed for coefficients d_k is a Beta(2, 20). Each numerical value given is obtained as an average over a set of $N = 500$ datasets that covers a broad range of levels of differentiation. See figures below for details.

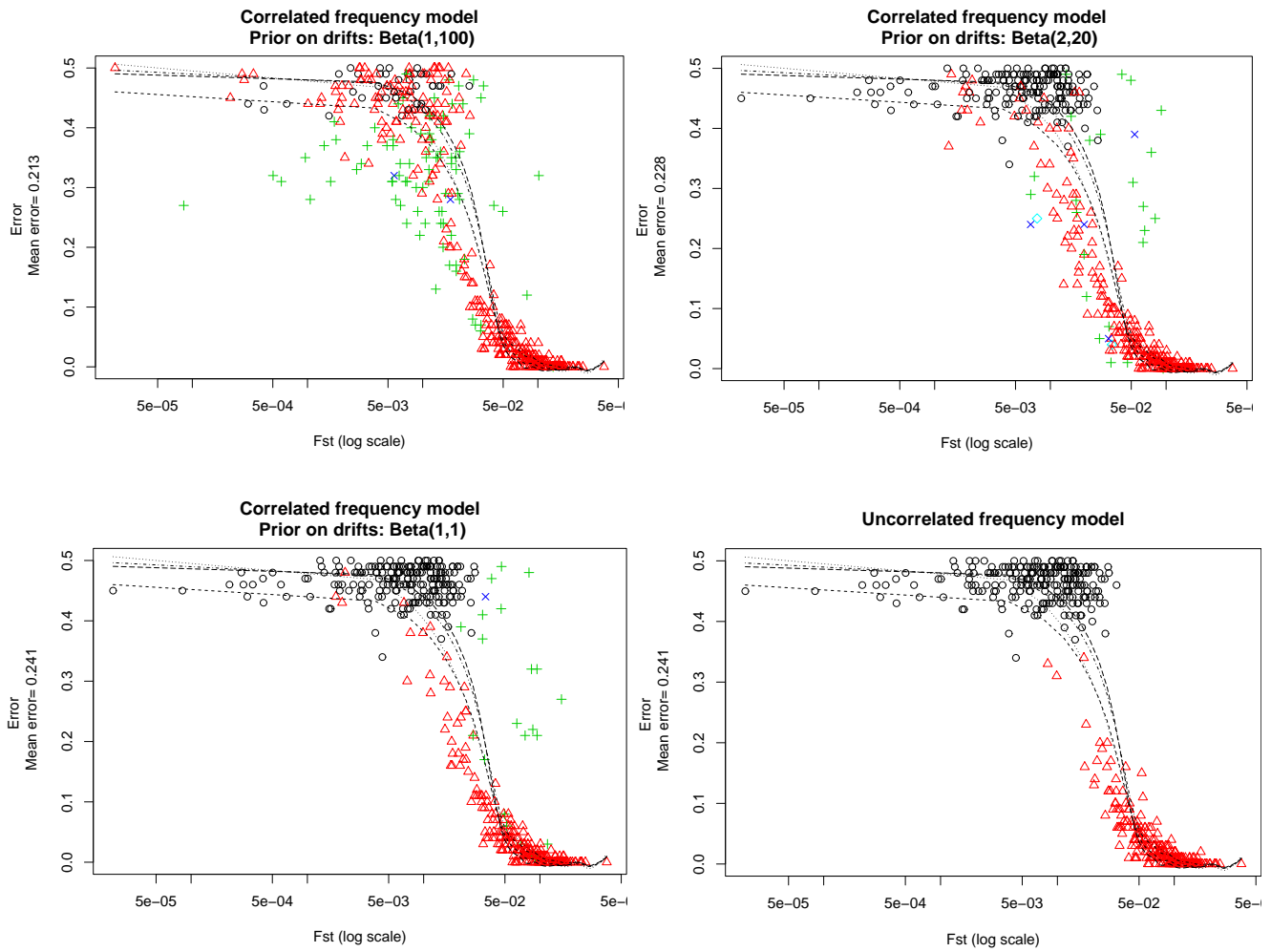


Fig. 1. Misassignment rates for $N = 500$ datasets simulated from the prior-likelihood model as a function of pairwise F_{ST} . Simulation and inference are carried out with a non-spatial prior. Each dataset consists of $n = 100$ individuals belonging to $K = 2$ populations with genotypes at $L = 10$ independent loci. The color and shape of the symbol stands for the number of populations inferred: \circ one population, \triangle two populations (correct result), $+$ three populations, \times four populations. The various dashed lines are non-parametric smoothing for the four clouds.

