Comparative Performance and Uncertainty of Gridded Atmospheric Products in Dzud Risk Assessments: A Case Study of the Selenge River Basin

Sural, K.E.[1*], Shiklomanov[1], N.I.,& Alexeev[2], V., Shiklomanov A.I.[3]

[1] eges@gwu.edu, The George Washington University, USA; [2] International Arctic Research Center, University of Alaska Fairbanks, USA; [3] University of New Hampshire, USA

Abstract

Accurate meteorological inputs are critical for climate risk modeling, yet sparse observations in complex terrain create significant data gaps. We evaluated four high-resolution atmospheric datasets (UAF WRF, ERA5-Land, MERRA-2, and NCEP-NCAR) across the topographically complex Selenge River basin (~447,000 km²) to determine how forcing variability affects the representation of extreme winter conditions linked to livestock mortality. Gridded climate indices were validated against local station records using ACC, MSSS, and ensemble-spread analyses. Results reveal substantial inter-dataset divergence and systematic cold-season biases, demonstrating that dataset selection strongly impacts permafrost sensitivity and regional climate assessments. Resolving these uncertainties is essential for reliably mapping hydroclimatological variables and guiding adaptation policy across sensitive Northern Eurasian ecosystems.

Methodology

To evaluate the accuracy and structural uncertainties of the reanalysis products (UAF WRF, ERA5-Land, MERRA-2, and NCEP–NCAR) against point-based station observations, data were spatially paired using bilinear interpolation. Two primary statistical validation pathways were then utilized:

1. Evaluation of Normals:
Mean Squared Skill Score (MSSS) measures the explanatory power of the gridded model relative to a baseline historical average (climatology):

$$MSSS = 1 - \frac{MSE_{model}}{MSE_{reference}}$$

MSSS = 1: Perfect forecast. The model error is zero.
MSSS > 0: The model is more skillful than the climatology.
MSSS = 0: The model has the same accuracy as simply predicting the historical average
MSSS < 0: The model performs worse than the historical average.

2. Evaluation of Trends:
Anomaly Correlation Coefficient (ACC) computes the linear association between modeled and observed anomalies, effectively factoring out the underlying seasonal cycle:

$$ACC = \frac{\left(\sum(f-c)(a-c)\right)}{\left(\sqrt{\sum(f-c)^2}\sqrt{\sum(a-c)^2}\right)}$$

ACC = 1: Perfect association between forecast and actual anomalies.
ACC > 0.6: Useful for numerical weather predictions.
ACC = 0: No linear association, meaning the forecast has no skill in predicting the anomaly.

Results

Dzud Risk Score — Anomaly Correlation Coefficient (ACC) (Stations × Reanalysis products)

Stations	ERA5	MERRA2	NCEP	WRF
Tariat	0.64	0.71	0.63	0.44
Tosontsengel	0.76	0.77	0.49	0.21
Tsetserleg	0.47	0.76	0.42	0.27
Khatgal	0.72	0.79	0.54	0.10
Khujirt	0.58	0.73	0.40	0.35
Zuunmod	0.42	0.48	0.32	0.35
Erdenemandal	0.76	0.69	0.52	0.34
Erdenesant	0.59	0.55	0.22	0.16
Ulaanbaatar	0.54	0.57	0.14	0.51
Erdenet-Ovoo	0.28	0.49	0.25	-0.02
Moron	0.76	0.83	0.37	0.70
Buyant-Ukhaa	0.20	0.28	0.09	-0.27
Bulgan	0.63	0.65	0.44	0.02
Tarialan	0.78	0.65	0.37	0.44
Ugtaal	0.57	0.66	0.32	0.06
Khutag	0.60	0.50	0.37	-0.23
Baruunkharaa	0.65	0.68	0.28	-0.06
Orkhon	0.57	0.60	0.27	0.13
Darkhan	0.70	0.56	0.25	0.33
Yeroo	0.21	0.52	0.11	-0.06
Sukhbaatar	0.32	0.41	0.15	-0.22

Dzud Risk Score — Mean Squared Skill Score (MSSS) (Stations × Reanalysis products; colour scale: Dzud Risk Score Metric Value, −1.00 to 1.00)

