Markov Chain Monte Carlo Sampling on Polar Sea Ice Classification

Shuyi Li
Microwave Earth Remote Sensing Laboratory
Brigham Young University
Provo, UT 84602
PH: 801.422.4884, FAX: 801.422.6586
April 15, 2004

ABSTRACT

The Remund-Long (RL) Multi-Sensor Sea Ice Classification algorithm[1] combines both radiometer and scatterometer data using Principle Component Analysis and reduces the dimensionality and noise level of the data. The algorithm uses an iterative Maximum a Posteriori (MAP) method based on a multi-variant Gaussian model with a temporal prior. As a result, the algorithm successfully classifies Winter Antarctic region into five different ice types. However, due to the nature of this pixel wise classification algorithm, the final classification is more likely to be corrupted by slat-pepper-shaped artifacts. Such artifacts are introduced by the Scatterometer Image Reconstruction (SIR) algorithm[2] which utilizes multi-swath raw scatterometer data to generate high resolution images. In order to resolve such problem in RL algorithm, posterior distribution function with spatial prior is embedded into the classification process. A Markov Chain Monte Carlo (MCMC) sampling method is one way to sample such posterior distribution of the state space in which each element of the space has the size of an entire image. This report gives a brief introduction to the concept of Metropolis-Hastings Markov Chain Monte Carlo (MH MCMC) algorithm, discusses its implementation on polar sea ice classification, and compares the result with the RL algorithm.

SYMBOLS

C = Ice class vector
Z = 9 dimensional observation images array
Z' = reduced dimension eigen images array after PCA
M = Markov chain state space.
X = ice class image

INTRODUCTION

Polar sea ice 1 plays many important roles in the global climate. It serves as an effective insulator between the ocean and the atmosphere, restricting exchanges of heat, mass, momentum, and chemical constituents.[3] During the Winter season, the heat transfered from the warm ocean body to the relatively cold atmosphere is limited by the open ocean area and thin ice pack. The heat flux from open ocean exceeds the heat flux through an adjacent thick ice by two orders of magnitude. The distribution of the sea ice as a result plays important role on the regional heat balance. The unique characteristics of microwave remote sensing provide numerous advantages over conventional sensors such as optical and infrared (IR) sensors. Its immunity to the lack of sun light and atmosphere turbulence make microwave remote sensing (MRS) an ideal way to observe regions that have unfriendly environment conditions to conduct *in-situ* measurements.

Figure 1: Satellite Picture of Antarctica

RL ALGORITHM

There are two types of spaceborne MRS instruments, scatterometers (NSCAT, QSCAT, ERS) and radiometers (SSM/I). Scatterometer is a type of imaging RADAR that sends microwave pulses and measures the normalized RADAR cross section σ^0, while a radiometer is a instrument that scans the brightness temperature, T_b, of the Earth surface. Brightness temperature is the physical temperature of a perfect black body that would emit the same amount of radiation as observed region at given frequency and polarization. A perfect back body emits 100 percent of energy it has absorbed. T_b of a perfect black body equals its physical temperature. In the following analysis the total 9-channel (7-channel SSM/I, v-pol and

h-pol QSCAT) of resolution enhanced satellite images, Z, are used to do daily classification. Z is denoted as

$$Z = [z_1 z_2 z_3 ... z_9]^T \quad (1)$$

where Z_i is the vectorized the image. The data is been normalized to zero mean unite-variance according to sensor type. The principle component method reduces the dimensionality of Z from nine to three where 90% of information is preserved and noise is significantly reduced. The new set of image array is denoted Z'.

$$Z' = [v_1 v_2 v_3]^T Z \quad (2)$$

where v_1, v_2, v_3 are eigenvectors corresponding to the first three largest eigenvalues of auto-covariance matrix R_Z, defined as

$$\begin{aligned} R_Z &= E\{ZZ^T\} && (3)\\ &= V\Lambda V^T && (4)\\ &= \sum_{i=1}^{12} \lambda_i v_i v_i^T && (5) \end{aligned}$$

V, and Λ are eigenvector matrix and diagonalized eigenvalues. As seen in Figure 2 that majority of information is concentrated in eigen images 9, 8, 7, which corresponds to the first 3 biggest eigenvalues.

