Improved models of seasonality for global food security and nutrition

Aishwarya Venkat, Anastasia Marshak, Elena N. Naumova

GEOMED 2024, University of Hasselt, Belgium

September 11, 2024

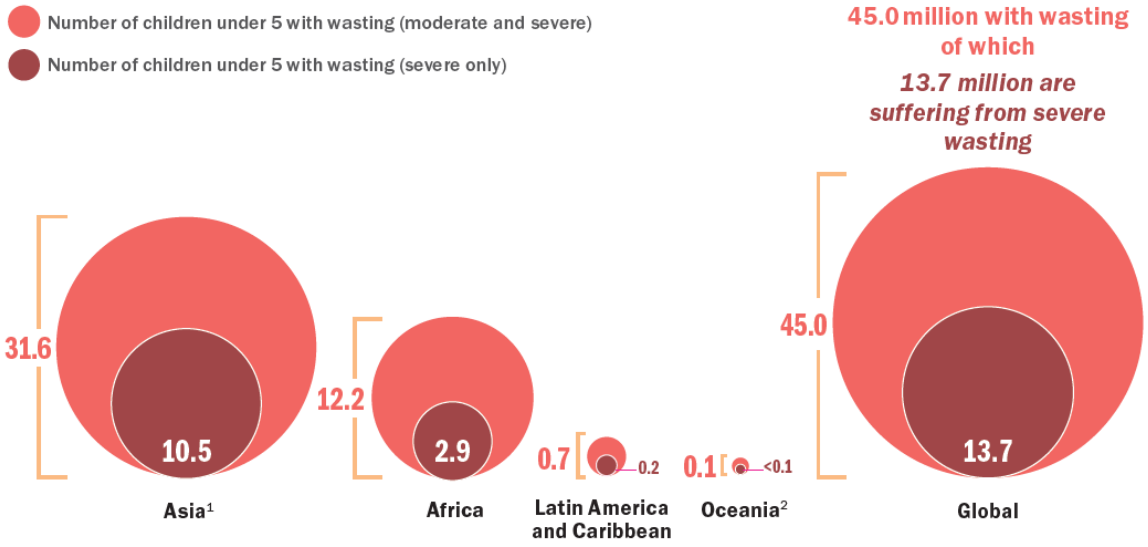
Tufts UNIVERSITY | Gerald J. and Dorothy R. Friedman School of Nutrition Science and Policy

InForMID
Initiative for the Forecasting and Modeling of Infectious Diseases

Feinstein International Center

Undernutrition among children under five

- Stunting: 148.1 million (22.3%)
  - Low height-for-age
  - Result of chronic undernutrition
  - Associated with lower cognition, lower earnings, and increased morbidity in adulthood
- Wasting: 45 million (6.8%)
  - Low weight-for-height
  - Result of recent weight loss due to disease or acute food insecurity
  - Associated with increased mortality risk



The JME does not currently adjust for seasonal or other factors that can affect wasting prevalence estimates

Source: UNICEF, WHO, World Bank. (2023). *Joint Child Malnutrition Estimates*.

Survey timing – practical constraints

- Measurement dependent on cross-sectional surveys
  - JME data sources: LSMS, MICS, DHS
    - “the primary source dataset contained 1100 data sources from 160 countries and territories”… “global estimates are highly representative of the majority of children across the globe for the most recent period”
  - Low frequency measurements, often during same months
    - Physical access for enumerators, year-round access difficult and expensive
    - Representative of population at a point in time
  - Data collection contingent on broad categories such as pre-harvest vs post-harvest, dry vs wet season, or hunger/lean vs plenty seasons
  - “Over 50% of studies [in African drylands] rely on 2-4 time points within the year and/or the inclusion of time as a categorical variable in the analysis” (Marshak et al, 2021)

Modeling temporal variation

- Seasons
  - Does acute malnutrition follow the agricultural calendar? **Not always!**
  - “Food first hypothesis”: malnutrition driven by lack of food (Pelletier, 1995)
  - Reflects assumptions and computational constraints of previous decades
- Discrete calendar month
  - Month as “fixed” effect or discrete category ignores temporal autocorrelation
  - Flexible reference category?
- Continuous time
  - High frequency signals – AR, MA, exponential smoothing
  - Interrupted time series, regression discontinuity designs
  - Sparse time series: harmonic regression

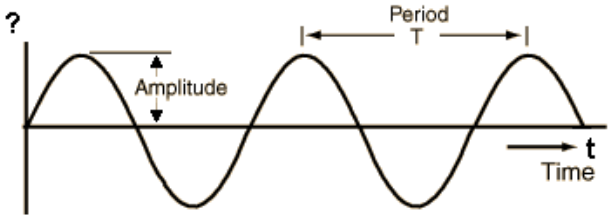
Challenges

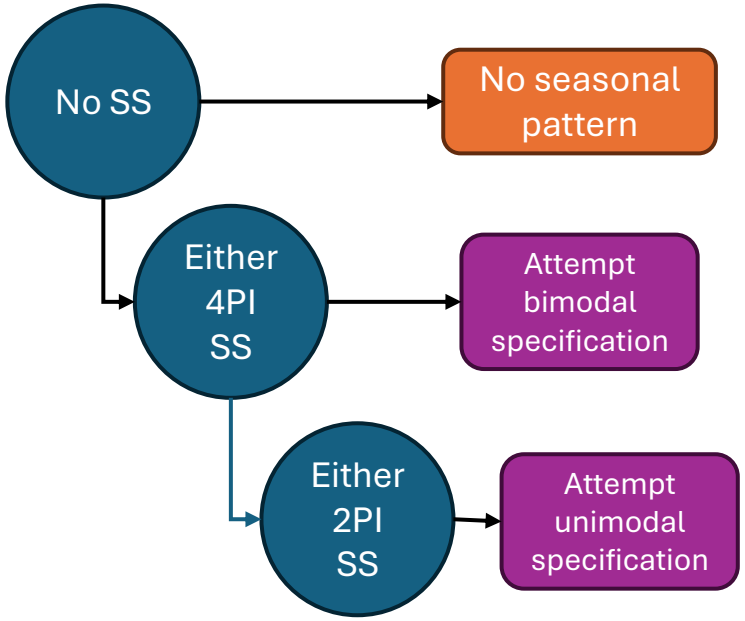
- How can we more accurately model geographic variability in wasting?
  - CAR, BYM, Fay-Herriot, hierarchical models
  - Inclusion of demographic and climate covariates or subsets
- How can we accurately model a continuous temporal outcome which is observed infrequently?
  - Multiple harmonic regression

Multiple Harmonic Regression

$$O_{i,k,p} = \beta_0 + \beta_1 \sin(2\pi\omega t_{i,k,p}) + \beta_2 \cos(2\pi\omega t_{i,k,p}) + \beta_3 \sin(4\pi\omega t_{i,k,p}) + \beta_4 \sin(4\pi\omega t_{i,k,p}) + \beta_5 T(t_{i,k,p})$$

Characteristic	**Unimodal (2π)**		**Bimodal (4π)**
	Gaussian Linear Model	*Log-Linear Model*	*Gaussian Linear or Log-Linear*
Regression Model	$Y_t = \beta_0 + \beta_1 \sin(2\pi\omega t) + \beta_2 \cos(2\pi\omega t) + \beta_3 T(t)$	$\ln(E[Y_t]) = \beta_0 + \beta_1 \sin(2\pi\omega t) + \beta_2 \cos(2\pi\omega t) + \beta_3 T(t)$	Y_t or $\ln(E[Y_t]) = \beta_0 + \beta_1 \sin(2\pi\omega t) + \beta_2 \cos(2\pi\omega t) + \beta_3 \sin(2\pi\omega t) + \beta_4 \cos(2\pi\omega t) + \beta_5 T(t)$
Amplitude (γ)	$\gamma = \sqrt{\beta_1^2 + \beta_2^2}$	$\gamma = e^{\sqrt{\beta_1^2 + \beta_2^2}}$	$A = P_G - N_G$
95% Confidence Interval of Amplitude ($CI(\gamma)$)	$Var(\gamma) = \frac{\beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2 + 2\sigma_{\beta_1\beta_2}\beta_1\beta_2}{\beta_1^2 + \beta_2^2}$ $CI(\gamma) = \gamma \pm 1.96\sqrt{Var(\gamma)}$	$Var(\gamma) = \gamma^2\left(\frac{\beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2 + 2\sigma_{\beta_1\beta_2}\beta_1\beta_2}{\beta_1^2 + \beta_2^2}\right)$ $CI(\gamma) = \gamma \pm 1.96\sqrt{Var(\gamma)}$	Estimated arithmetically from 999 simulations which randomly drop up to 50% of dataset $CI(\hat{\gamma}) = \sum_{n=1}^{n=999} P_G - N_G$
Peak (P)	$P = \beta_0 + \gamma$	$P = e^{\beta_0} + \gamma$	Estimated arithmetically from first, second, and third differences of the predicted seasonal curve. P_L = local maximum where C′ = 0 and C″ < 0 P_G = global maximum, largest value of all P_Ls
Nadir (P)	$N = \beta_0 - \gamma$	$N = e^{\beta_0} - \gamma$	Estimated arithmetically from first, second, and third differences of the predicted seasonal curve. N_L = local minimum where C′ = 0 and C″ > 0 N_G = global minimum, smallest value of all N_Ls
Peak Timing (P_T)	Phase shift $\Theta = \arctan\left(\frac{\beta_1}{\beta_2}\right)$ If $\beta_1 > 0$ and $\beta_2 > 0$, $P_T = (\Theta)\frac{M}{2\pi}$ If $\beta_2 < 0$, $P_T = (\Theta + \pi)\frac{M}{2\pi}$ If $\beta_1 < 0$ and $\beta_2 > 0$, $P_T = (\Theta + 2\pi)\frac{M}{2\pi}$		Estimated arithmetically from first, second, and third differences of the predicted seasonal curve. $P_{T,L}$ = Timing of P_L, $P_{T,G}$ = Timing of P_G
95% Confidence Interval of Peak Timing ($CI(\Theta)$)	$Var(\Theta) = \frac{\beta_1^2\sigma_2^2 + \beta_2^2\sigma_1^2 - 2\sigma_{\beta_1\beta_2}\beta_1\beta_2}{(\beta_1^2 + \beta_2^2)^2}$ $CI(\Theta) = \Theta \pm 1.96\sqrt{Var(\Theta)}$		Estimated arithmetically from 999 simulations which randomly drop up to 50% of dataset $CI(\widehat{P_T}) = \sum_{n=1}^{n=999} P_{T,G}$





Complete code available on Github!

