# You keep using Coefficient of Variation
# *I do not think it means what you think it means*

Devan Allen McGranahan [1] Samuel D. Fuhlendorf [2]

[1]USDA Agricultural Research Service, Miles City, MT; [2]Oklahoma State University, Stillwater, OK

## Overview

- Variability—statistically, *variance*, the second data moment—can be a more important ecological property than the mean
- Coefficient of Variation is popular but problematic under multiple sources of variability
- **We demonstrate *variance partitioning* as a spatially-explicit alternative to CV**

## Defining variability

### Coefficient of Variation

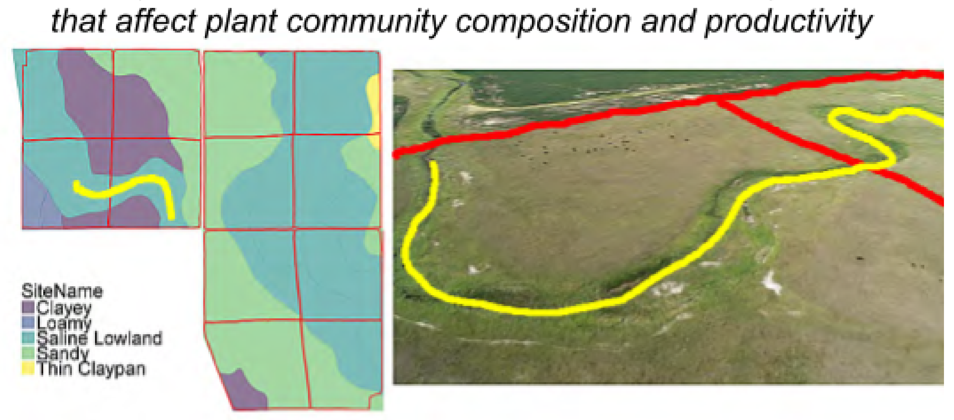
- Variability can be abstract—simple measures are understandable
- Coefficient of Variation is calculated as $CV = \frac{\sigma}{\mu} \cdot 100$ where $\sigma$ = standard deviation and $\mu$ = mean

### Handling multiple sources of variability

- *CV is only reliable when mean is constant*
- When variability effects differences in both $\mu$ and $\sigma$, CV is misleading
- Both *inherent* and *disturbance-driven heterogeneity* can case variability in $\mu$ and $\sigma$:

**Inherent environmental heterogeneity**

*Ecological sites summarize topoedaphic variability in soils that affect plant community composition and productivity*




**Disturbance-driven heterogeneity**

*Spatially-discrete burns create a mosaic of patches with asynchronous plant successional stages*

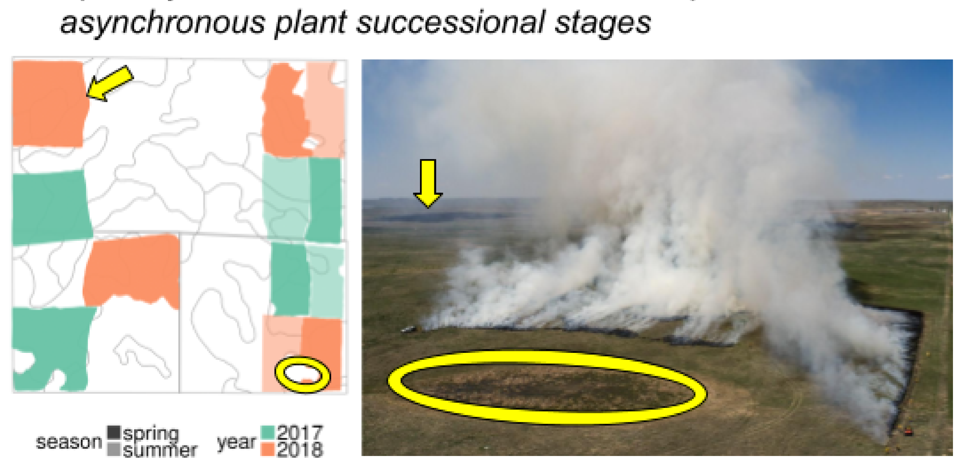



Fig. 1: Examples of heterogeneity sources.