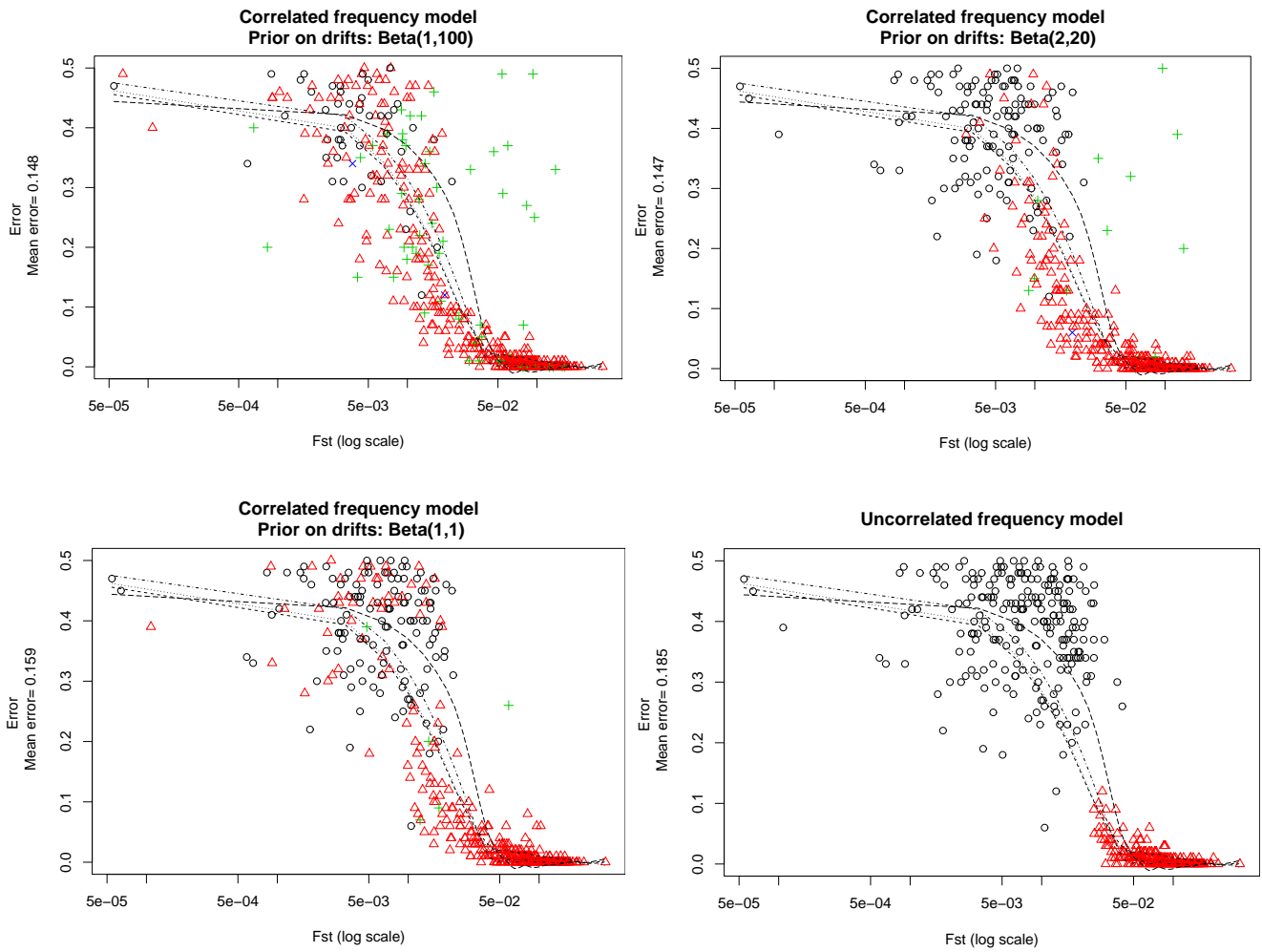


Fig. 2. Misassignment rates for $N = 500$ datasets simulated from the prior-likelihood model as a function of pairwise F_{ST} . Simulation and inference are carried out with a spatial prior. Each dataset consists of $n = 100$ individuals belonging to $K = 2$ populations with genotypes at $L = 10$ independent loci. The color and shape of the symbol stands for the number of populations inferred: \circ one population, \triangle two populations (correct result), $+$ three populations, \times four populations. The various dashed lines are non-parametric smoothing for the four clouds.