Stations	ERA5	MERRA2	NCEP	WRF
Tariat	0.37	0.43	0.32	0.01
Tosontsengel	0.57	0.55	0.13	-0.13
Tsetserleg	-0.32	0.27	-0.45	-0.46
Khatgal	0.48	0.60	0.13	-0.64
Khujirt	0.21	0.49	-0.10	-0.13
Zuunmod	-0.02	0.08	-0.30	-0.37
Erdenemandal	0.54	0.34	0.02	-0.19
Erdenesant	0.03	0.07	-0.70	-0.66
Ulaanbaatar	-0.04	-0.01	-1.44	-0.39
Erdenet-Ovoo	-0.41	0.03	-0.58	-0.78
Moron	0.58	0.69	0.01	0.48
Buyant-Ukhaa	-1.11	-0.18	-1.39	-1.60
Bulgan	0.34	0.38	-0.01	-0.77
Tarialan	0.46	0.27	-0.26	-0.10
Ugtaal	0.23	0.37	-0.34	-0.75
Khutag	0.13	-0.03	-0.39	-1.51
Baruunkharaa	0.26	0.31	-0.41	-0.98
Orkhon	0.21	0.22	-0.28	-0.79
Darkhan	0.11	-0.43	-0.98	-0.81
Yeroo	-0.37	0.17	-0.55	-1.04
Sukhbaatar	-0.50	-0.27	-0.91	-2.15

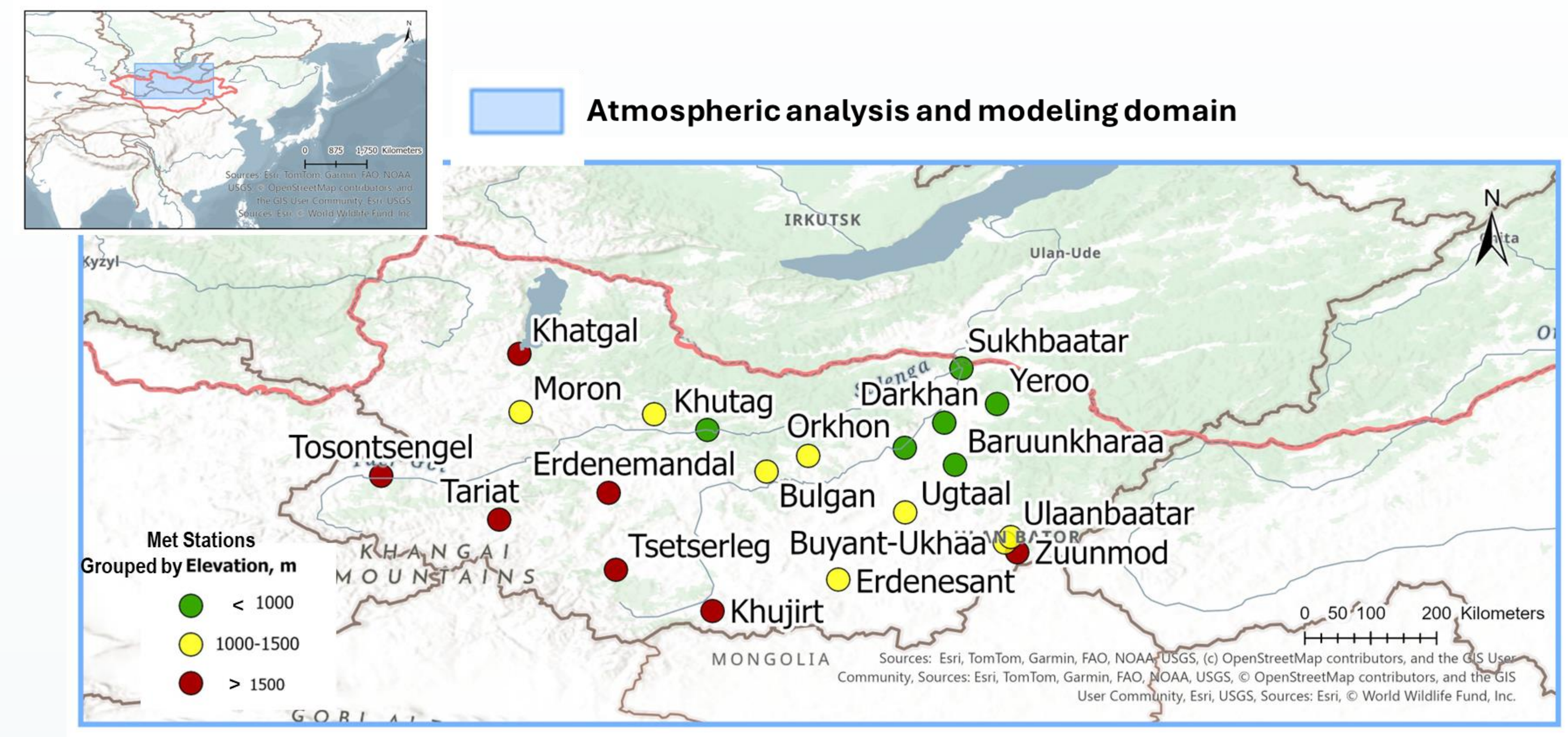
Average ACC across Stations (Variables & Indices)