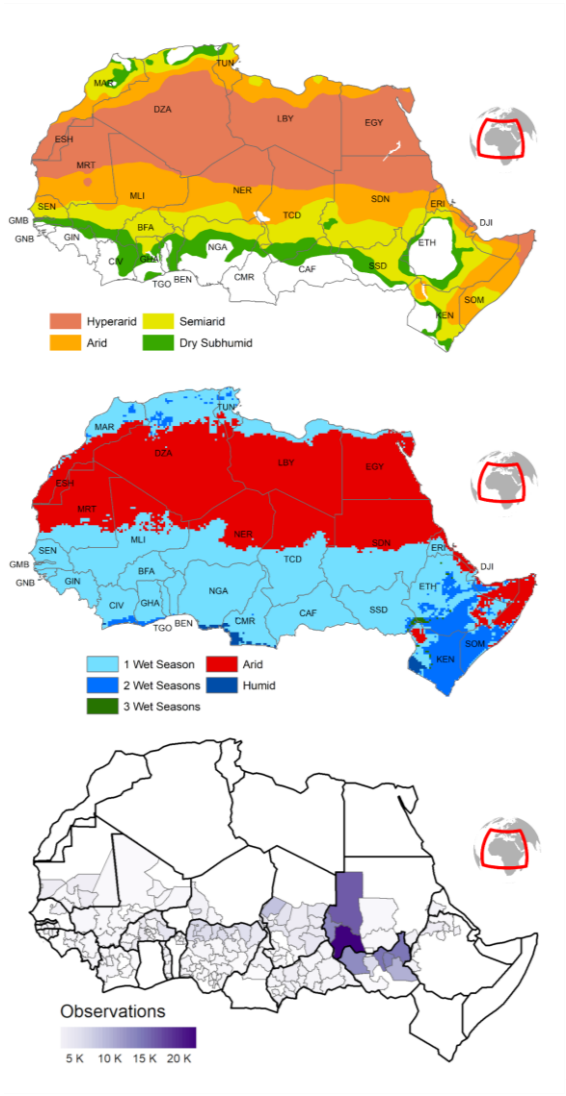
Source: Naumova, EN and MacNeill, IB (2006) *Seasonality assessment for biosurveillance systems*. Advances in Statistical Methods for the Health Sciences. Boston, MA: Birkhauser, pp 437–450.

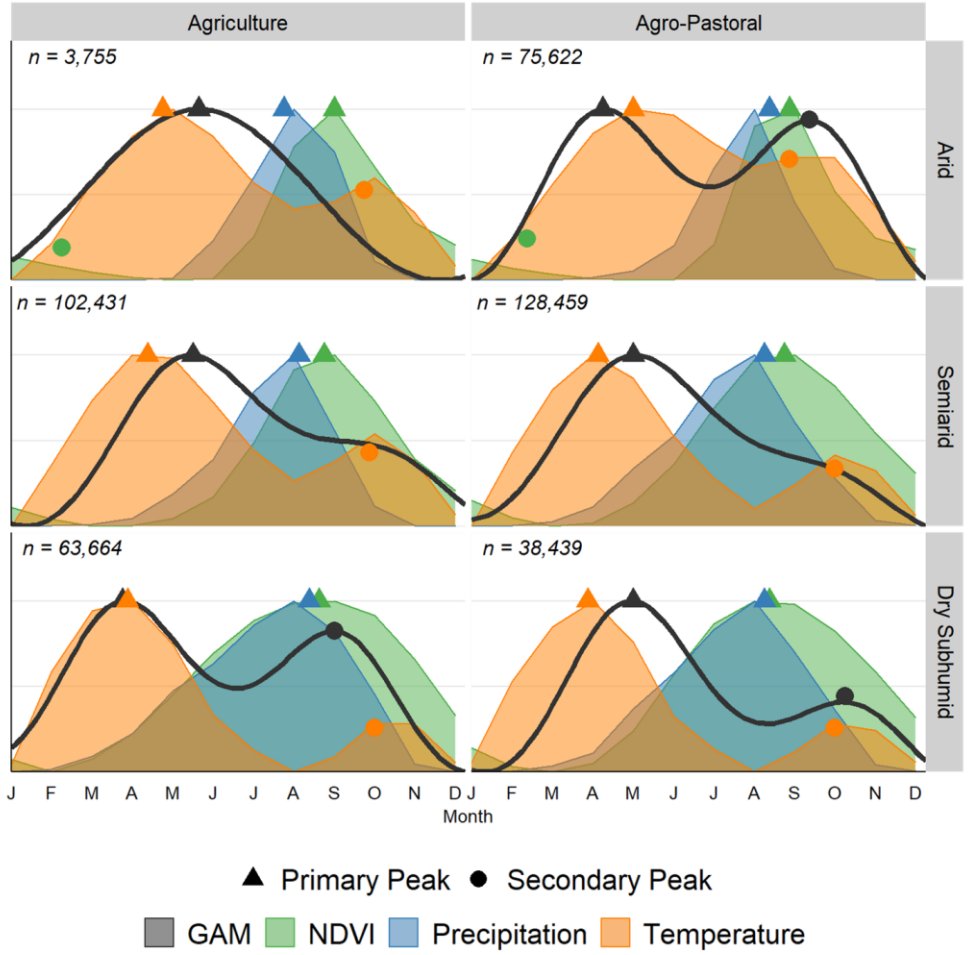
Multiple Harmonic Regression

- Key advantages
  - Simple, established, low technical knowledge barrier to interpretation
  - Can calculate confidence intervals of peak timing and peak values
  - Tractable 'validation' by changing frequency (month/day of survey) and outcome specification (%, binary, counts)
  - Stable to missing data
- Some disadvantages
  - Pre-specification of functional form – also present in non-regression based methods
  - Symmetric shape of curve - incorporates highest penalty for incorrectly predicting peaks

Characteristic	**Unimodal (2π)**		**Bimodal (4π)**
	Gaussian Linear Model	*Log-Linear Model*	*Gaussian Linear or Log-Linear*
Peak Timing (P_T)	Phase shift $\Theta = \arctan\left(\frac{\beta_1}{\beta_2}\right)$ If $\beta_1 > 0$ and $\beta_2 > 0$, $P_T = (\Theta)\frac{M}{2\pi}$ If $\beta_2 < 0$, $P_T = (\Theta + \pi)\frac{M}{2\pi}$ If $\beta_1 < 0$ and $\beta_2 > 0$, $P_T = (\Theta + 2\pi)\frac{M}{2\pi}$		Estimated arithmetically from first, second, and third differences of the predicted seasonal curve. $P_{T,L}$ = Timing of P_L, $P_{T,G}$ = Timing of P_G
95% Confidence Interval of Peak Timing ($CI(\Theta)$)	$Var(\Theta) = \frac{{\beta_1}^2{\sigma_2}^2 + {\beta_2}^2{\sigma_1}^2 - 2\sigma_{\beta_1\beta_2}\beta_1\beta_2}{({\beta_1}^2 + {\beta_2}^2)^2}$ $CI(\Theta) = \Theta \pm 1.96\sqrt{Var(\Theta)}$		Estimated arithmetically from 999 simulations which randomly drop up to 50% of dataset $CI(\widehat{P_T}) = \sum_{n=1}^{n=999} P_{T,G}$