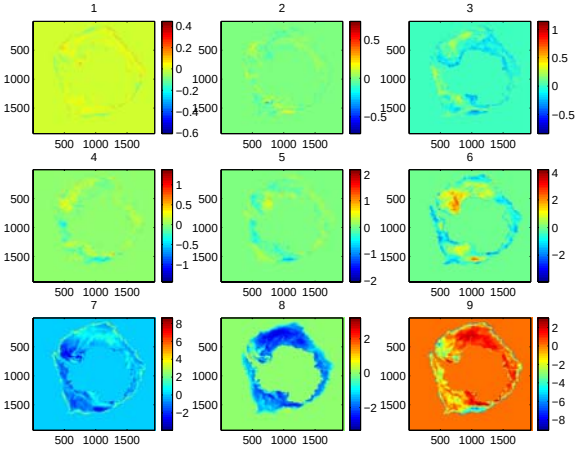

Figure 2: Projected eigen images using principle component analysis

There are five ice types exist in the winter Antarctic region: *smooth first year*, *rough first year*, *perennial*, *pancake*, and *marginal* ice. The fallowing discussing is based on the general behavior of scattering and emission from each type of ice. Under the assumption of a multivariant Gaussian model for given ice type c_i, the probability density function can be expressed as

$$P(Z'|c_i) = \frac{1}{\sqrt{2\pi}|K_{c_i}|^{\frac{1}{2}}} e^{-\frac{1}{2}(Z'-\mu_{c_i})^T K_{c_i}{}^{-1}(Z'-\mu_{c_i})} \quad (6)$$

A maximum a Posteriori (MAP) classifier is implemented on Z' with temporal prior distribution $P(C)$.

$$\begin{aligned} C_{MAP} &= argmax_{C_i}\{P(C|Z')\} && (7)\\ &= argmax_{C_i}\{\frac{P(Z'|C)P(C_i)}{P(Z')}\} && (8) \end{aligned}$$

An iterative method is carried on the MAP algorithm. The initial estimation of the mean μ_i is obtained by taking the average of a homogeneous region for each ice type is selected based on the *in situ* observation.[4] Since K_i is harder to find initially, a nearest neighborhood classification is implemented. K_i, in return, can be calculated based the result of the first iteration; μ_i is also updated based on previous classification. The initial prior probability $P(C_i)$ is solely found based on intuition and previous study on the Antarctica sea ice. For example, perennial ice is more rare than pancake ice, while smooth first year is more abundant than marginal ice zone. Each prior probability $P(c_i)$ is estimated based on the classified data at every iteration and fed back into the next iteration. The classified sea ice image of Julian day 251 of 2002 after 15th iteration is shown in Figure 3.

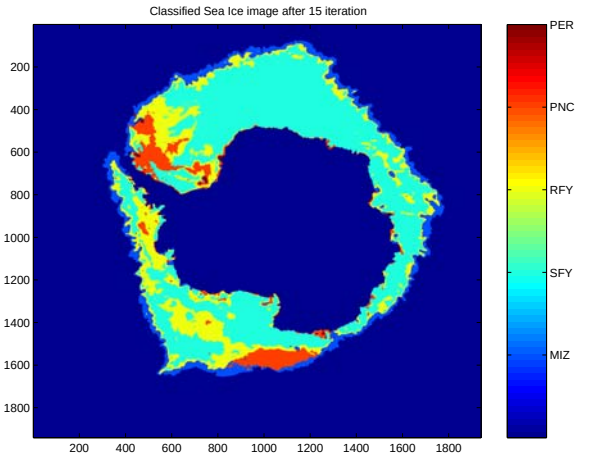

Figure 3: Classified Antarctic sea ice using MAP algorithm

The algorithm converges in terms of the norm of auto-covariance matrix K_i, shown in the Figure 4.

