Sampling Errors in Some Global Climate Sampling Schemes

Ta-Hsin Li

Department of Statistics
Texas A&M University
College Station, Texas 77843

Joint work with G. R. North and S. S. Shen

Prepared for presentation at the American Geophysical Union Meeting
December 8–12, 1997, San Francisco, CA

Work supported in part by the Global Program of NOAA

OVERVIEW

- Spherical Harmonic Representation
- Linear Estimation and Aliasing Effects
- Measures of Sampling Error
- Sampling by Point-Gauge Networks
- Sampling by Polar-Orbiting Satellites

SPHERICAL HARMONIC REPRESENTATION

- *Laplace Series Expansion:* for an anomaly field $T(\boldsymbol{n}, t)$,

$$T(\boldsymbol{n}, t) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} T_{\ell m}(t) \, Y_{\ell m}(\boldsymbol{n})$$

ℓ = degree

m = zonal wavenumber

- *Applications:*
  - Spectral methods in some GCMs' numerical schemes
  - Data archiving and data compression
  - Estimation and change detection

ESTIMATION OF S. H. COEFFICIENTS

- *Spherical Harmonic Coefficients:*

$$T_{\ell m}(t) = \int T(\boldsymbol{n}, t)\, Y^*_{\ell m}(\boldsymbol{n})\, d\Omega(\boldsymbol{n})$$

- *Point-Gauge Observations:* simultaneous measurements

$$T(\boldsymbol{n}_j) \qquad (j = 1, \cdots, J)$$

- *Linear Estimation:* weighted average of observations

$$\widehat{T}_{\ell m} = \sum_{j=1}^{J} w_j\, T(\boldsymbol{n}_j)\, Y^*_{\ell m}(\boldsymbol{n}_j)$$

ALIASING EFFECTS

- *Aliasing Effects:* weighted sum of the aliases of T_{lm}

$$\hat{T}_{\ell m} = \sum_{\ell'=0}^{\infty} \sum_{m'=-\ell'}^{\ell'} \Gamma(\ell, m, \ell', m')\; T_{\ell' m'}$$

- *Aliasing Coefficients:* depend only on the sampling design

$$\begin{aligned} \Gamma(\ell, m, \ell', m') &= \sum_{j=1}^{J} w_j \, Y_{\ell' m'}(\boldsymbol{n}_j) \, Y^*_{\ell m}(\boldsymbol{n}_j) \\ &\neq \delta_{\ell' - \ell} \, \delta_{m' - m} \end{aligned}$$

GAUSS-LEGENDRE VERSUS UNIFORM

- *Uniform Network:* $J = MN$,

$$\begin{aligned} \text{Longitude:} \quad \theta_p &= -\pi + 2\pi p/M \quad (p = 1, \cdots, M) \\ \text{Latitude:} \quad \phi_q &= -\pi/2 + \pi q/(N+1) \quad (q = 1, \cdots, N) \\ \text{Weights:} \quad w_j &= 4\pi/(MN) \quad \text{(constant)} \end{aligned}$$

- *Gauss-Legendre Network:*

$$\begin{aligned} \phi_q &= \text{roots of the Legendre polynomial } P_N(\sin\phi) \\ w_j &= \text{Gaussian weights} \times 2\pi/M \end{aligned}$$

STRUCTURE OF ALIASING COEFFICIENTS

- *General Case:* unif. longitude, symm. latitude, even M

$$\Gamma(\ell, m, \ell', m') = 0 \quad \text{if } \ell' \neq \ell + 2r \text{ or } m' \neq m + \tfrac{1}{2}Ms$$

- *Gauss-Legendre Case:* with $N = \frac{1}{2}M$,

$$\Gamma(\ell, m, \ell + 2r, m) = 0 \quad \text{if } -\tfrac{1}{2}\ell \leq r < N - \ell,\ r \neq 0$$

$$\Gamma(\ell, m, \ell, m) = 1 \quad \text{if } |m| \leq \ell < N \quad \text{(unbiased)}$$

MEASURES OF SAMPLING ERROR

- *Mean-Squared Error:* a statistical index of error

$$\epsilon_{\ell m}^2 = E\Big\{|\widehat{T}_{\ell m} - T_{\ell m}|^2\Big\}$$

- *Relative MSE:* MSE as a fraction of the "size" of $T_{\ell m}$

$$e_{\ell m}^2 = \frac{\epsilon_{\ell m}^2}{\mathrm{Var}\{T_{\ell m}\}} \qquad \text{(dimensionless)}$$