Seasonal wasting in North African drylands

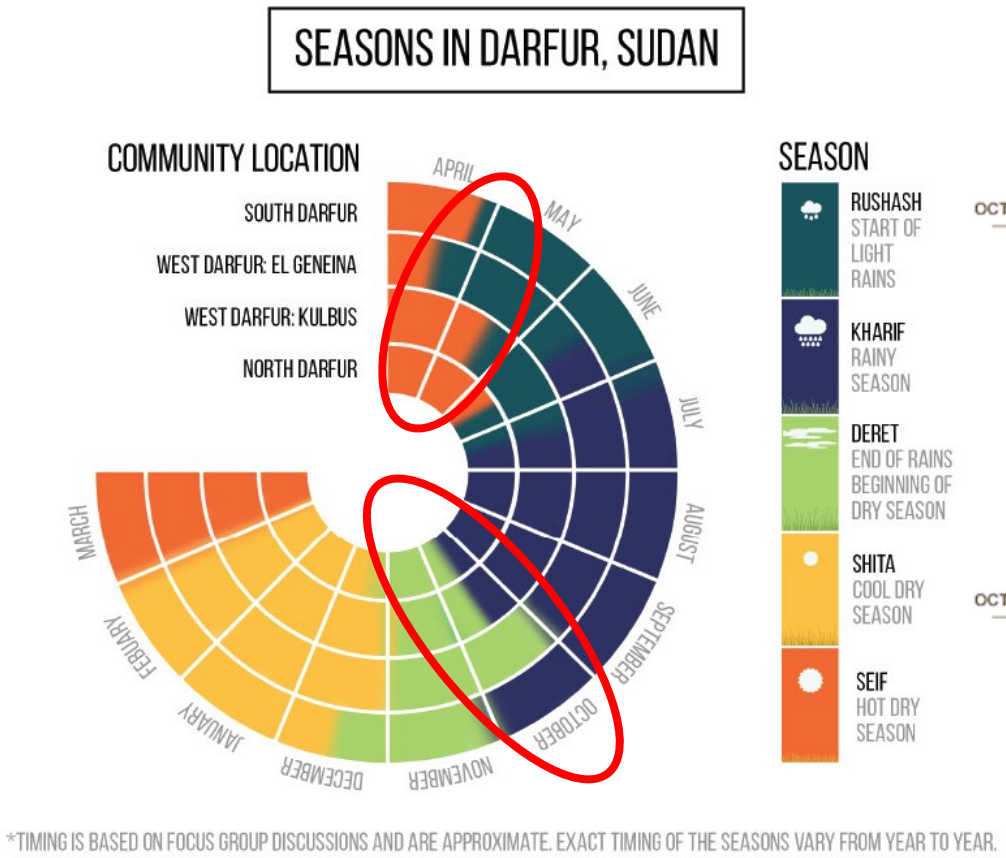


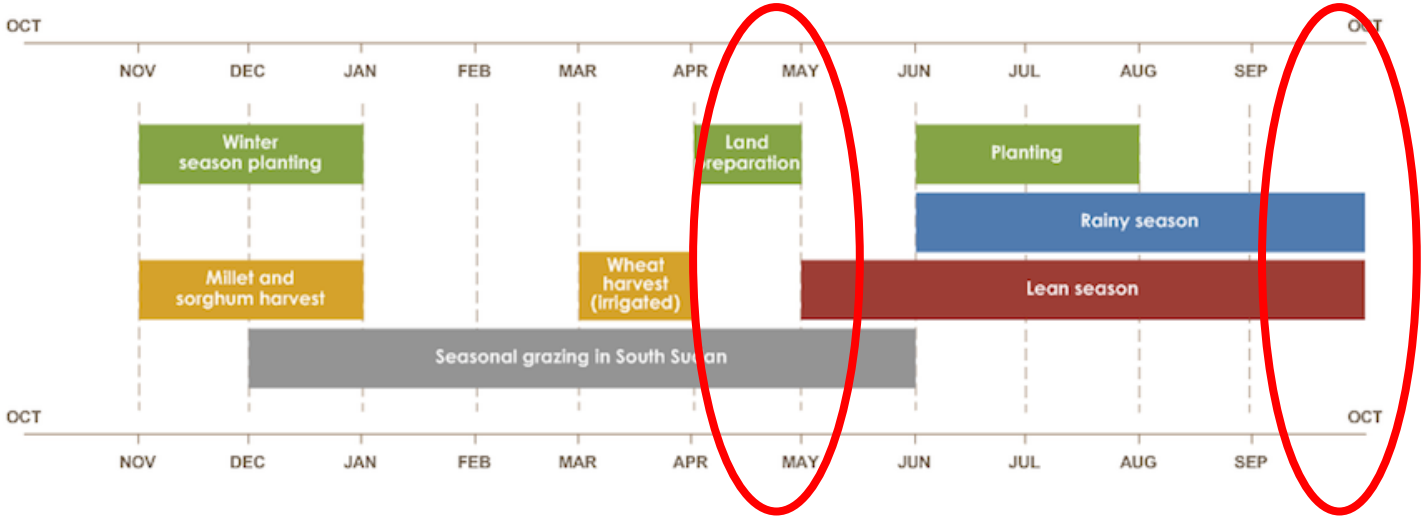


Finding: largest wasting peak is associated with highest temperatures; smaller wasting peak occurs after peaks of rainfall + vegetation

Source: Venkat A., Marshak, A., Young, H., and Naumova, E. N. (2023). Seasonality of Acute Malnutrition in African Drylands: Evidence from 15 Years of SMART Surveys. Food and Nutrition Bulletin.

Local vs. livelihoods-based calendars

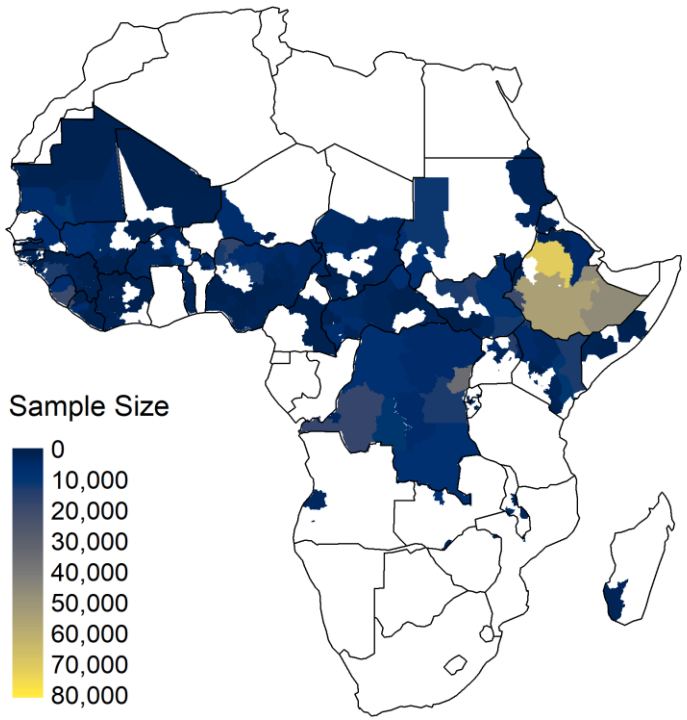




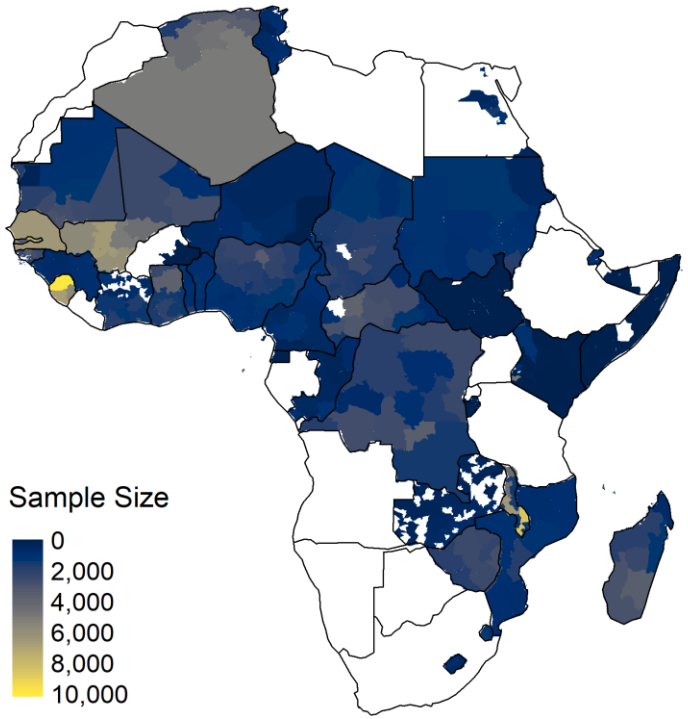
Sources: (*left*) Young, H. & Ismail, M.A. 2019. *Complexity, continuity and change: livelihood resilience in the Darfur region of Sudan*. Disasters, 43(S3): S318–S344; (right) FEWSNET