Variables & Indices	ERA5	MERRA2	NCEP	WRF
Total Winter Precipitation	0.71	0.76	0.54	-0.02
Maximum Snow Depth	0.22	0.08	-0.06	0.22
Cold Spell Duration	0.54	0.55	0.50	0.23
Freeze-Thaw Cycles	0.31	0.32	0.15	0.08
Extreme Cold Days	0.69	0.70	0.64	0.24
Wind Chill Index	0.10	0.04	0.11	0.60
Extreme Wind Days	0.16	0.24	0.16	0.26
SPEI Drought Index	0.72	0.79	0.45	0.40
Composite Dzud Risk Score	0.56	0.61	0.33	0.17

Average MSSS across Stations (colour scale: Variable Average Metric Value, −1.00 to 1.00)

Variables & Indices	ERA5	MERRA2	NCEP	WRF
Total Winter Precipitation	0.42	0.51	0.08	-1.05
Maximum Snow Depth	-0.42	-0.80	-1.16	-0.42
Cold Spell Duration	0.10	0.11	0.02	-0.55
Freeze-Thaw Cycles	-0.38	-0.35	-0.69	-0.83
Extreme Cold Days	0.38	0.41	0.30	-0.52
Wind Chill Index	-0.79	-0.91	-0.79	0.20
Extreme Wind Days	-0.69	-0.51	-0.69	-0.49
SPEI Drought Index	0.45	0.57	-0.10	-0.21
Composite Dzud Risk Score	0.08	0.21	-0.40	-0.66