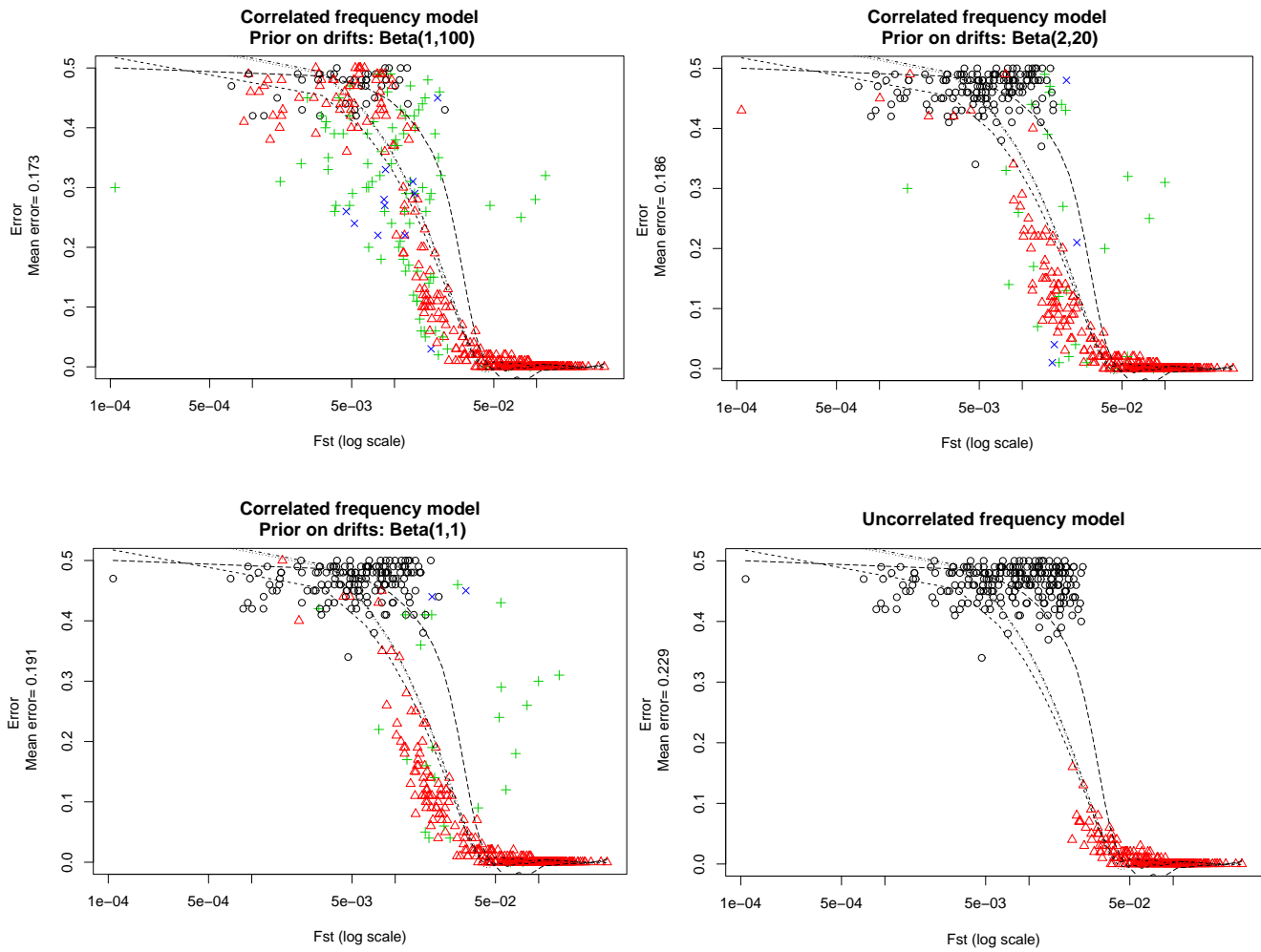


Fig. 3. Misassignment rates for $N = 500$ datasets simulated from the prior-likelihood model as a function of pairwise F_{ST} . Simulation and inference are carried out with a non-spatial prior. Each dataset consists of $n = 100$ individuals belonging to $K = 2$ populations with genotypes at $L = 20$ independent loci. The color and shape of the symbol stands for the number of populations inferred: ○ one population, △ two populations (correct result), + three populations, × four populations. The various dashed lines are non-parametric smoothing for the four clouds.

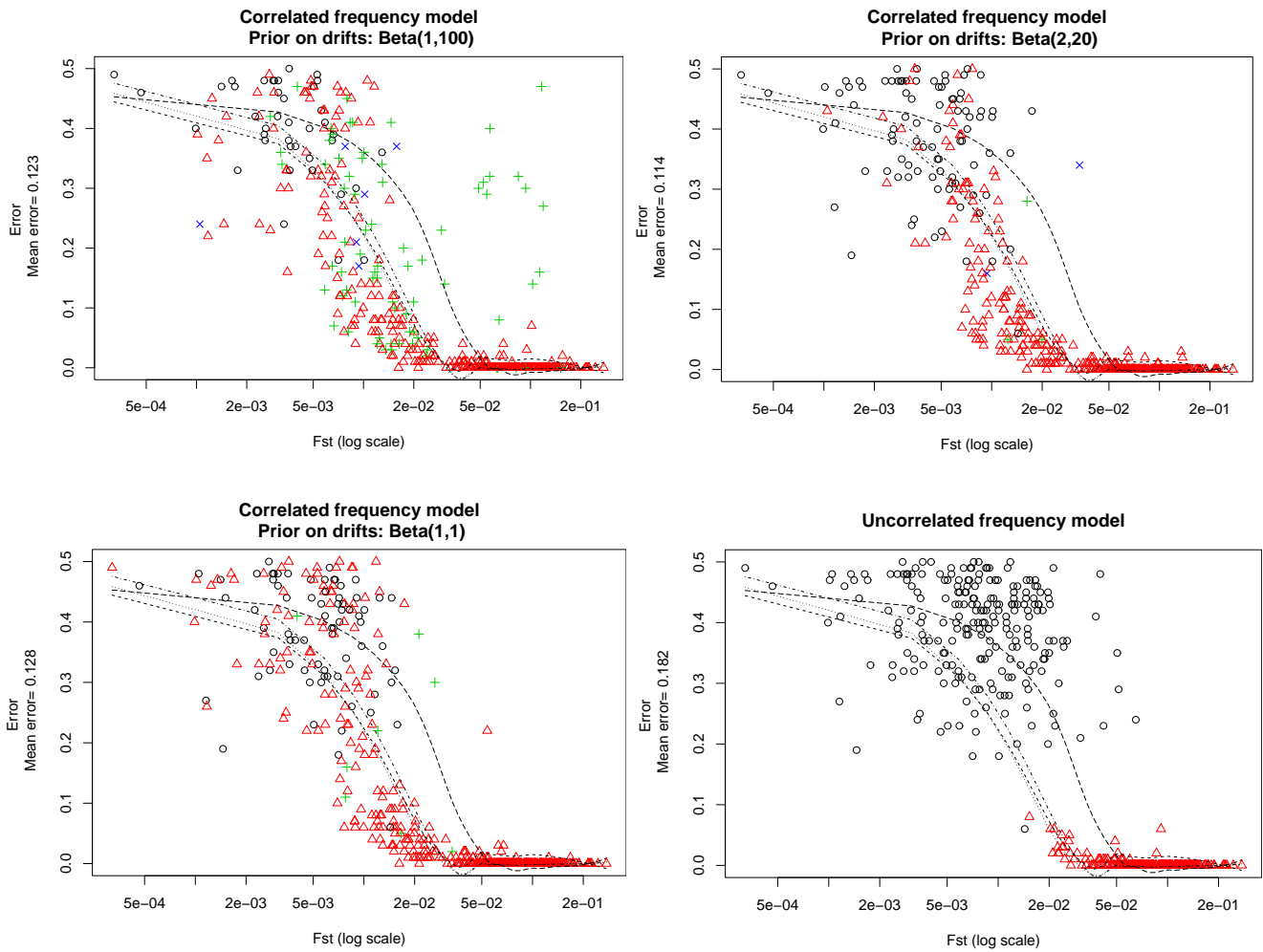


Fig. 4. Misassignment rates for $N = 500$ datasets simulated from the prior-likelihood model as a function of pairwise F_{ST} . Simulation and inference are carried out with a spatial prior. Each dataset consists of $n = 100$ individuals belonging to $K = 2$ populations with genotypes at $L = 20$ independent loci. The color and shape of the symbol stands for the number of populations inferred: \circ one population, \triangle two populations (correct result), $+$ three populations, \times four populations. The various dashed lines are non-parametric smoothing for the four clouds.