Scaling up sample size by pooling datasets

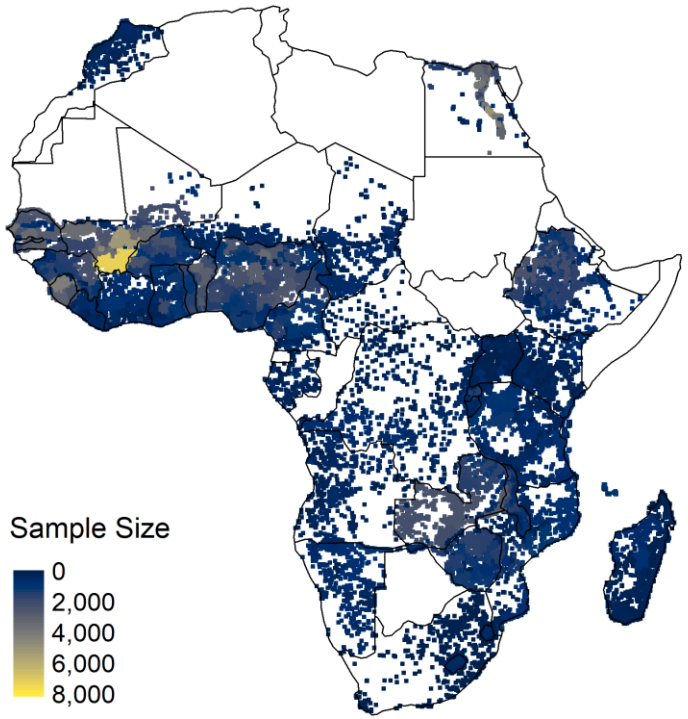
SMART



MICS

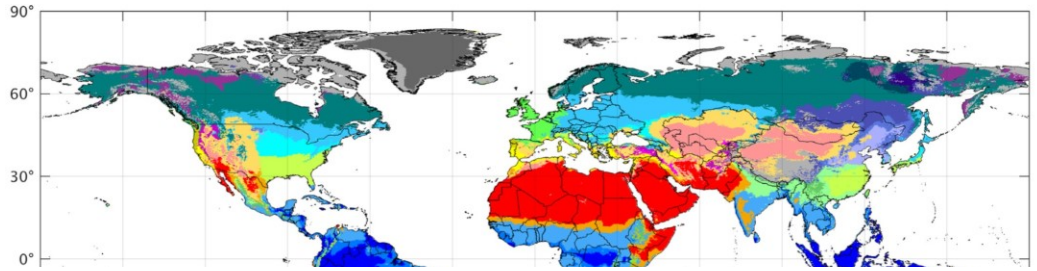


DHS

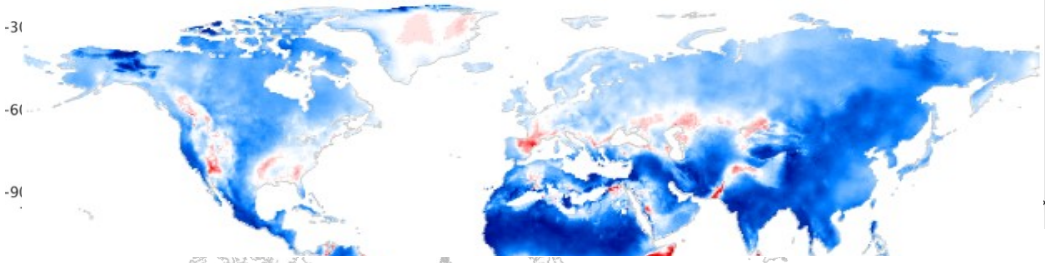


Total *n* = 2,591,633 children in 49 countries

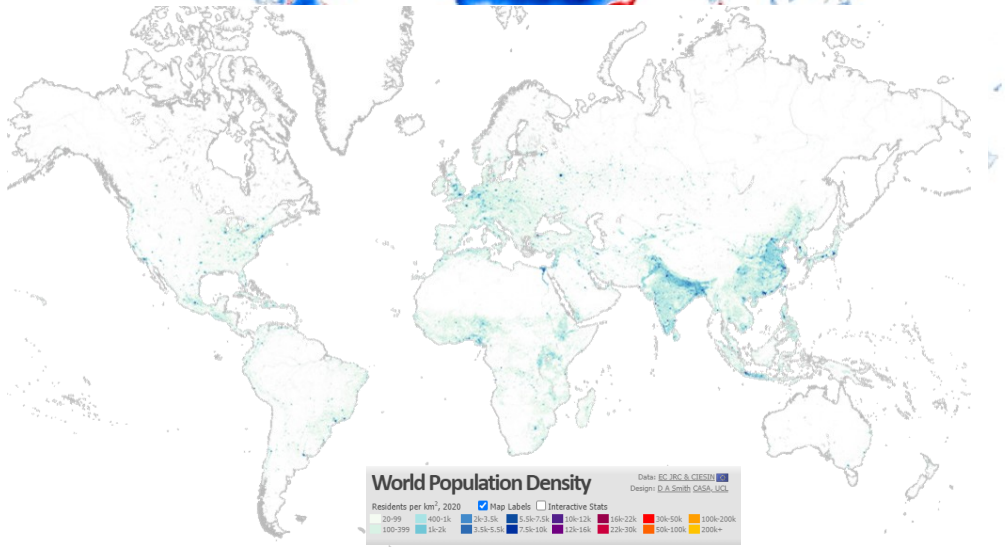
Improving spatial specificity



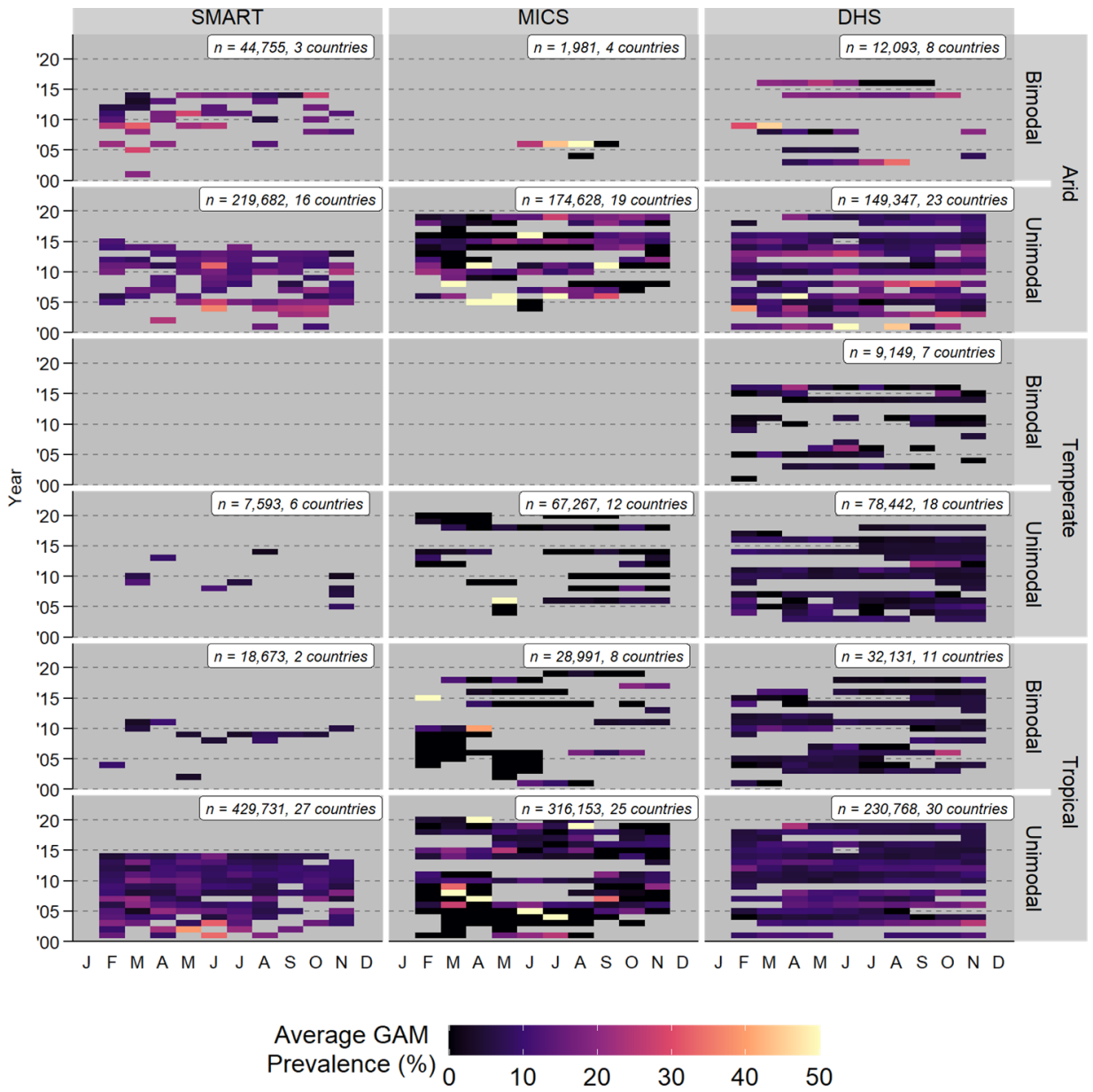
Koppen climate classifications (Beck, 2018)



Global rainfall modality (Knoben, 2019)