## Data and methods

*Productivity Scenarios* defined by quantiles of the range of productivity across US National Grasslands as determined by Rangeland Analysis Platform data on perennial + annual herbaceous biomass

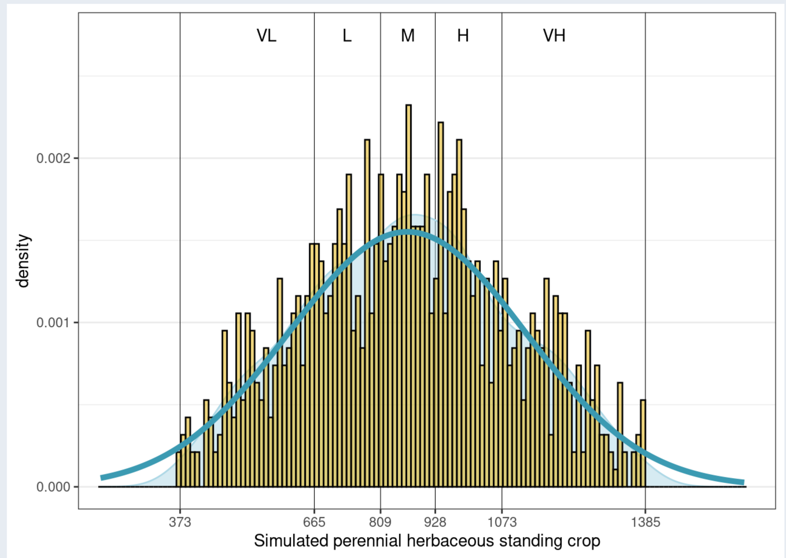



Fig. 2: Productivity classes (kg ha$^{-1}$).

| Productivity scenario | Mean | SD | CV |
|---|---|---|---|
| Very Low (VL) | 553 | 78.8 | 0.14 |
| Low (L) | 739 | 42.3 | 0.06 |
| Medium (M) | 869 | 33.6 | 0.04 |
| High (H) | 993 | 39.4 | 0.04 |
| Very High (VH) | 1194 | 81.7 | 0.07 |

### Non-spatial simulations

Calculate CV for each productivity scenario:

- Actual mean ($\mu$) and standard deviation ($\sigma$) from data in Fig. 2.
- Constant $\mu$ across productivity classes (VH & VL) with $\sigma$ from each class
- Constant $\sigma$ across productivity classes (VH & VL) with $\mu$ from each class

### Spatial simulations

Four landscape scenarios (Fig. 3):

- Randomly-distributed productivity
- Disturbance only: Time-since-fire gradient
- Inherent variability only: soil types
- Both disturbance gradient and soil types

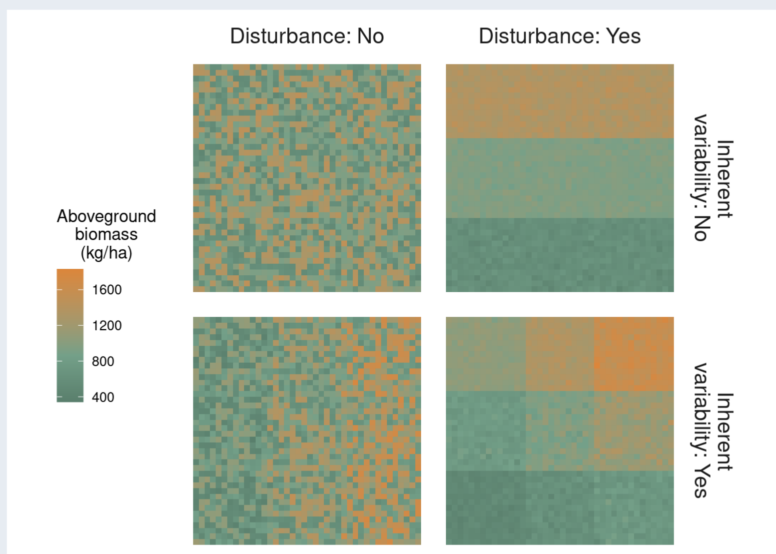



Fig. 3: Four simulated landscapes.