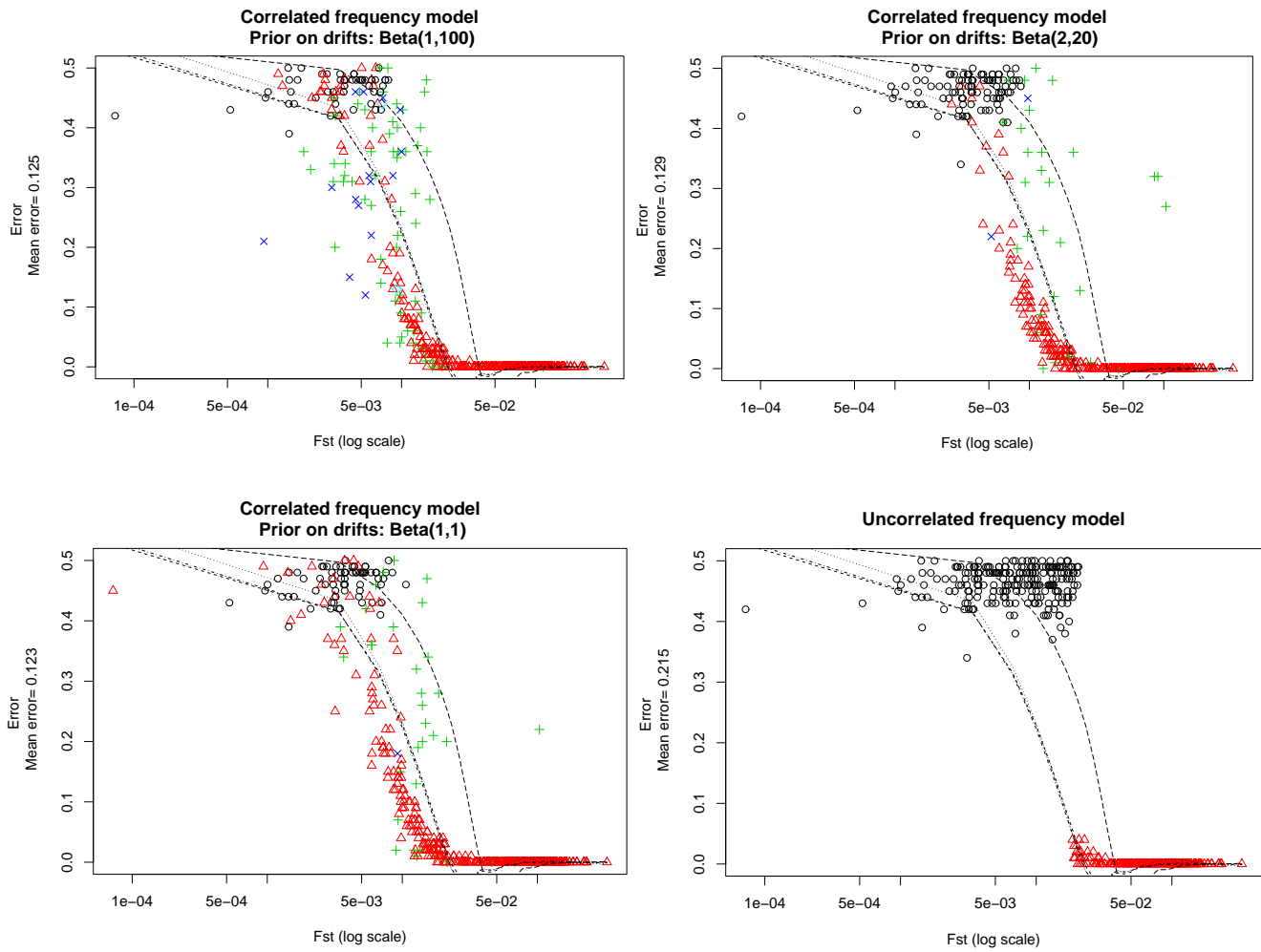


Fig. 5. Misassignment rates for $N = 500$ datasets simulated from the prior-likelihood model as a function of pairwise F_{ST} . Simulation and inference are carried out with a non-spatial prior. Each dataset consists of $n = 100$ individuals belonging to $K = 2$ populations with genotypes at $L = 50$ independent loci. The color and shape of the symbol stands for the number of populations inferred: \circ one population, \triangle two populations (correct result), $+$ three populations, \times four populations. The various dashed lines are non-parametric smoothing for the four clouds.