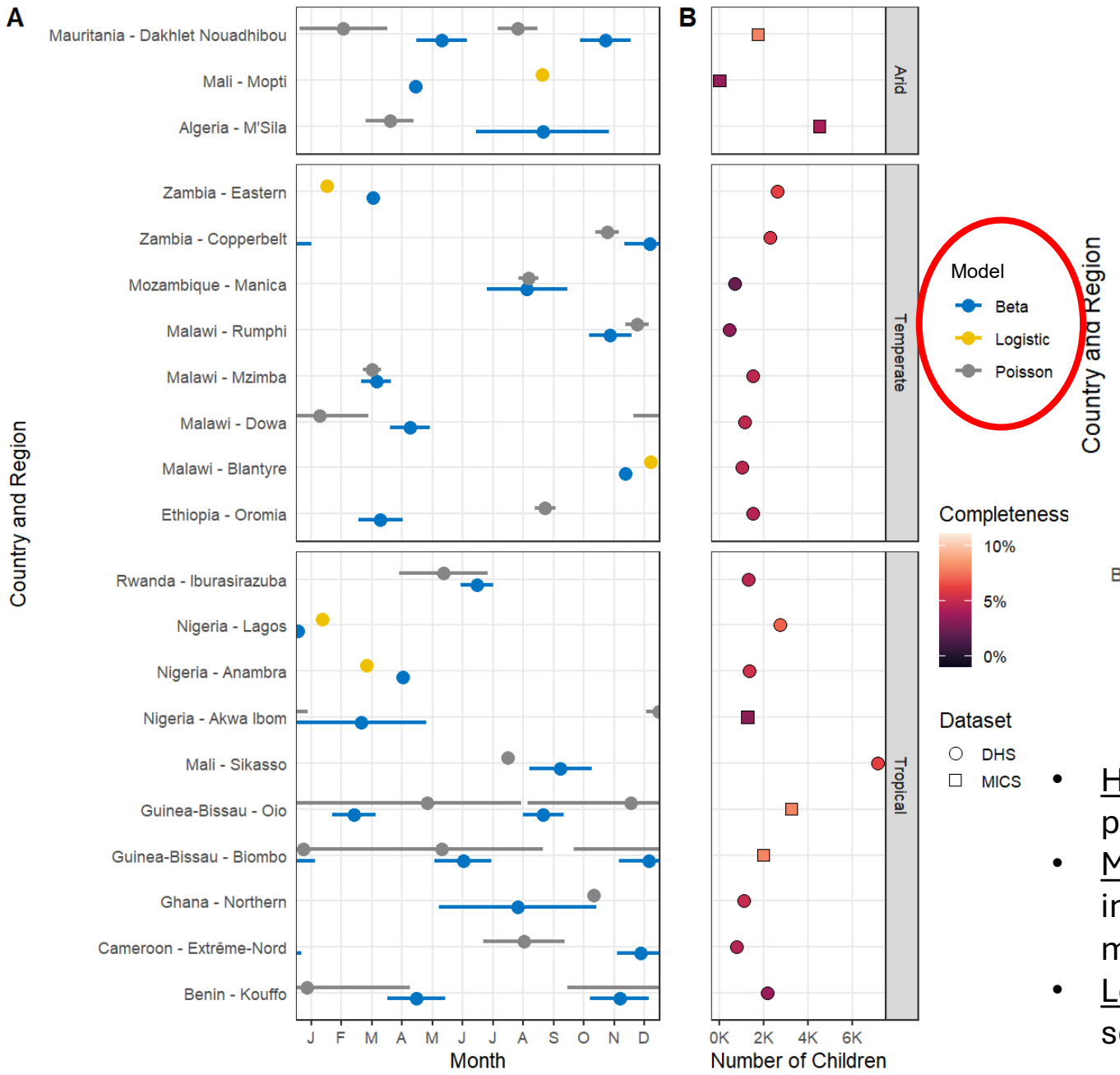


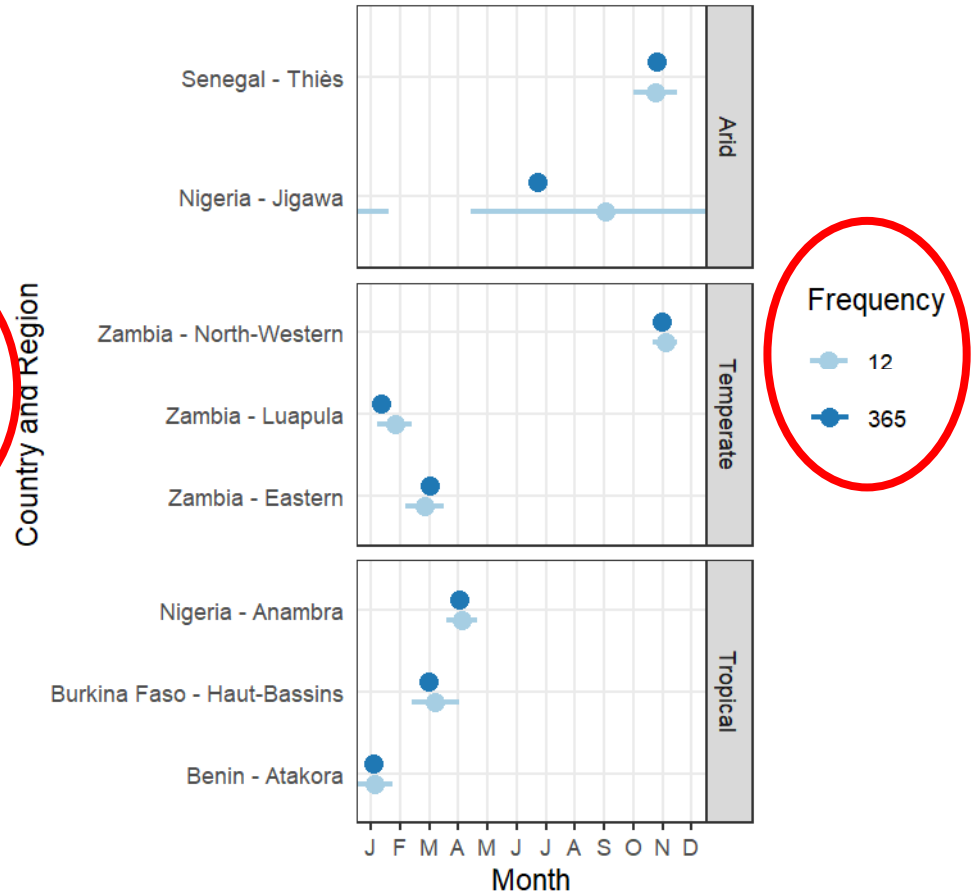
Pixels that have population > 0 (Global Human Settlement Layer)



Completeness

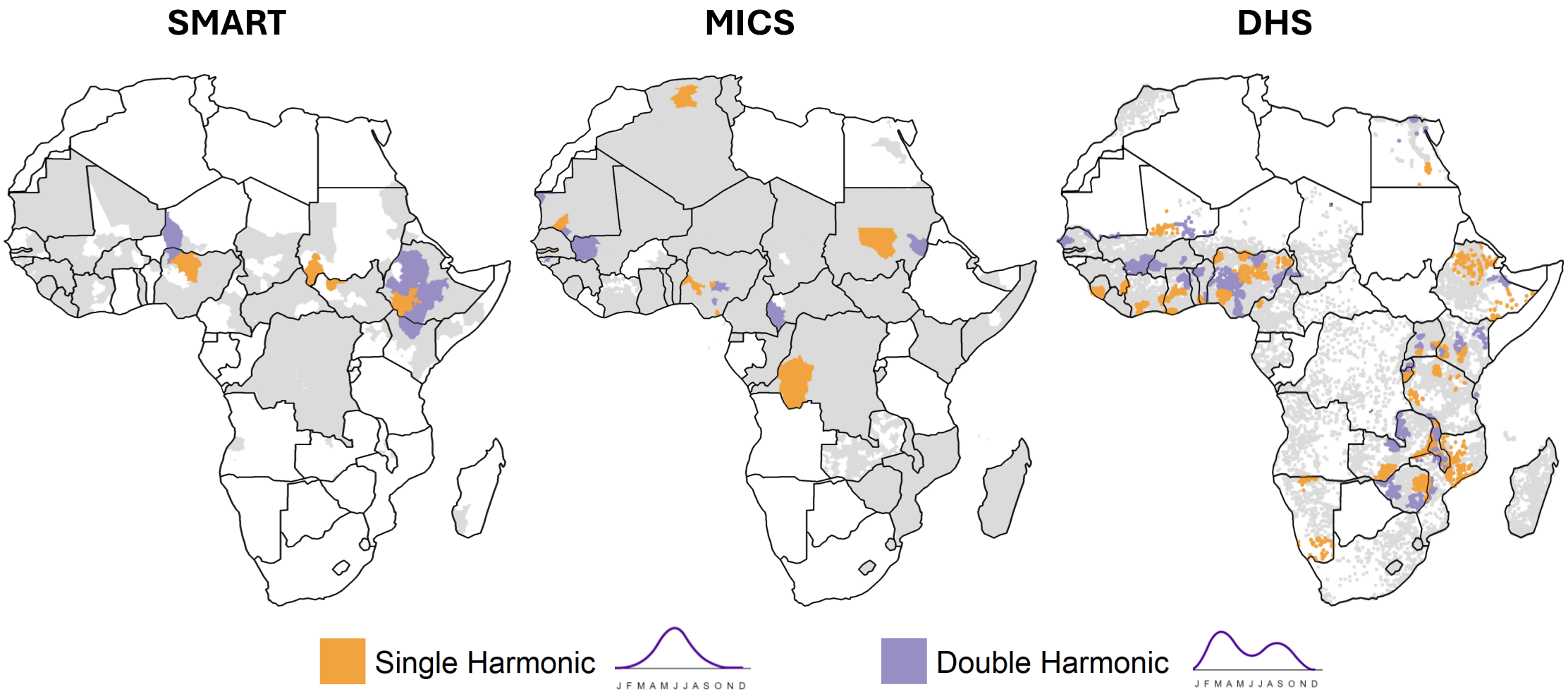
- Despite best efforts, many partitions have no/sparse data
- Minimal empirical completeness of ~7% required for plausible confidence intervals and peak timings
- Threshold of minimum six unique months of data – often lacking at subnational scale