## Problems with CV

### Non-spatial example

- When means vary, *CV declines as mean increases*
- Low mean, high CV relationship *independent of SD*

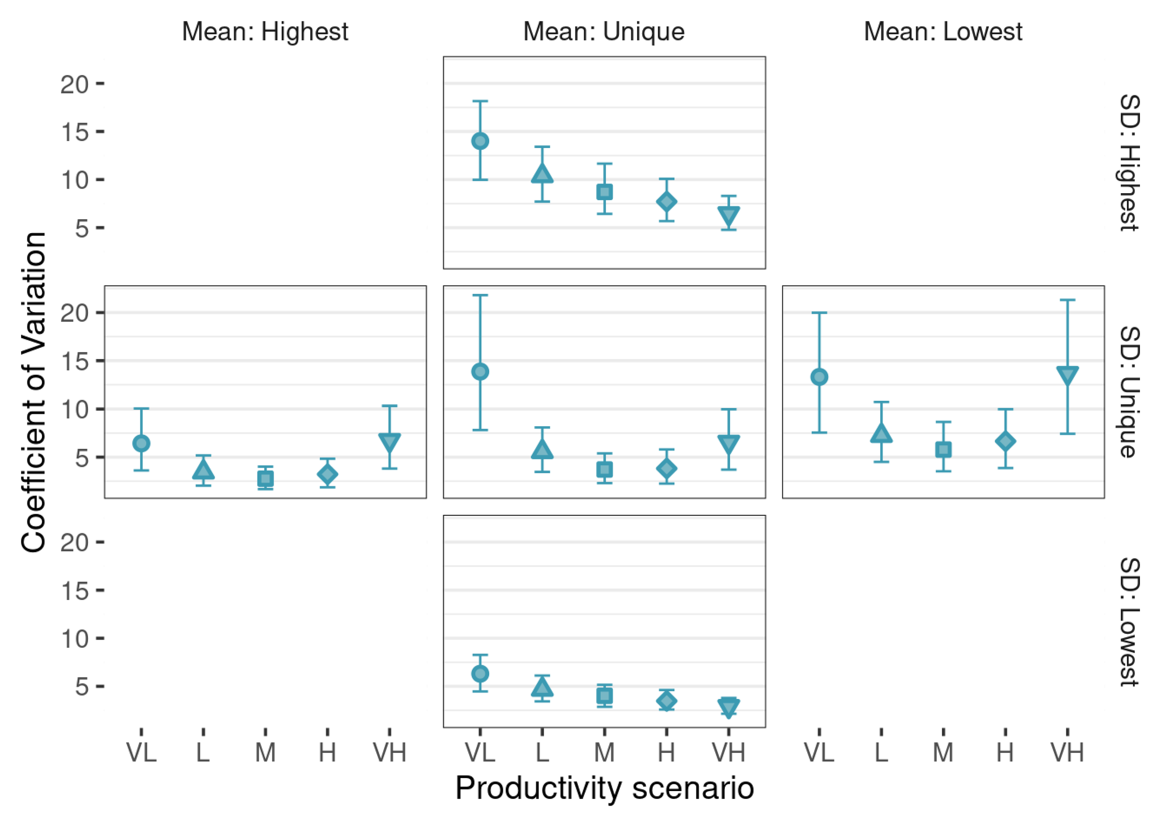



Fig. 4: CV calculations from non-spatial example. *Center:* CV with actual mean and SD for each productivity scenario (Fig. 2). *L & R:* CV with unique SD for each scenario and constant mean, highest and lowest classes. *Top & Bottom:* CV with unique mean for each scenario and constant SD, highest and lowest classes.