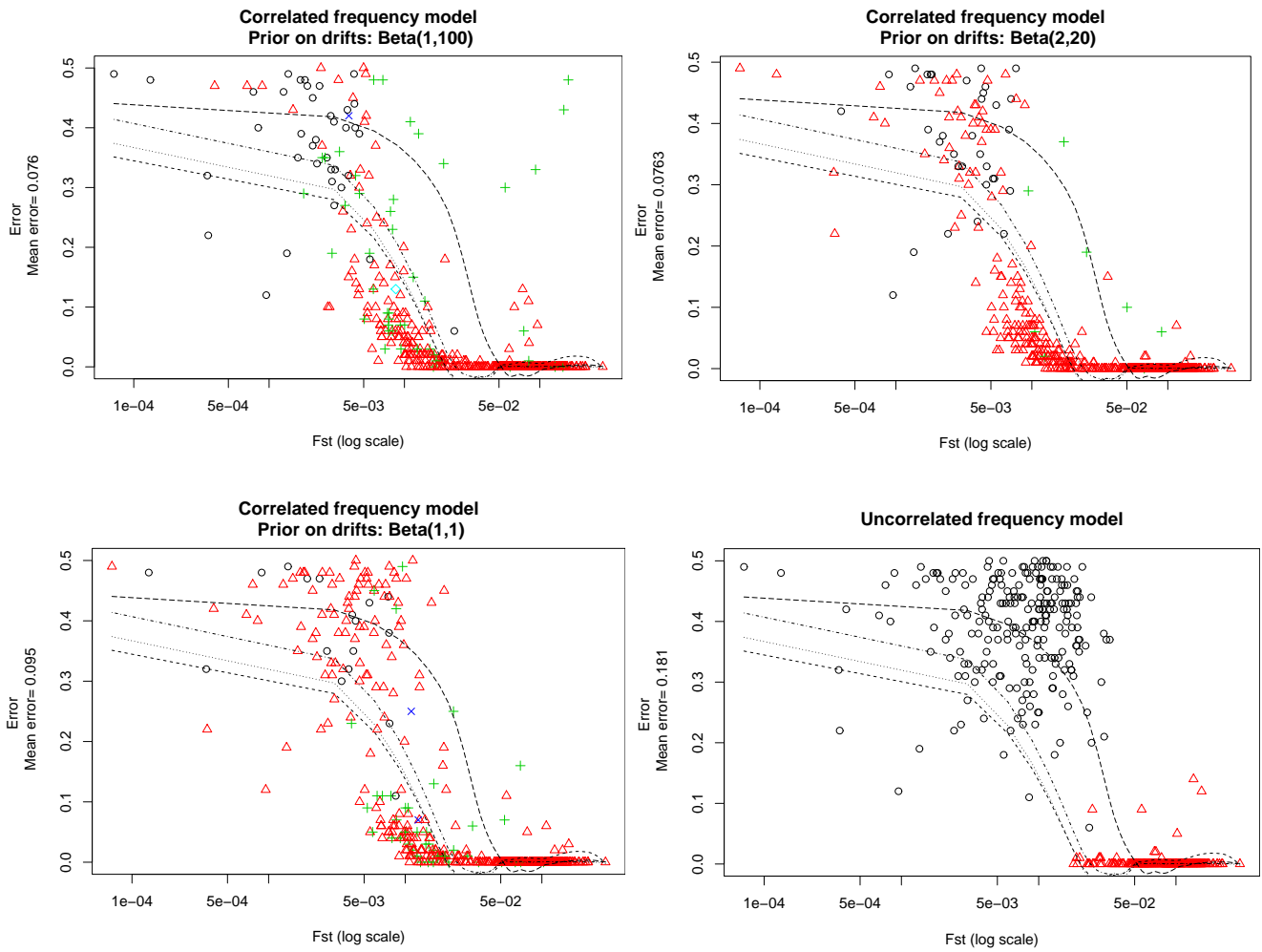


Fig. 6. Misassignment rates for $N = 500$ datasets simulated from the prior-likelihood model as a function of pairwise F_{ST} . Simulation and inference are carried out with a spatial prior. Each dataset consists of $n = 100$ individuals belonging to $K = 2$ populations with genotypes at $L = 50$ independent loci. The color and shape of the symbol stands for the number of populations inferred: ○ one population, △ two populations (correct result), + three populations, × four populations. The various dashed lines are non-parametric smoothing for the four clouds.

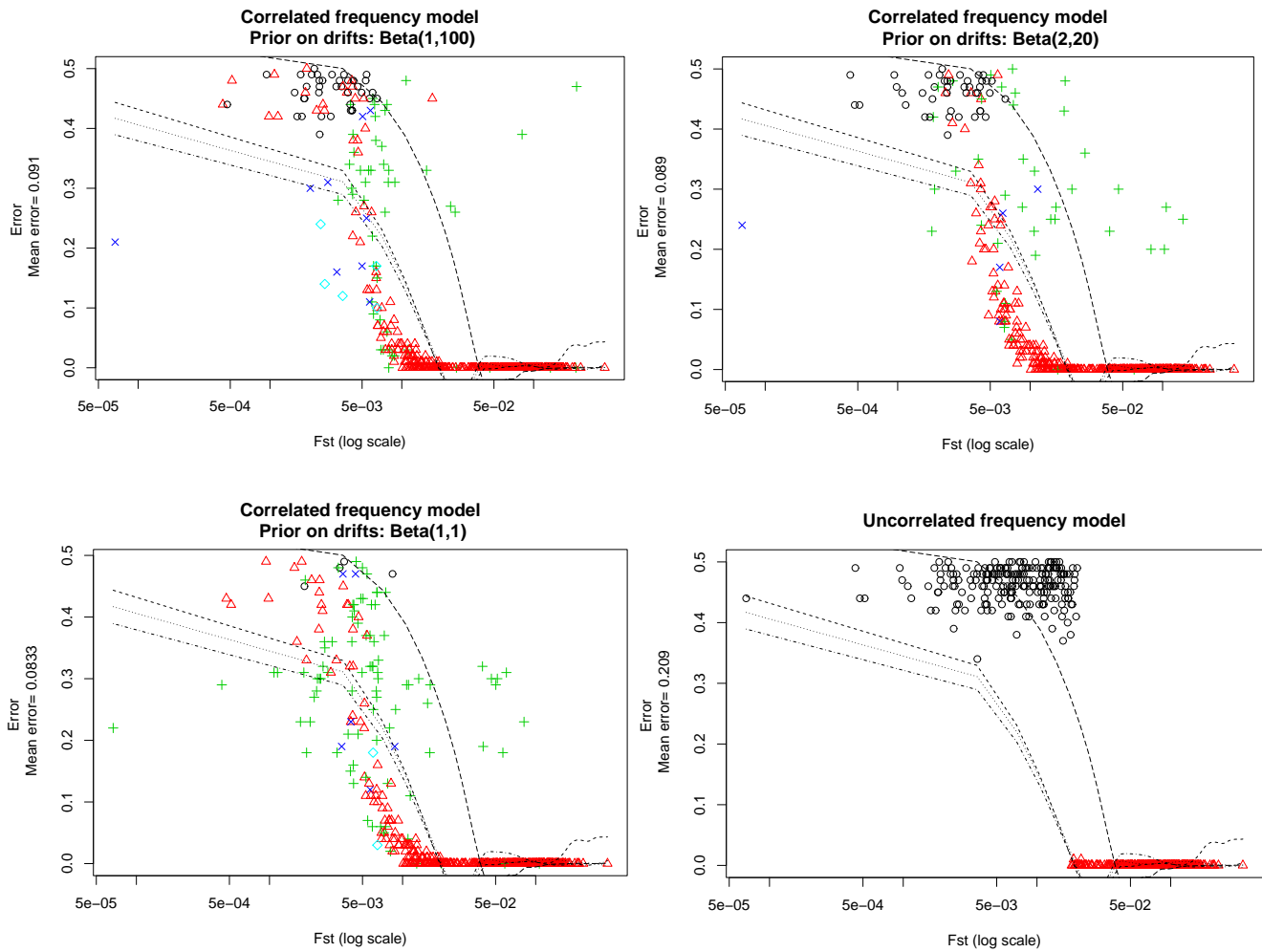


Fig. 7. Misassignment rates for $N = 500$ datasets simulated from the prior-likelihood model as a function of pairwise F_{ST} . Simulation and inference are carried out with a non-spatial prior. Each dataset consists of $n = 100$ individuals belonging to $K = 2$ populations with genotypes at $L = 100$ independent loci. The various dashed lines are non-parametric smoothing for the four clouds. The color and shape of the symbol stands for the number of populations inferred: ○ one population, △ two populations (correct result), + three populations, × four populations.