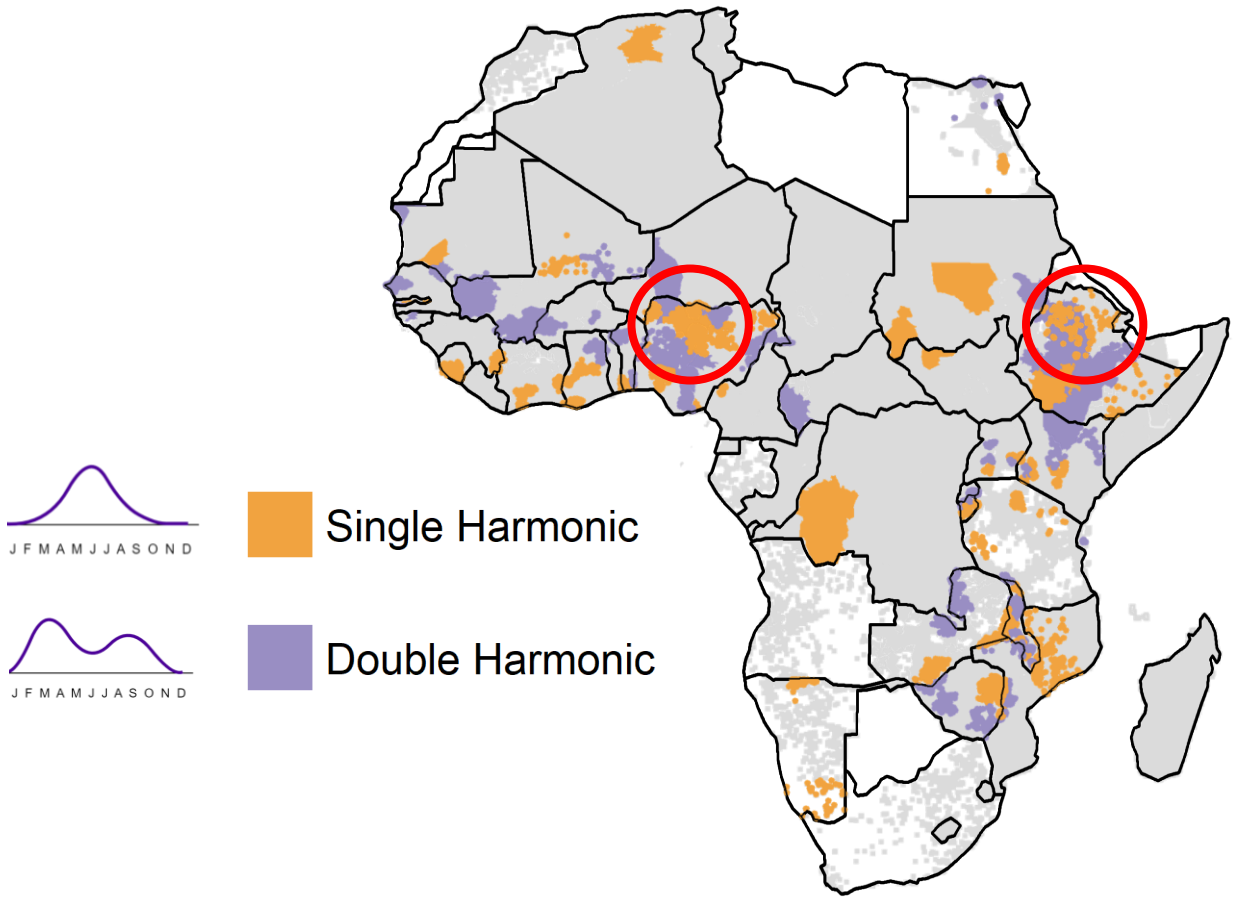


- <u>High Confidence:</u> overlapping confidence intervals of peak timings
- <u>Medium Confidence:</u> non-overlapping confidence intervals but statistically significant seasonality across multiple specifications
- <u>Low Confidence:</u> only one significant estimate of seasonality

Distribution of significant harmonics



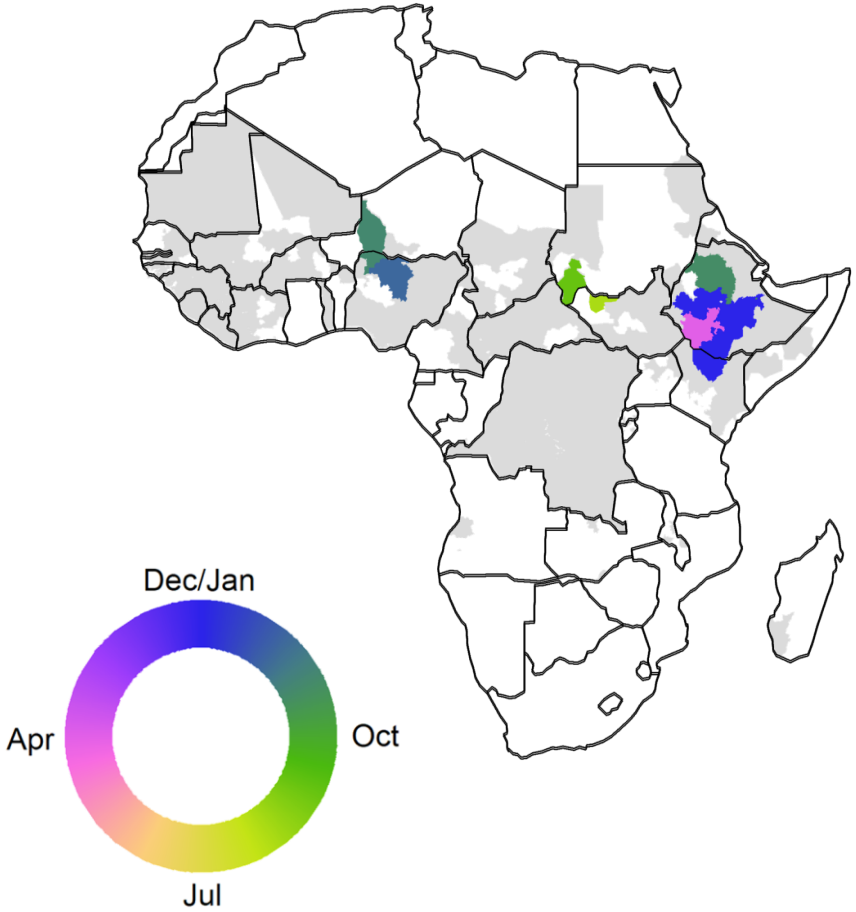
Distribution of significant harmonics



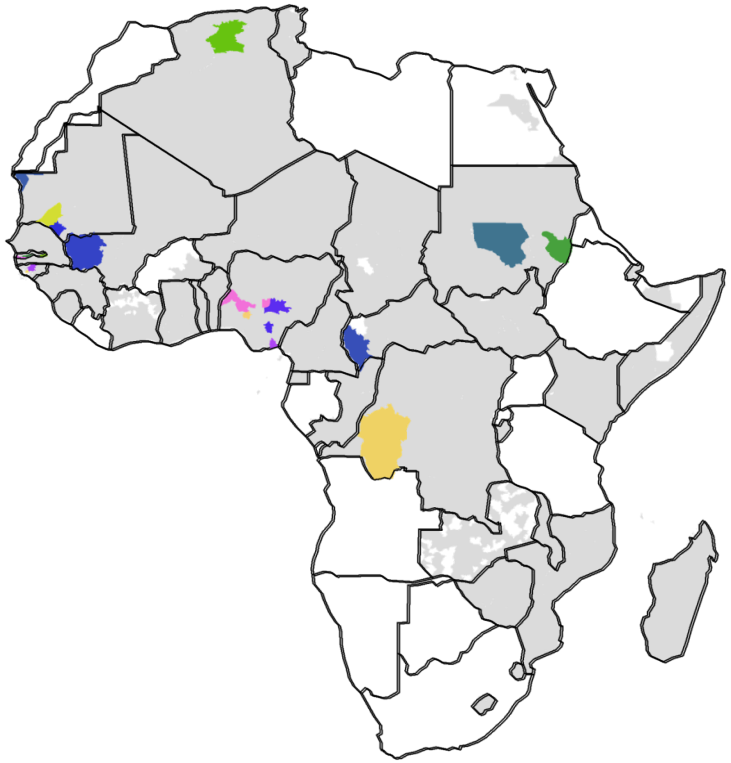
- Mix of significant single and double harmonics indicates heterogeneity
- Datasets can be utilized to validate or distinguish demographic nuances
  - E.g. Northern Nigeria and Ethiopian highlands
- Baseline map for other regions to contribute own analyses to fill in the gap