Without utilizing any prior information, a maximum likelihood (ML) classification can be implemented. Figure 5 shows the classified ice type result using ML algorithm. As we can see that the perennial ice zone is mis-classified into much larger region in northwest and

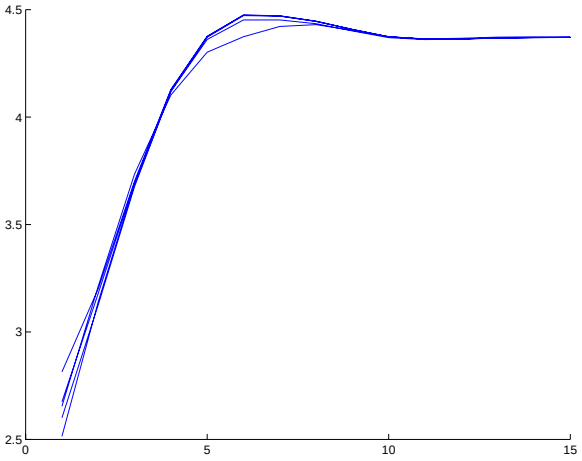

Figure 4: Norm of the K_{c_i} V.S. number of iterations

south area, since there are hardly any sea ice can survive through warn weather in Summer Antarctica.

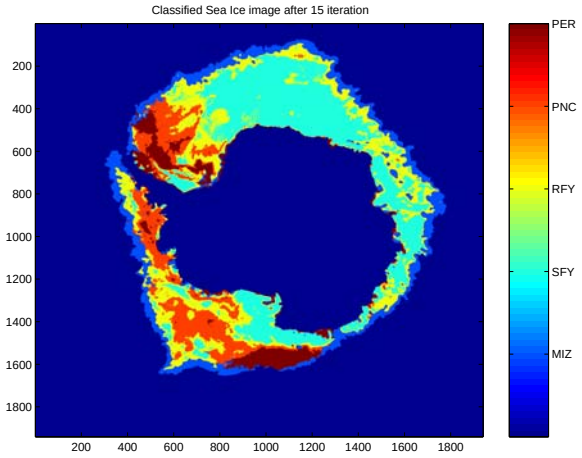

Figure 5: Classified Antarctic sea ice using ML algorithm

MH-MCMC METHOD

The temporal prior information imposed in RL iterative MAP algorithm has shown great improvement over ML algorithm, however heavy salt-and-pepper-shaped artifact can be found around northern Antarctica region using both algorithms. Comparing the zoomed in versions of this particular region between the MAP result and original SIR image, Figure 6 shows the artifact in both images shares similar structure. Thus the artifact in the final classified result is introduced by the SIR algorithm used to generate high resolution source images.

There are two possible ways to remove such defect from the final classified image: First, median filtering on the final classified image preserves the mean and variance, while reducing salt and pepper noise. The secondly approach is to use a *blob* shape spatial constrain as the prior distribution $P(C)$ in MAP algorithm. To impose such spatial prior requires classifying the image as a whole. First we start with the minimum mean squared error Bayesian estimation for X is $E\left[P(X|Z')\right]$ where X is the ice class image (instead of single-pixel ice class C). By Bayes' rule, we can express the posterior distribution as