COMPOSITION OF SAMPLING ERROR

- *Aliased Power:* (assuming homogeneity)

$$\epsilon_{\ell m}^2 = \sum_{\ell'=0}^{\infty} \sum_{m'=-\ell'}^{\ell'} A(\ell, m, \ell', m')\ \sigma_{\ell'}^2$$

- *Factorization Property:*

$$A(\ell, m, \ell', m') = |\Gamma(\ell, m, \ell', m') - \delta_{\ell'-\ell}\delta_{m'-m}|^2$$

  – squared aliasing function (SAF)
  – depends only on the sampling design

$$\sigma_{\ell'}^2 = \text{Var}\{T_{\ell',m'}\}$$

  – power of the aliases $T_{\ell' m'}$
  – depends only on the field's statistics

COMPOSITION OF MSE (CONT'D)

- *MSE Spectrum:* distribution of the aliased power

$$d_{\ell m}(\ell') = \frac{1}{\epsilon_{\ell m}^2} \sum_{m'=-\ell'}^{\ell'} A(\ell, m, \ell', m') \; \sigma_{\ell'}^2$$

$$\sum_{\ell'=0}^{\infty} d_{\ell m}(\ell') = 1 \qquad \text{(dimensionless)}$$

- *Interpretation:* the fraction of MSE that can be attributed to the $2\ell' + 1$ spherical harmonic components of degree ℓ'

A SIMPLE CLIMATE MODEL

- *Energy Balance Model (EBM):*

$$-\mu_0^2 \nabla^2 T(\boldsymbol{n}) + T(\boldsymbol{n}) = F(\boldsymbol{n})$$

$F(\boldsymbol{n})$: zero-mean white noise, $\mathrm{Var}\{F(\boldsymbol{n})\} = \sigma_F^2$

μ_0 : spatial scale parameter

- *Statistical Parameters:*

$$\sigma_\ell^2 = \frac{4\pi\sigma_F^2}{\{1 + \mu_0^2 \ell(\ell+1)\}^2}$$

Relative MSE for Estimating T_{00}

Network	Number of Stations ($J = MN, N = M/2$)									
	2	8	18	32	50	72	98	128	162	200
Gauss-Legendre	4.2	.60	.18	.067	.030	.015	.008	.005	.003	.002
Lat.-Long. Uniform	4.2	.62	.17	.071	.048	.045	.049	.054	.060	.066

Relative MSE of 98-Station Networks ($N = \frac{1}{2}M = 7$)

	Gauss-Legendre Network					Lat.-Long. Uniform Network				
m	$\ell=1$	$\ell=2$	$\ell=3$	$\ell=4$	$\ell=5$	$\ell=1$	$\ell=2$	$\ell=3$	$\ell=4$	$\ell=5$
0	.009	.021	.045	.099	.22	.134	.379	.461	.330	.30
1	.016	.021	.042	.092	.20	.042	.057	.287	.679	1.22
2		.029	.045	.087	.17		.059	.048	.142	.34
3			.061	.096	.17			.082	.082	.16
4				.131	.20				.124	.15
5					.28					.21

SAMPLING BY SATELLITE

- *Satellite Observations:* geo-asynchronous satellites

$$T(\boldsymbol{n}(t), t) \quad t \in \mathcal{I} = \text{sampling interval}$$

- *Sampling Properties:*
  - non-complete (short-time) coverage in space
  - non-simultaneous (global) sampling in time

ESTIMATION USING SATELLITE DATA

- *Linear Estimation:*

$$\widehat{T}_{\ell m} = \int_{\mathcal{I}} T(\boldsymbol{n}(t), t)\, Y^{*}_{\ell m}(\boldsymbol{n}(t))\, w(t)\, dt$$

- *Estimands:*
  - instantaneous coefficients $T_{\ell m}(t_0)$ for some $t_0 \in \mathcal{I}$
  - time-average coefficients

$$\bar{T}_{\ell m} = \frac{1}{|\mathcal{I}|} \int_{\mathcal{I}} T_{\ell m}(t)\, dt$$

SAMPLING ERRORS AND ALIASING EFFECTS

- *Mean-Squared Error:* composed of aliased power

$$\epsilon_{\ell m}^2 = \sum_{\ell'=0}^{\infty} \sum_{m'=-\ell'}^{\ell'} \int_{-\pi}^{\pi} A(\ell, m, \ell', m', \omega)\, f(\ell', \omega)\, d\omega$$

(assuming homogeneity and stationarity)