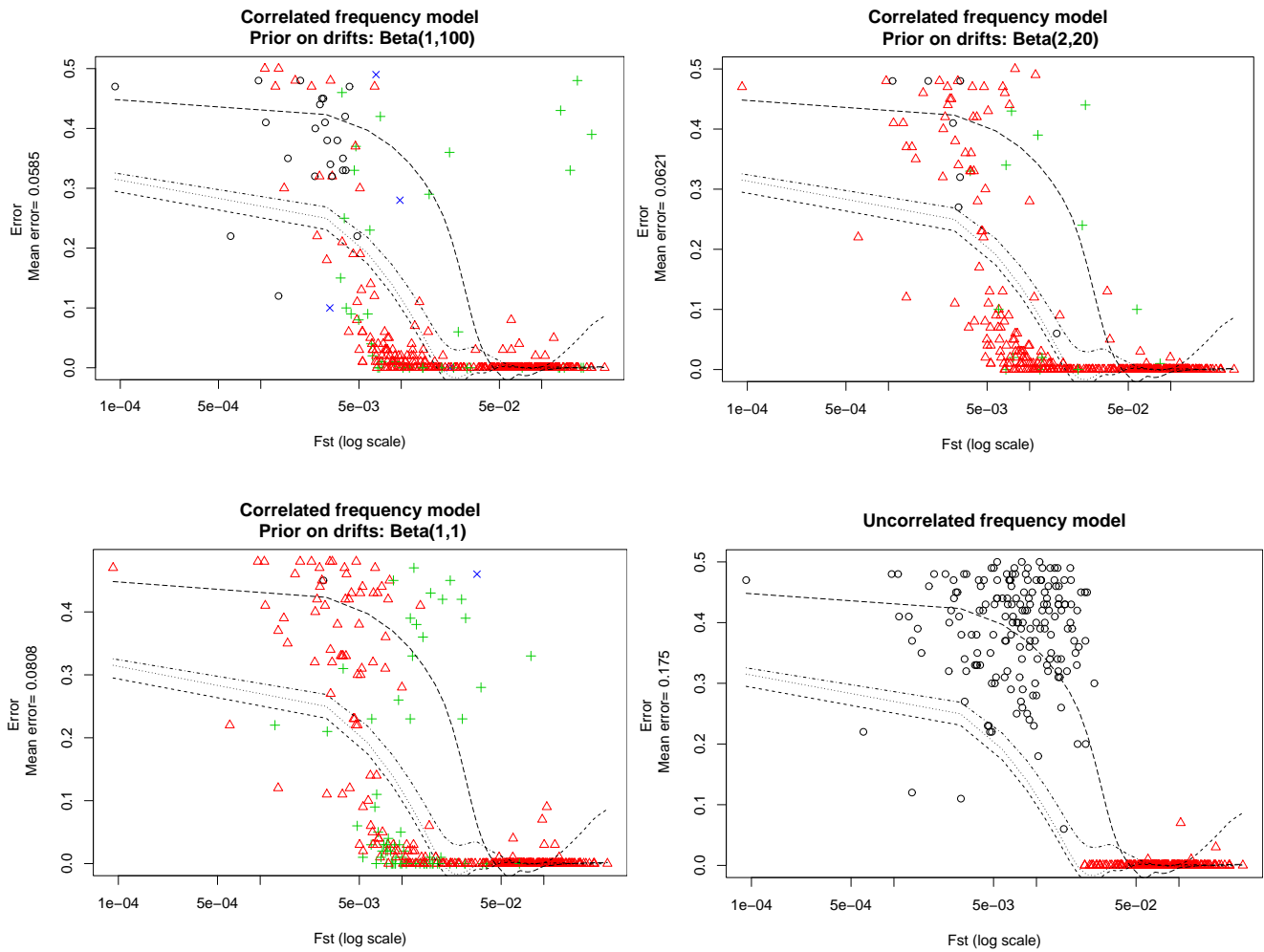


Fig. 8. Misassignment rates for $N = 500$ datasets simulated from the prior-likelihood model as a function of pairwise F_{ST} . Simulation and inference are carried out with a spatial prior. Each dataset consists of $n = 100$ individuals belonging to $K = 2$ populations with genotypes at $L = 100$ independent loci. The various dashed lines are non-parametric smoothing for the four clouds. The color and shape of the symbol stands for the number of populations inferred: \circ one population, \triangle two populations (correct result), $+$ three populations, \times four populations.