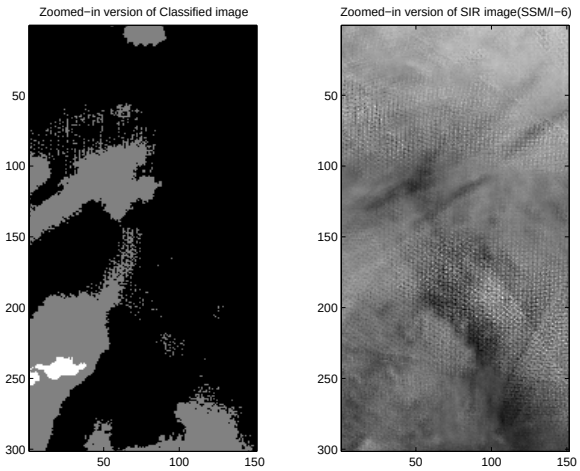

Figure 6: Artifact comparison between classified image and SIR image

$$P(X|Z') \propto P(Z'|X)P(X) \tag{9}$$

In many cases, the full posterior probability distribution is analytically intractable because the large number of possible combination of X. For instance: a typical high resolution microwave backscatter SIR image is 1940X1940 and there are 5 possible ice classes for each pixel, then the size of the state space, Ω, is 1940^{10}. And also the spatial prior distribution $P(X)$ may involve information that is difficult to express in analytical way, for example the shape of particular ice type may be described as a "blob" based on the assumption that the ice type of a given pixel shares similar ice type with its neighboring pixels. Such problem is analytically intractable and may even seem like unsolvable, however it can be treated as a simulation process: N samples can be drawn from state space Ω and each sample is obtained by simulating the physical process from Z' to X with probability proximately $P(X|Z')$. As a result, a sampled posterior distribution is acquired based the sub state space $\Theta = \{X_1, X_2 ... X_N\}$. With posterior distribution in hand, mean, mode even higher order statistics of X can be calculated. Such sampling-based method is called Monte-Carlo method. A first-order Markov Chain model applied with Monte-Carlo method is discussed and implemented on polar sea ice classification in this report.

A *Markov chain* of a sequence of correlated random variable denoted as $M = \{X_n, n = 0 \to \infty\}$, which is called *state space* and denoted as Ω satisfies *Markov condition*

$$\begin{aligned} &P(X_{n+1} = x_{n+1}|X_n = x_n, ..., X_0 = x_0) \\ &= P(X_{n+1} = x_{n+1}|X_n = x_n) \end{aligned} \tag{10}$$

As see in the equation that the distribution of random variable X_{n+1} is only depends on previous one X_n. So a transitional probability for X_{n+1} given X_n is defined as $P(X_{n+1}|X_n)$. If the transitional probability is independent with time index n when fixing the initial distribution X_0, then M is said to be *homogeneous*. Thus the transition probability can be written in matrix form P, where $P_{i,j} = P(X_{n+1} = i|X_n = j)$. Define that $\pi_i^{(n)} = P(X_n = i)$. When $n = 1$,

$$\begin{aligned} \pi_i^{(1)} &= \sum_{j\in\Omega} P(X_1 = i, X_0 = j) \\ &= \sum_{j\in\Omega} P(X_1 = i|X_0 = j)P(X_0 = j) \end{aligned}$$

It can be written in matrix form,

$$\begin{aligned} \pi^{(1)} &= P\pi^{(0)} \\ \pi^{(n+1)} &= P\pi^{(n)} \end{aligned}$$

where $\pi^{(n)} = [\pi_j^{(n)}, j = 1...N]^T$. Suppose that for some π that

$$\pi = P\pi \tag{11}$$

e.t., π is the eigenvector of P with eigenvalue 1. A Markov chain is said to be *ergodic* is when the $\pi^{(n)} \to \pi$ as $n \to \infty$ for any $\pi^{(n)}$. And π is the *equilibrium distribution* of the chain M.[5]

Metropolis-Hastings Markov Chain Monte Carlo is a certain type of algorithm which generates a Markov Chain with equilibrium distribution π, Let $X_n = i$ and X_{n+1} is determined in the following way.