- *Factorization Property:*

  $A(\ell, m, \ell', m', \omega)$: spherical-temporal SAF

  $f(\ell', \omega)$: spherical-temporal PSD of the aliases

MSE SPECTRA

- *Joint MSE Spectrum:*

$$d_{\ell m}(\ell', \omega) = \frac{1}{\epsilon_{\ell m}^2} \sum_{m'=-\ell'}^{\ell'} A(\ell, m, \ell', m', \omega)\ f(\ell', \omega)$$

- *Spherical MSE Spectrum:* distribution over degrees

$$d_{\ell m}(\ell') = \int_{-\pi}^{\pi} d_{\ell m}(\ell', \omega)\, d\omega$$

- *Temporal MSE Spectrum:* distribution over frequencies

$$d_{\ell m}(\omega) = \sum_{\ell'=0}^{\infty} d_{\ell m}(\ell', \omega)$$

POLAR-ORBITING SATELLITES

- *Sampling Design:* with $\alpha = 2\pi M/N$,

$$\begin{aligned} \phi(t) &= \text{arcsin}(\sin(\alpha t)) \\ \theta(t) &= -2\pi t \\ w(t) &= \alpha\pi \cos(\alpha t)/M \end{aligned}$$

- *Remark:* the polar-orbiting satellite is assumed to encircle the Earth M times in N days to complete its orbital pattern without repetition

A SIMPLE CLIMATE MODEL

- *Time-Dependent EBM*:

$$\tau_0 \frac{\partial}{\partial t} T(\boldsymbol{n}, t) - \mu_0^2 \nabla^2 T(\boldsymbol{n}, t) + T(\boldsymbol{n}, t) = F(\boldsymbol{n}, t)$$

$F(\boldsymbol{n}, t)$:	white noise, $\mathrm{Var}\{F(\boldsymbol{n}, t)\} = \sigma_F^2$
τ_0 :	temporal scale parameter
μ_0 :	spatial scale parameter

- *Power Spectral Density*:

$$f_\ell(\omega) = \frac{\sigma_F^2 \tau_0 / \pi}{\tau_0^2 \omega^2 + \{1 + \mu_0^2 \ell(\ell + 1)\}^2}$$

Relative MSE for Estimating $\bar{T}_{00}$ of EBM Fields

orbit parameters	ave period	time scale τ_0 (days)						
(M, N)	(days)	1	2	5	10	20	30	40
(4,1)	1	0.861	0.575	0.411	0.362	0.340	0.333	0.329
(8,2)	2	1.089	0.638	0.381	0.314	0.294	0.292	0.293
(16,4)	4	1.389	0.844	0.446	0.320	0.275	0.268	0.269
(32,8)	8	1.207	0.905	0.536	0.360	0.275	0.254	0.249
(16,1)	1	0.678	0.351	0.150	0.083	0.050	0.039	0.033
(32,2)	2	0.976	0.508	0.212	0.111	0.060	0.044	0.036

- Relative MSE decreases as the time scale τ_0 increases
- Relative MSE decreases for long-term averages only if the time scale of the field is sufficiently large

Relative MSE for Estimating $\bar{T}_{\ell m}$ of EBM Fields ($M = 16, N = 1$)

	$\tau_0 = 1$ day					$\tau_0 = 10$ days				
m	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
0	0.778	1.011	1.352	1.789	2.280	0.096	0.124	0.171	0.235	0.313
1	0.827	1.079	1.451	1.919	2.464	0.102	0.134	0.179	0.240	0.319
2		1.102	1.483	1.966	2.512		0.136	0.188	0.454	0.331
3			1.505	1.995	2.559			0.188	0.463	0.346
4				2.019	2.590				0.259	0.352
5					2.612					0.350

	$\tau_0 = 30$ days				
m	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
0	0.043	0.054	0.074	0.104	0.139
1	0.048	0.063	0.083	0.107	0.141
2		0.065	0.090	0.120	0.153
3			0.091	0.125	0.167
4				0.126	0.171
5					0.172

- Relative MSE decreases as the time scale τ_0 increases
- Relative MSE increases as the wavenumbers (ℓ, m) increase
- Relative MSE increases faster with the number of zonal waves $(m = \ell)$ than with the number of meridional waves $(m = 0)$: the lack of zonal resolution is largely responsible

CONCLUSIONS

- *Methodology*: a unified method to analyze sampling errors for point-gauge networks and asynchronous satellites

- *Future Research*: (a) analyzing observations from multiple sources, (b) understanding the influence of diurnal and seasonal variations on sampling errors