Distribution of peak timings

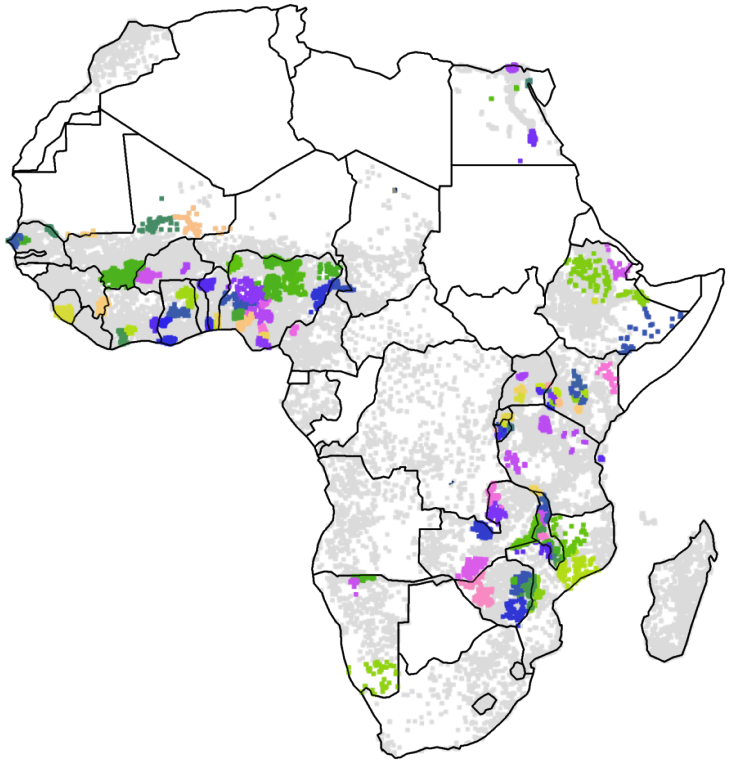
SMART



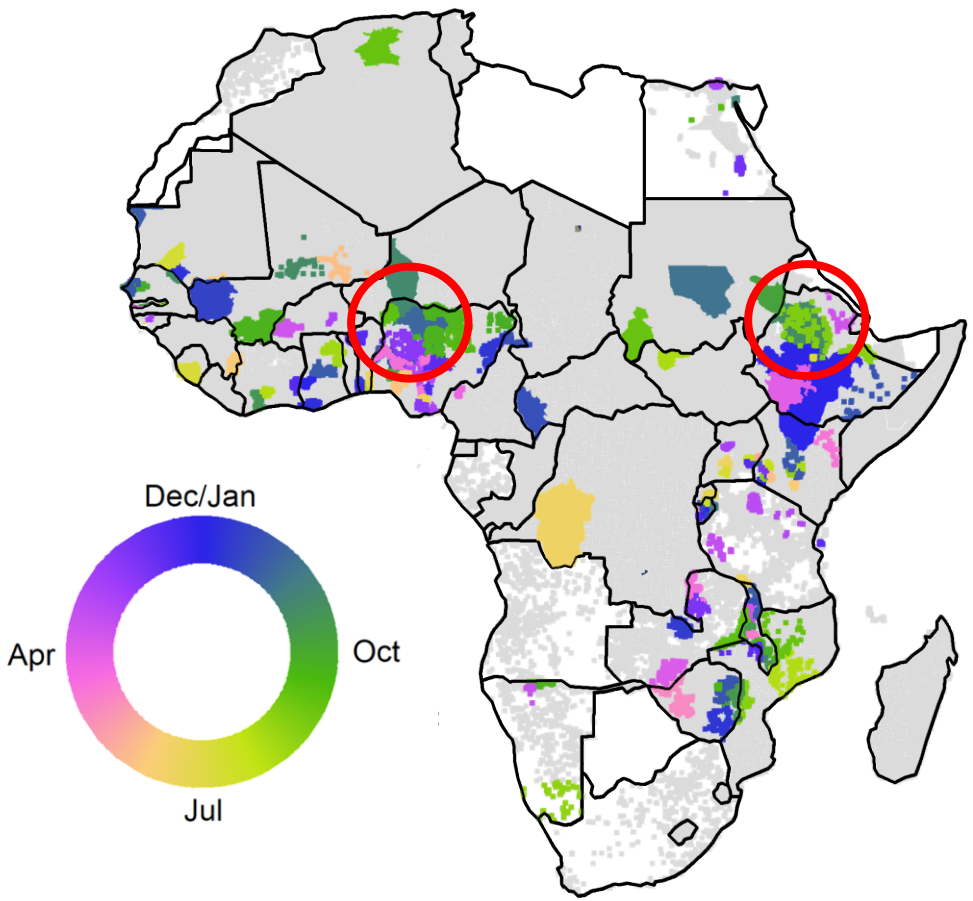
MICS



DHS



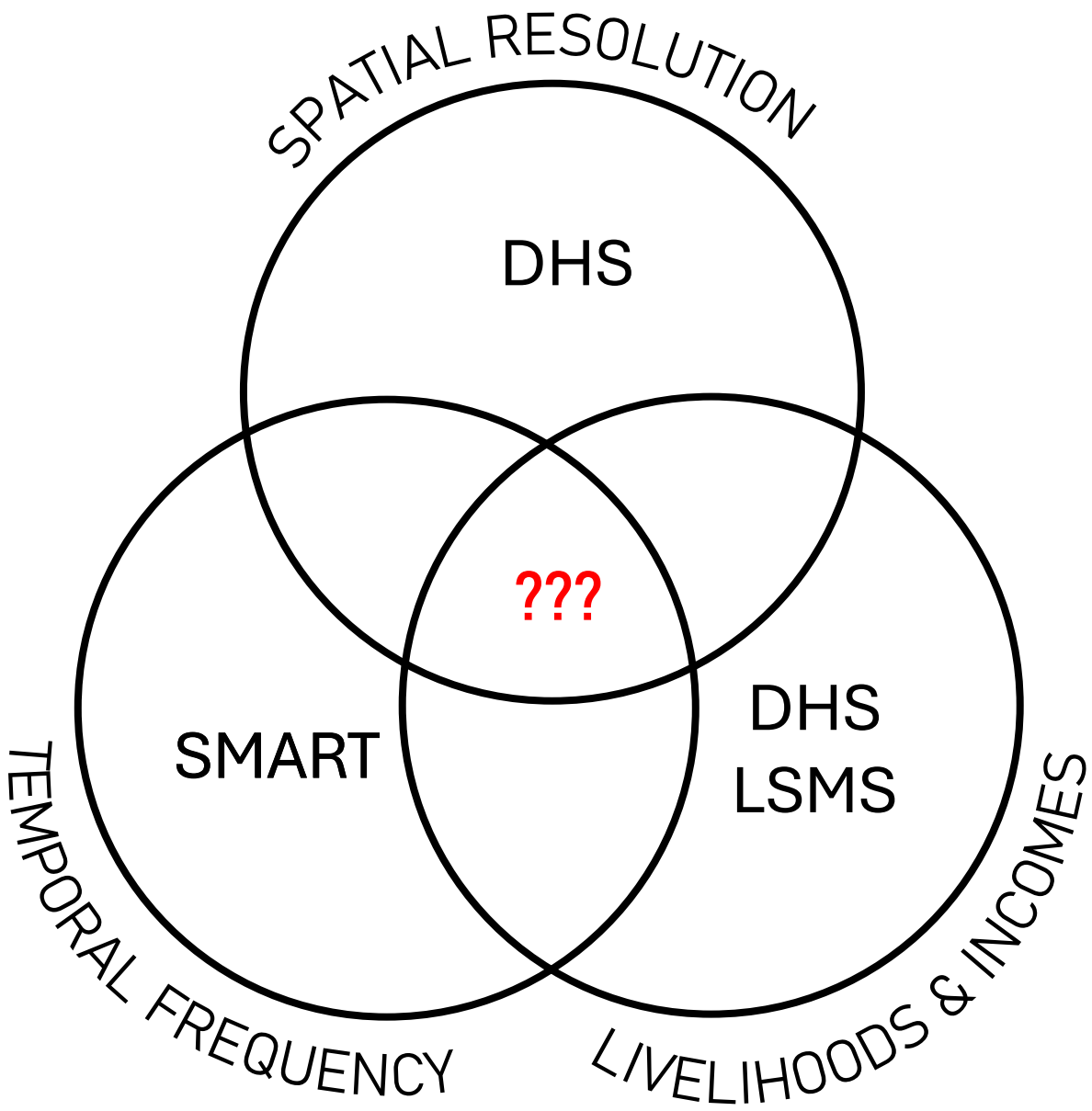
Distribution of peak timings



- Heterogeneity
- Estimated peak values can help prioritize regions for nutrition surveillance
- While colors indicate statistical significance, peak values/magnitudes of wasting may be different, not necessarily actionable

Ongoing challenges

- Spatial data availability
  - MICS, LSMS: most spatial data available after 2005
  - Boundary changes
  - Warehousing of SMART data
- Validation of cross-sectional findings with nutrition surveillance data
- Drivers
  - Food prices/affordability?
  - Livelihood changes?
  - Conflict?

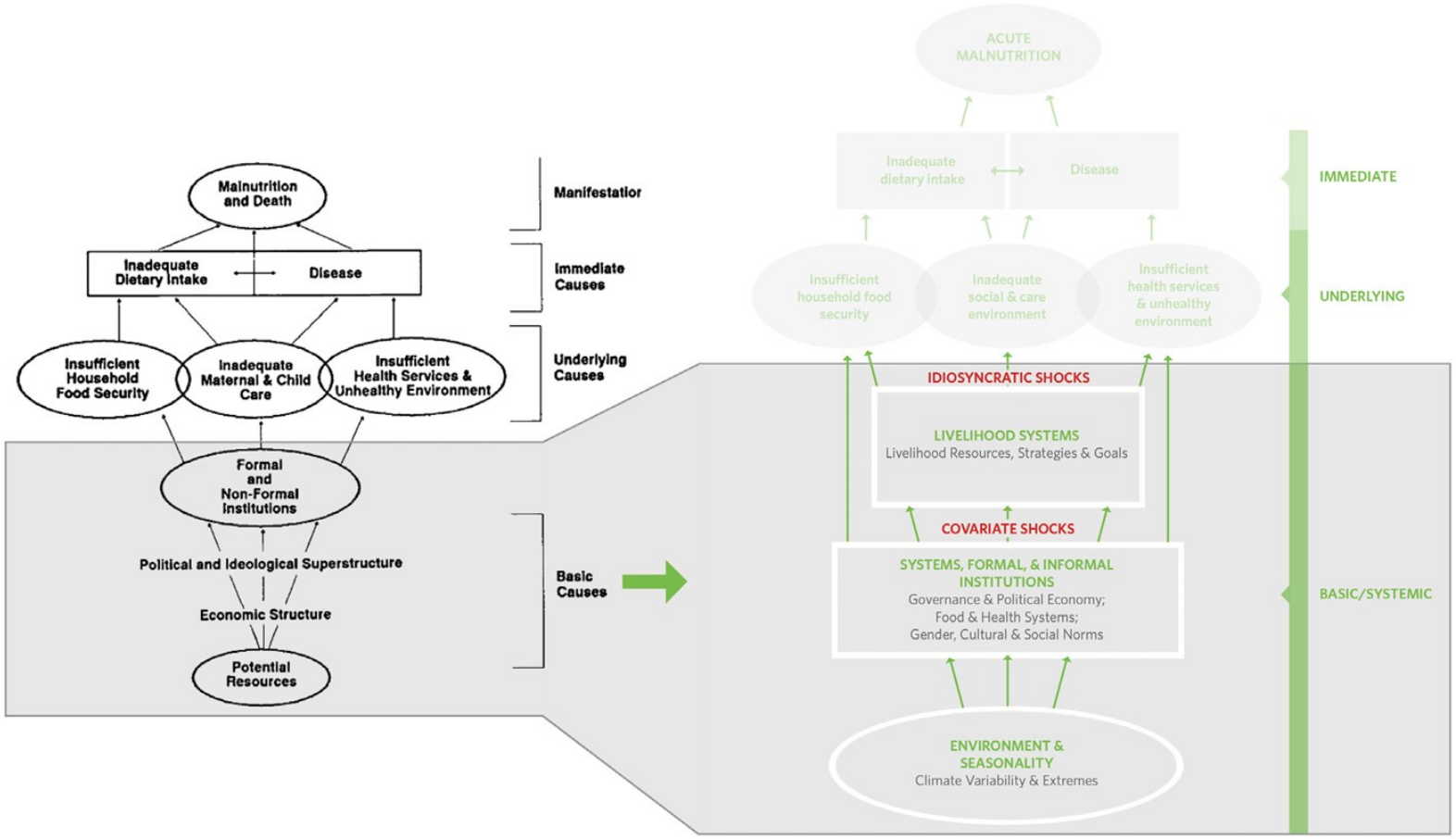


Future Directions

- Seasonality is malleable, contextual, and complex
  - Model results of wasting need to be considered alongside livelihoods, community dynamics, and broader food security indicators
  - Local environmental calendars can help identify pathways, e.g. infectious disease, herd movements, mobility
  - Underlying drivers may vary across partitions
- A renewed seasonality agenda
  - Prioritize high and medium confidence regions for 'ground-truthing' seasonality
  - Improve primary data collection
    - SMART surveys after extreme weather events or disruptions
    - Improve routine nutrition surveillance, particularly in 'medium confidence' regions
  - Develop guidelines to harness continuous temporal variables
  - Disentangle individual or communal vulnerability from access to protective policies and programs