1. *Generation setup*: Generate a candidate state j from i with some distribution $g(j|i)$. $g(j|i)$ is a fixed distribution that we are free to choose, so long that it satisfies the conditions
   - $g(j|i) = 0) \to g(i|j) = 0$; (can't go forward implies can't go back)
   - $g(j|i)$ is the transition matrix of an irreducible Markov chain on Ω.
2. *Acceptance step*: With probability

$$\alpha(j|i) = min\left\{1, \frac{\pi_j g(i|j)}{\pi_i g(j|i)}\right\} \tag{12}$$

   set $X_{n+1} = j$(i.e., accept j), otherwise set $X_n = i$(i.e., reject j).

Metropolis (1953) and Hastings (1970) have proved the uniqueness of equilibrium distribution π using MH-MCMC algorithm under the condition that the ergodic Markov Chain M is irreducible and aperiodic.[5]

IMPLEMENTATION OF MH-MCMC

Before the implementing MH-MCMC algorithm, preliminary works including data normalization and principle component analysis can be done in similar fashion as in RL algorithm. The data used in following discuss is Z', the reduced-dimension eigen image set. In general, the ultimate goal in many parameter estimation scenarios is to obtain the posterior distribution, e.t., $P(X|Z')$ in ice classification application. By applying MH-MCMC sampling algorithm we can acquire a sampled version of Ω, which has equilibrium distribution $P(X|Z')$. By the uniqueness theorem, several assumptions have to be made for implementing such algorithm: First, the state space M, formed by all possible combination of ice class image (X), can be model as a random Markov process; second, the process is ergodic, irreducible and aperiodic.

An implementation of the Metropolis-Hastings Markov Chain Monte Carlo (MH MCMC) on ice classification involves the following steps:

1. Suppose current classified ice type image is $X_n = f$, a pixel at location (k, l) is randomly selected with uniform distribution $\frac{1}{KL}$, where K, L is the dimension of the image and $f_{kl} = \{1, 2, 3, 4, 5\}$ each number represent one ice type. The four candidate states are $X_{n+1} = f^{(i)}(i = 1, 2, 3, 4)$ defined as

$$f_{mn}^{(i)} \begin{cases} \neq f_{kl} & m = k, n = l \\ = f_{mn} & m \neq k, n \neq l \end{cases}$$

   e.t., the only change for the candidate states is at position (k, l), all other pixels stay unchanged.

2. Calculate the acceptance probability $\alpha(f'|f)$ using

$$d = argmax_{(i)}\left\{\frac{P(f^{(i)}|Z')g(f|f^{(i)})}{P(f|Z')g(f^{(i)}|f)}\right\}$$

$$\alpha(f^{(d)}|f) = min\left\{1, \frac{P(f^{(d)}|Z')g(f|f^{(d)})}{P(f|Z')g(f^{(d)}|f)}\right\}$$

   since the generation distribution is uniform, $g(f|f^{(i)}) = g(f^{(i)}|f)$. The posterior distribution by the Bayes' rule is

$$P(f^{(d)}|Z') \propto P(Z'|f^{(d)})P(f^{(d)}),$$

   where $P(Z'|f^{(d)})$ is multi-variant Gaussian distribution (Equation 6) and $P(X = j)$ is the spatial

prior. A possible prior as is the *blob*-shaped prior that is the ice type of a given pixel is more likely to be identical to the ice types of its neighbors. In mathematical expression,

$$P(X = f) = \frac{1}{K_\gamma} e^{-\gamma \# f}, \qquad (13)$$

$\# f$ is the number of edges linking disagreeing pixels in f, $\gamma > 0$ is a constant and K_γ is the normalization factor. Then

$$\frac{P(f^{(d)}|Z')}{P(f|Z')} = \frac{P(Z'|f^{(d)})}{P(Z'|f)} e^{-\gamma(\# f^{(d)} - \# f)}. \qquad (14)$$

3. The candidate state $f^{(d)}$ is accepted with the probability $\alpha(f^{(d)}|f)$. If the state is accepted, $X_{n+1} = f^{(n)}$, otherwise the state stays unchanged.