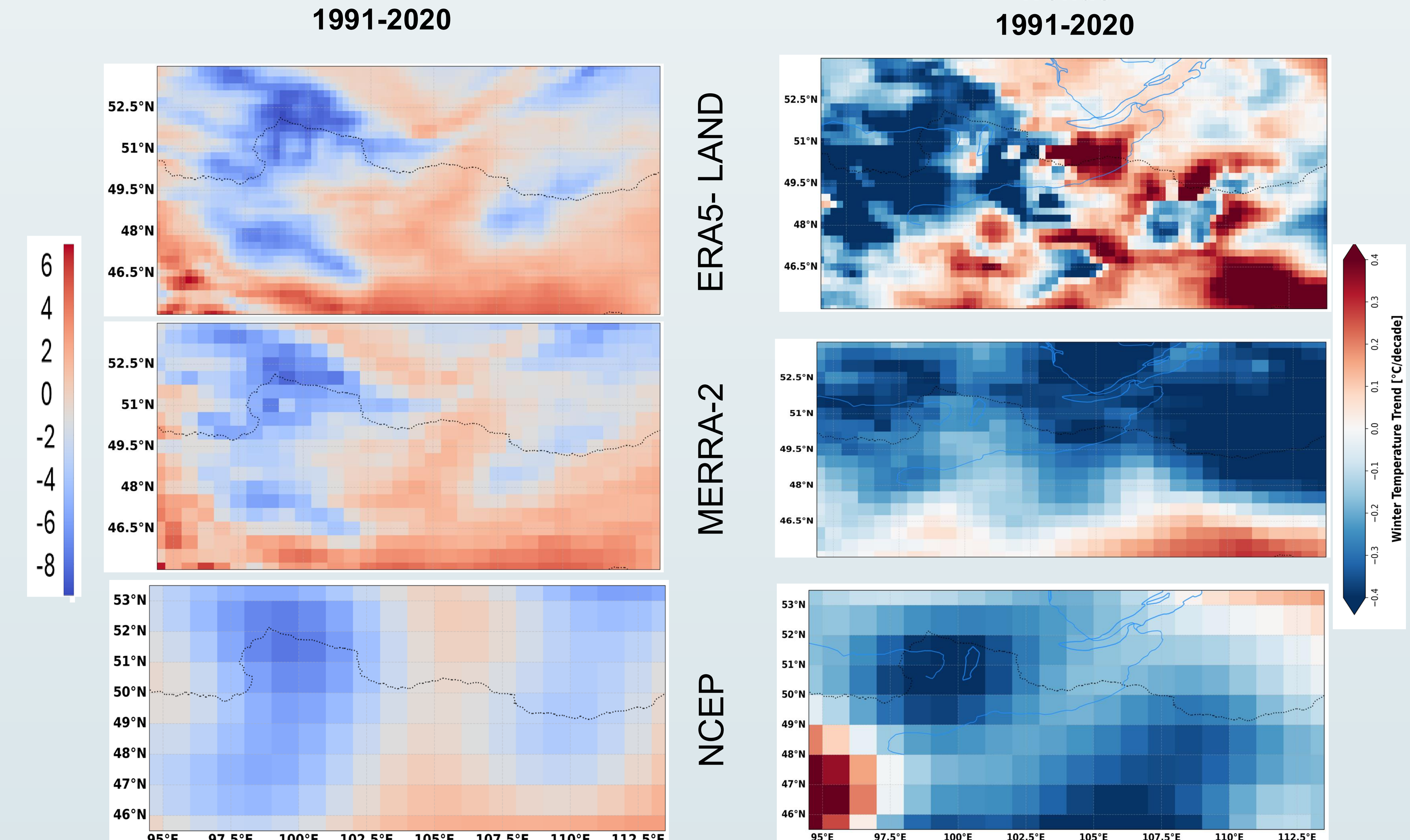
Study Area & Data



The Selenge River Basin (~447,000 km², Mongolia) features complex terrain with heterogeneous permafrost covering 29.5% of the area. Since these marginal zones hover near 0°C and are highly sensitive to climatic shifts, accurate forcing data is vital for subsurface modeling. This study validates gridded atmospheric fields from **UAF WRF, ERA5-Land, MERRA-2, and NCEP–NCAR** against ground observations stratified across low (<1000 m), mid (1000–1500 m), and high (>1500 m) elevation tiers to systematically quantify terrain-driven errors.

	ERA5-Land	NCEP / NCAR (R1)	MERRA-2
Developing Institution	European Centre for Medium-Range Weather Forecasts (ECMWF) / Copernicus	National Centers for Environmental Prediction (NOAA-NCEP)	Global Modeling and Assimilation Office (NASA GMAO)
Spatial Resolution	~9 km (0.1° × 0.1°)	~250 km (2.5° × 2.5°)	~50 km (0.5° lat × 0.625° long)
Temporal Resolution	Hourly, Monthly	6-hourly, Daily, Monthly	Hourly, 3-hourly, Monthly
Temporal Coverage	1950 – Present	1948 – March 2026 (Production Ended)	1980 – Present

Annual Air Temperature Climatic Normals 1991-2020 | **Annual Air Temperature Trends 1991-2020**



Conclusion

The Macro Winner (MERRA-2 > ERA5): MERRA-2 delivers the strongest composite Dzud fidelity ($ACC = 0.61$, $MSSS = 0.21$), driven by superior tracking of moisture anomalies (Precipitation ACC = 0.76, SPEI ACC = 0.79).

The Wind Chill Inversion: Global products fail completely at modeling deadly wind-temperature coupling ($MSSS \leq -0.79$), whereas dynamically downscaled WRF uniquely captures surface Wind Chill Index (ACC = 0.60, MSSS = 0.20).

NCEP Magnitude Failure: NCEP loses all predictive value for Dzud risk ($MSSS = -0.40$) due to large variance and magnitude biases across stations.

Acknowledgements

This research is supported by the U.S. National Science Foundation under the following projects: "NNA Research: Collaborative Research: Frozen Commons: Change, Resilience and Sustainability in the Arctic," with Awards #2127343 to The George Washington University and #2127344 to the University of New Hampshire; and "NNA Research: Collaborative Research: Socio-Ecological Systems Transformation in River basins of the sub-Arctic under climate change (SESTRA)," with Awards #2318381 to The George Washington University, #2318380 to the University of New Hampshire, and #2318383 to the University of Alaska, Fairbanks.