Simulations from a Wright-Fisher neutral model

M	θ	Bias				Misassignment rate			
		Prior on allele frequencies				Prior on allele frequencies			
		Low	Medium	Flat	Uncor.	Low	Medium	Flat	Uncor.
0.5	0.5	0	0	0	0	0	0	0	0
0.5	1	0	0	0	0	0	0	0	0
0.5	2	0	0	0	0	0	0	0	0
0.5	4	0	0	0.03	0	0	0	0.005	0
1	0.5	0	0	0	0	0.001	0.001	0.001	0.001
1	1	0	0	0	0	0	0	0	0
1	2	0	0	0.01	0	0	0	0	0
1	4	0	0	0.02	0	0	0	0.01	0
2	0.5	0	0.02	0.02	-0.01	0.008	0.008	0.008	0.013
2	1	0	0	0	0	0.001	0.001	0.001	0.001
2	2	0	0	0.01	0	0	0	0.005	0
2	4	0	0	0.01	0	0	0	0	0
4	0.5	0	0.14	0.16	-0.1	0.034	0.053	0.054	0.077
4	1	0	0.01	0.05	0	0.007	0.012	0.012	0.008
4	2	0.01	0	0.03	0	0.001	0.001	0.002	0.002
4	4	0	0	0.01	0	0	0	0	0
10	0.5	0.01	0.36	0.19	-0.85	0.183	0.186	0.218	0.436
10	1	0.09	0.48	0.43	-0.52	0.095	0.121	0.133	0.285
10	2	0.06	0.64	0.64	-0.07	0.04	0.15	0.132	0.06
10	4	0.02	0.56	0.65	0	0.013	0.083	0.079	0.006
20	0.5	-0.18	-0.18	-0.45	-1	0.414	0.38	0.409	0.5
20	1	0.3	0.44	0.1	-1	0.31	0.262	0.324	0.5
20	2	0.45	0.71	0.65	-0.95	0.216	0.221	0.213	0.479
20	4	0.63	1.06	1.13	-0.36	0.166	0.223	0.21	0.208
40	0.5	-0.21	-0.43	-0.7	-1	0.45	0.452	0.474	0.5
40	1	0.11	-0.17	-0.38	-1	0.421	0.415	0.432	0.5
40	2	0.4	0.19	-0.12	-1	0.381	0.372	0.382	0.5
40	4	0.64	0.47	0.62	-1	0.393	0.338	0.308	0.5

Table 3. Accuracy of inferences on data simulated according to a Wright-Fisher model. Each value of the table is estimated from $N = 100$ independently simulated datasets consisting of $n = 100$ individuals belonging to $K = 2$ populations with genotypes at $L = 10$ unlinked loci, and analyzed with four different methods (columns). Simulations and inferences are based on a non-spatial model.

M	θ	Bias				Misassignment rate			
		Prior on allele frequencies				Prior on allele frequencies			
		Low	Medium	Flat	Uncor.	Low	Medium	Flat	Uncor.
0.5	0.5	0.07	0.02	0.02	0	0.025	0	0.005	0
0.5	1	0.06	0	0	0	0.015	0	0	0
0.5	2	0	0	0.05	0	0	0	0.005	0
0.5	4	0	0	0.26	0	0	0	0.04	0
1	0.5	0.01	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0
1	2	0	0	0.06	0	0	0	0.005	0
1	4	0	0	0.3	0	0	0	0.04	0
2	0.5	0.03	0.03	0.03	0.01	0.01	0.005	0	0.005
2	1	0	0	0.04	0	0	0	0.005	0
2	2	0	0	0.16	0	0	0	0.035	0
2	4	0	0	0.47	0	0	0	0.1	0
4	0.5	0.01	0.04	0.05	-0.01	0.012	0.011	0.015	0.012
4	1	0	0	0.05	0	0	0	0	0
4	2	0	0	0.17	0	0	0	0.03	0
4	4	0	0	0.55	0	0	0	0.145	0
10	0.5	0	0.38	0.41	-0.7	0.065	0.12	0.113	0.362
10	1	0.04	0.3	0.37	-0.22	0.021	0.055	0.059	0.119
10	2	0	0.08	0.08	0	0.002	0.016	0.015	0.002
10	4	0	0.03	0.23	0	0	0.005	0.048	0
20	0.5	-0.01	0.42	0.3	-1	0.247	0.225	0.223	0.5
20	1	0.19	0.82	0.8	-1	0.122	0.188	0.203	0.5
20	2	0.22	1.02	1.04	-0.89	0.068	0.148	0.159	0.447
20	4	0.14	0.96	1.11	-0.04	0.035	0.138	0.114	0.026
40	0.5	-0.13	-0.05	-0.36	-1	0.421	0.395	0.422	0.5
40	1	0.45	0.31	0.35	-1	0.35	0.319	0.339	0.5
40	2	0.78	0.78	0.97	-1	0.285	0.284	0.275	0.5
40	4	0.8	1.28	1.71	-0.99	0.303	0.242	0.234	0.495

Table 4. Accuracy of inferences on data simulated according to a Wright-Fisher model. Each value of the table is estimated from $N = 100$ independently simulated datasets consisting of $n = 100$ individuals belonging to $K = 2$ populations with genotypes at $L = 20$ unlinked loci. Simulations and inferences are based on a non-spatial model.

M	θ	Bias				Misassignment rate			
		Prior on allele frequencies				Prior on allele frequencies			
		Low	Medium	Flat	Uncor.	Low	Medium	Flat	Uncor.
0.5	0.5	0.55	0.1	0.11	0.05	0.13	0.016	0.01	0.001
0.5	1	0.62	0	0.01	0	0.135	0	0.005	0
0.5	2	0.36	0	0.17	0	0.07	0	0.035	0
0.5	4	0.1	0.01	0.66	0	0.02	0	0.13	0
1	0.5	0.23	0.09	0.11	0.04	0.06	0.021	0	0
1	1	0.23	0	0.08	0	0.06	0	0.015	0
1	2	0.07	0	0.21	0	0.03	0	0.045	0
1	4	0	0	0.84	0	0	0	0.17	0
2	0.5	0.14	0.12	0.17	0.08	0.04	0.01	0.035	0.02
2	1	0.02	0	0.14	0	0.005	0	0.03	0
2	2	0	0	0.51	0	0	0	0.115	0
2	4	0	0.02	0.95	0	0	0	0.22	0
4	0.5	0.1	0.11	0.19	0.04	0.02	0.016	0.03	0.006
4	1	0	0.01	0.29	0	0	0	0.05	0
4	2	0	0.02	0.74	0	0	0	0.19	0
4	4	0	0.05	0.99	0	0	0	0.225	0
10	0.5	0.09	0.13	0.26	-0.4	0.049	0.015	0.043	0.236
10	1	0	0	0.29	-0.02	0	0	0.061	0.01
10	2	0	0.02	0.84	-0.01	0	0.005	0.215	0.005
10	4	0	0.25	1	0	0	0.05	0.27	0
20	0.5	-0.49	0.53	0.64	-0.92	0.324	0.115	0.117	0.499
20	1	0.05	0.27	0.77	-1	0.024	0.056	0.073	0.5
20	2	0	0.04	0.59	-0.78	0.006	0.009	0.132	0.39
20	4	0	0.2	0.96	-0.03	0	0.049	0.214	0.015
40	0.5	-0.44	0.84	0.91	-0.89	0.421	0.269	0.256	0.499
40	1	0.46	1.15	1.37	-1	0.164	0.218	0.22	0.5
40	2	0.32	1.33	1.83	-1	0.117	0.193	0.172	0.5
40	4	0.1	0.56	0.96	-1	0.036	0.118	0.22	0.5