RESULT

Using MH-MCMC algorithm with $\gamma = 0.5$, the final classified Antarctica sea ice on Julian date 251 of year2002 is shown in Figure 7.

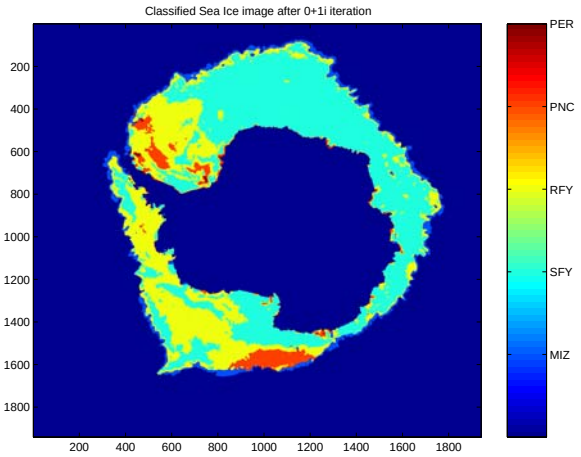

Figure 7: Antarctic sea ice classification using MH-MCMC algorithm

Comparing with Figure 3 generated using RL iterative MAP algorithm, Figure 7 displays generally similar orientation for each ice class region. In Figure 7, rough first year ice is more abundant in southern Antarctica, while pancake ice is shrunk in the northern region. Take closer look at Figure 8, the artifact in the left generated by RL algorithm has been removed in the image on the right using MH-MCMC algorithm. As a result of applying spatial *blob*-shaped prior in MH-MCMC algorithm, ice region such as smooth first year ice is bulkier than using RL algorithm. Questions may rise: is the spatial prior in MH-MCMC a legitimate assumption even though it can significantly reduce noise, and how accurate both algorithm really are?

Further investigation emphasizing on validation of both algorithms is to be carried on using other data source, such as Synthetic Aperture Radar (SAR), visible light or infrared imagery.

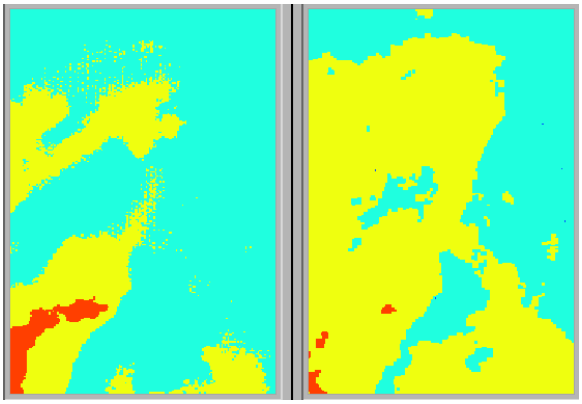

Figure 8: Classified ice type using RL (left) and MH-MCMC (right) algorithm

References

[1] David G. Long Quinn P. Remund and Mark R. Drinkwater, "An Iterative Approach to Multisensor Sea Ice Classification", *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 38, NO. 4*, July 2000.

[2] Q.P. Remund and D.G. Long, "Validation of the SIRF Resolution Enhancement Algorithm for Scatterometer Data Using SAR Imagery", vol. 2, pp. 1309–1311, 1999.

[3] W. Campbell D. Cavalieri, P. Gloersen, *Artic and Antarctic Seaice, 1978-1987*, NASA, Washington, D.C., 1992.

[4] M. Chenny E. Cherkaeva et. M. Golden, D. Borup, "Invers Electromagnetic Scattering Models for Sea Ice", *IEEE Trans. on Geosci. and Rem. Sens.*, vol. 36, no. 5, pp. 1675–1704, 1998.

[5] S. Tan and C. Fox, Eds., *Invers Problem Course Notes*, http://www.phy.auckland.ac.nz/Staff/smt/453707SC.html, Auckland University.