Seasonality as a basic driver of acute malnutrition

F(x)

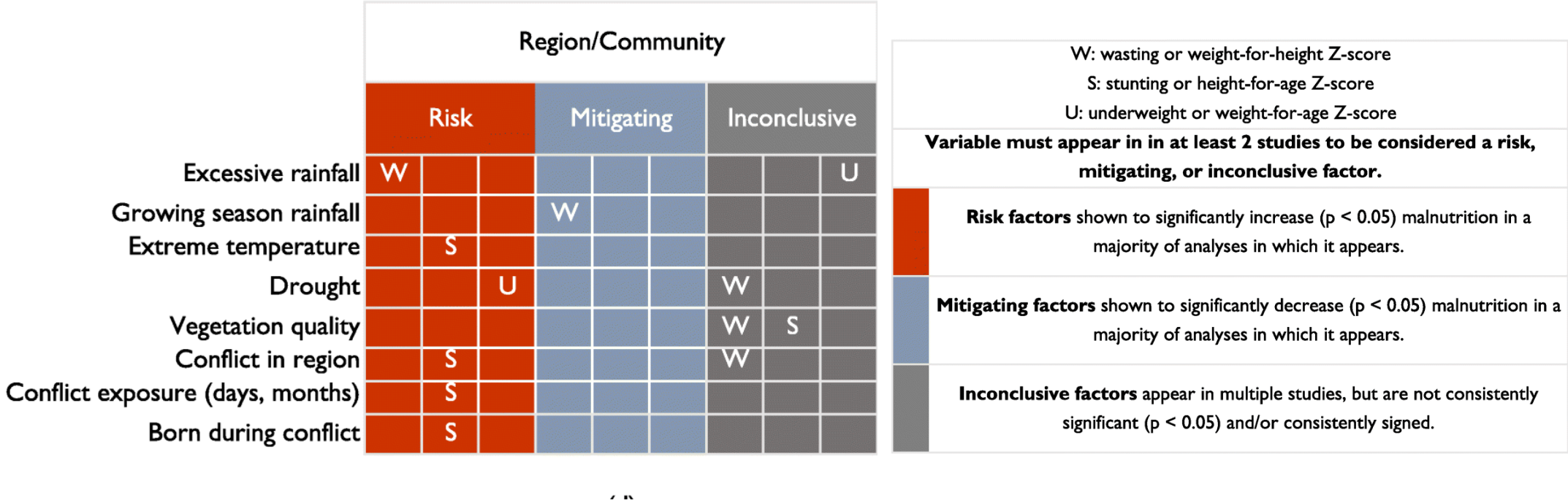


???

x

Source: Young, Y. (2020). *Nutrition in Africa's drylands: A conceptual framework for addressing acute malnutrition*. Boston: Feinstein International Center, Tufts University.

Temporal aggregation extends to covariates



Source: Brown, M. E., Backer, D., Billing, T., White, P., Grace, K., Doocy, S., & Huth, P. (2020). Empirical studies of factors associated with child malnutrition: highlighting the evidence about climate and conflict shocks. *Food Security*, *12*, 1241-1252.

References

- UNICEF, WHO, World Bank. (2023). *Joint Child Malnutrition Estimates*.
- Marshak, A., Venkat, A., Young, H., & Naumova, E. N. (2021). *How seasonality of malnutrition is measured and analyzed*. International Journal of Environmental Research and Public Health, 18(4), 1828.
- Naumova, EN and MacNeill, IB (2006) *Seasonality assessment for biosurveillance systems*. Advances in Statistical Methods for the Health Sciences. Boston, MA: Birkhauser, pp 437–450.
- Venkat A., Marshak, A., Young, H., and Naumova, E. N. (2023). Seasonality of Acute Malnutrition in African Drylands: Evidence from 15 Years of SMART Surveys. Food and Nutrition Bulletin.
- Young, H. & Ismail, M.A. 2019. *Complexity, continuity and change: livelihood resilience in the Darfur region of Sudan*. Disasters, 43(S3): S318–S344.
- Young, Y. (2020). *Nutrition in Africa's drylands: A conceptual framework for addressing acute malnutrition*. Boston: Feinstein International Center, Tufts University.
- Brown, M. E., Backer, D., Billing, T., White, P., Grace, K., Doocy, S., & Huth, P. (2020). Empirical studies of factors associated with child malnutrition: highlighting the evidence about climate and conflict shocks. *Food Security*, *12*, 1241-1252.

Thank you! Questions?

Aishwarya Venkat

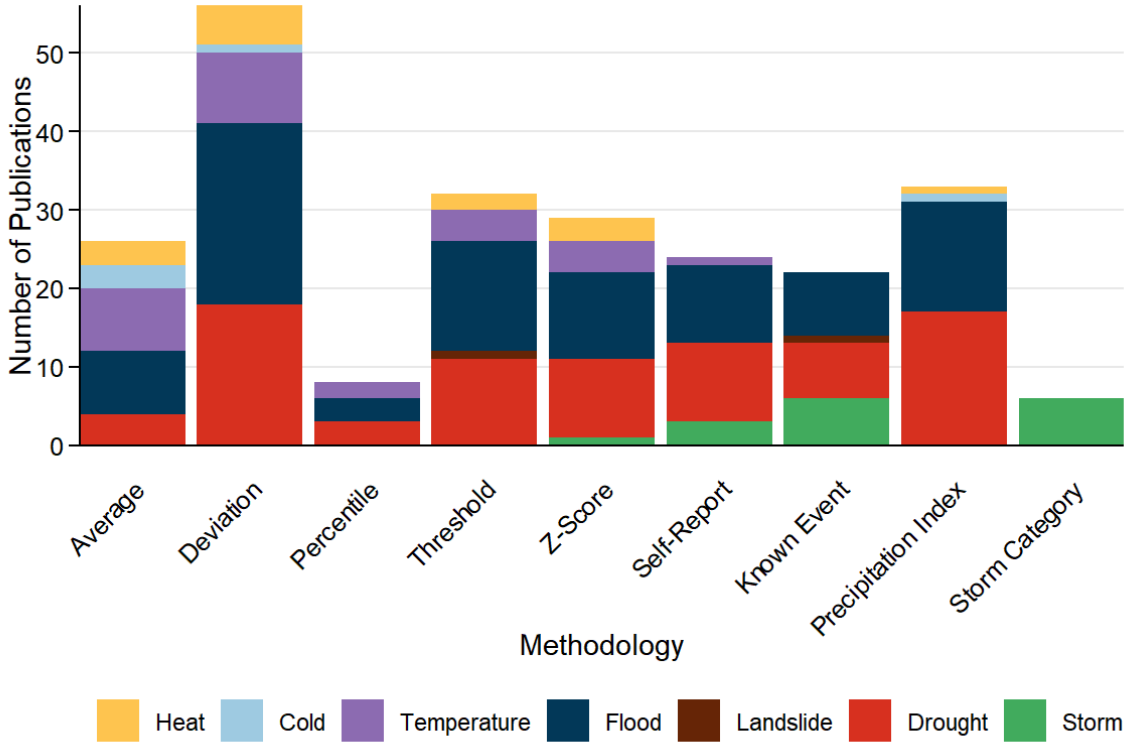
Aishwarya.Venkat@tufts.edu | aish-venkat.github.io

Tufts UNIVERSITY | Gerald J. and Dorothy R. Friedman School of Nutrition Science and Policy

InForMID
Initiative for the Forecasting and Modeling of Infectious Diseases

Feinstein International Center

Beyond time – what is the right measure?



Source: Venkat, A. (2024). *Chapter 2: Measures of Climate Hazards and Human Health*. In: Climate and Health: Extreme Events, Food Systems, and Nutrition. [Doctoral Dissertation]. Tufts University, Boston, MA, USA.

- Relevant criteria
  - Remotely sensed
  - Long time series
  - High spatial and temporal resolution
- Considerations
  - Global rules vs. local range of exposures
  - What is environmental or biological basis of threshold selection?

Extreme weather

- Operational definitions
  - **Heatwave**: values exceeding 95th percentile of monthly maximum temperature, calculated from Terraclimate (Abatzoglou et al, 2018)
  - **Coldwave:** values below 5th percentile of monthly minimum temperature, calculated from Terraclimate (Abatzoglou et al, 2018)
  - **Flood**: values exceeding 95th percentile of 1-month SPI time series, calculated from CHIRPS (Funk et al, 2015)
  - **Drought**: values below 5th percentile of 6-month Standardized Precipitation and Evapotranspiration Index (Dalezios et al, 2017; Vicente-Serrano et al, 2010), calculated from CHIRPS monthly precipitation (Funk et al, 2015) and MOD11C3 v061 monthly temperature (Wan et al, 2021)
  - **Storm**: average radius of storm-force winds or higher, from IBTrACS (Knapp et al, 2010)