Table 5. Accuracy of inferences on data simulated according to a Wright-Fisher model. Each value of the table is estimated from $N = 100$ independently simulated datasets consisting of $n = 100$ individuals belonging to $K = 2$ populations with genotypes at $L = 50$ unlinked loci, and analyzed with four different methods (columns). Simulations and inferences are based on a non-spatial model.

M	θ	Bias				Misassignment rate			
		Prior on allele frequencies				Prior on allele frequencies			
		Low	Medium	Flat	Uncor.	Low	Medium	Flat	Uncor.
0.5	0.5	0.88	0.15	0.16	0.1	0.206	0.021	0.025	0.01
0.5	1	0.86	0.01	0.05	0	0.2	0	0.01	0
0.5	2	0.64	0.08	0.31	0	0.115	0.005	0.05	0
0.5	4	0.36	0.01	0.86	0	0.09	0.005	0.21	0
1	0.5	0.77	0.23	0.27	0.22	0.18	0.051	0.037	0.035
1	1	0.81	0	0.13	0	0.235	0	0.02	0
1	2	0.39	0.03	0.57	0	0.12	0	0.085	0
1	4	0.18	0.05	0.95	0	0.045	0	0.265	0
2	0.5	0.41	0.29	0.37	0.12	0.071	0.05	0.069	0.019
2	1	0.33	0	0.33	0	0.075	0	0.09	0
2	2	0.14	0.03	0.76	0	0.04	0.005	0.2	0
2	4	0.05	0.07	0.99	0	0.005	0.005	0.215	0
4	0.5	0.3	0.31	0.51	0.18	0.046	0.05	0.07	0.03
4	1	0	0	0.65	0	0	0	0.15	0
4	2	0.02	0.01	0.94	0	0	0	0.2	0
4	4	0	0.27	1	0	0	0.025	0.245	0
10	0.5	0.24	0.32	0.6	-0.17	0.069	0.075	0.115	0.226
10	1	0	0.03	0.88	0	0	0.01	0.185	0
10	2	0	0.18	0.99	0	0	0.03	0.28	0
10	4	0	0.75	1	0	0	0.145	0.25	0
20	0.5	-0.46	0.3	0.53	-0.76	0.376	0.077	0.124	0.497
20	1	-0.01	-0.01	0.64	-1	0.005	0.005	0.176	0.5
20	2	0	0.37	0.99	-0.83	0	0.065	0.24	0.415
20	4	0	0.89	1	-0.02	0	0.195	0.255	0.01
40	0.5	-0.61	1.08	1.15	-0.76	0.478	0.223	0.254	0.497
40	1	0.06	0.88	1.48	-1	0.062	0.155	0.168	0.5
40	2	0.01	0.26	0.95	-1	0.019	0.06	0.259	0.5
40	4	0	0.79	0.87	-1	0.001	0.162	0.24	0.5

Table 6. Accuracy of inferences on data simulated according to a Wright-Fisher model. Each value of the table is estimated from $N = 100$ independently simulated datasets consisting of $n = 100$ individuals belonging to $K = 2$ populations with genotypes at $L = 100$ unlinked loci, and analysed with four different methods (columns). Simulations and inferences are based on a non-spatial model.

ANALYSIS OF REAL DATA

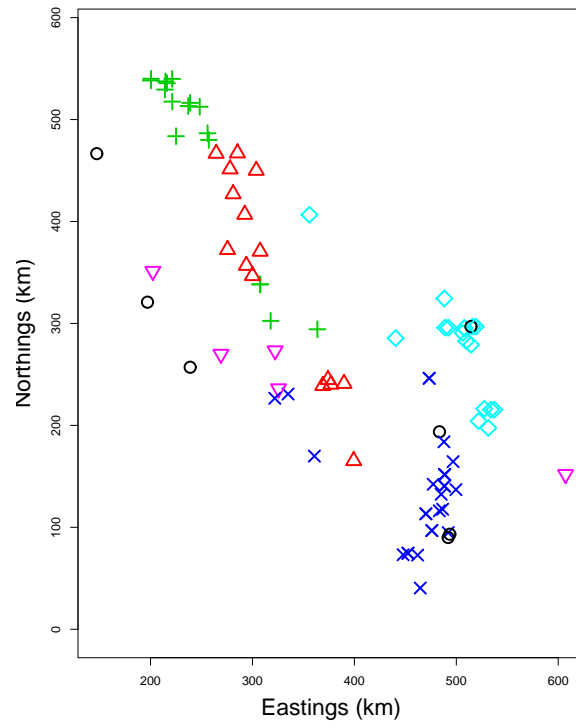


Fig. 9. Spatial spread of the six inferred wolverines sub-populations. Color and shape of the symbols refer to the for inferred population label: ○ population one, △ population two, + population three × population four, ◇ population five, ▽ population six.

Pop. label	1	2	3	4	5	6
		0.12	0.15	0.15	0.15	0.12
			0.073	0.12	0.16	0.063
				0.19	0.17	0.11
					0.15	0.2
						0.16
	-0.097	0.037	0.068	0.082	0.13	0.13

Table 7. Estimated F statistics for the six inferred wolverines sub-populations. Lines 1-5. pairwise F_{ST} , bottom line F_{IS} .