### Spatial example

- CV is not sensitive to landscape-level heterogeneity
- Similar trends as in Fig. 3 when homoscedastic

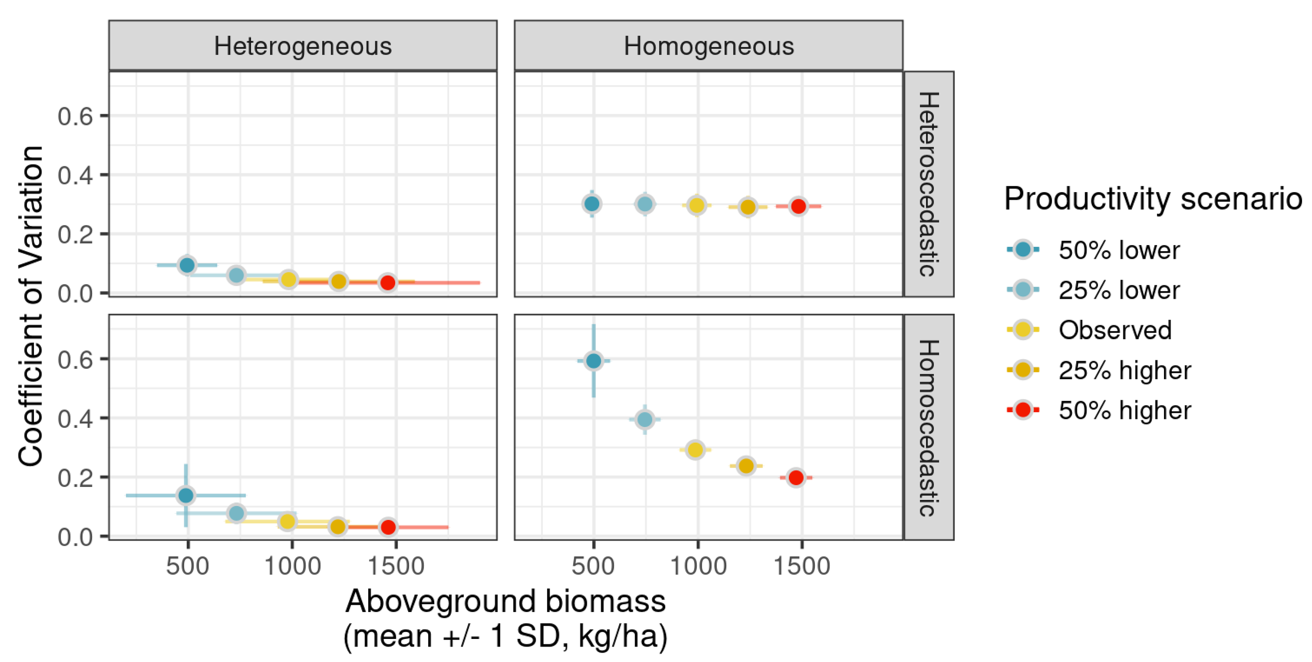



Fig. 5: *Columns:* CV across heterogeneous and homogeneous landscapes (Fig. 3). *Rows:* Assuming the variance in productivity data remain constant as means increase (homoscedastic, and assumption of linear regression ) vs increasing variance with mean (heteroscedastic, a reality of many ecological data.)

## Solution: Variance Partitioning

- Random-effect regression parses variance into spatially-relevant scales
- Spatial heterogeneity *quantified as patch contrast*—the degree of difference among landscape units
- Coarse distinction between heterogeneous (patchy) and homogeneous (uniform) landscapes
- Fine distinction among degrees of variability among patches

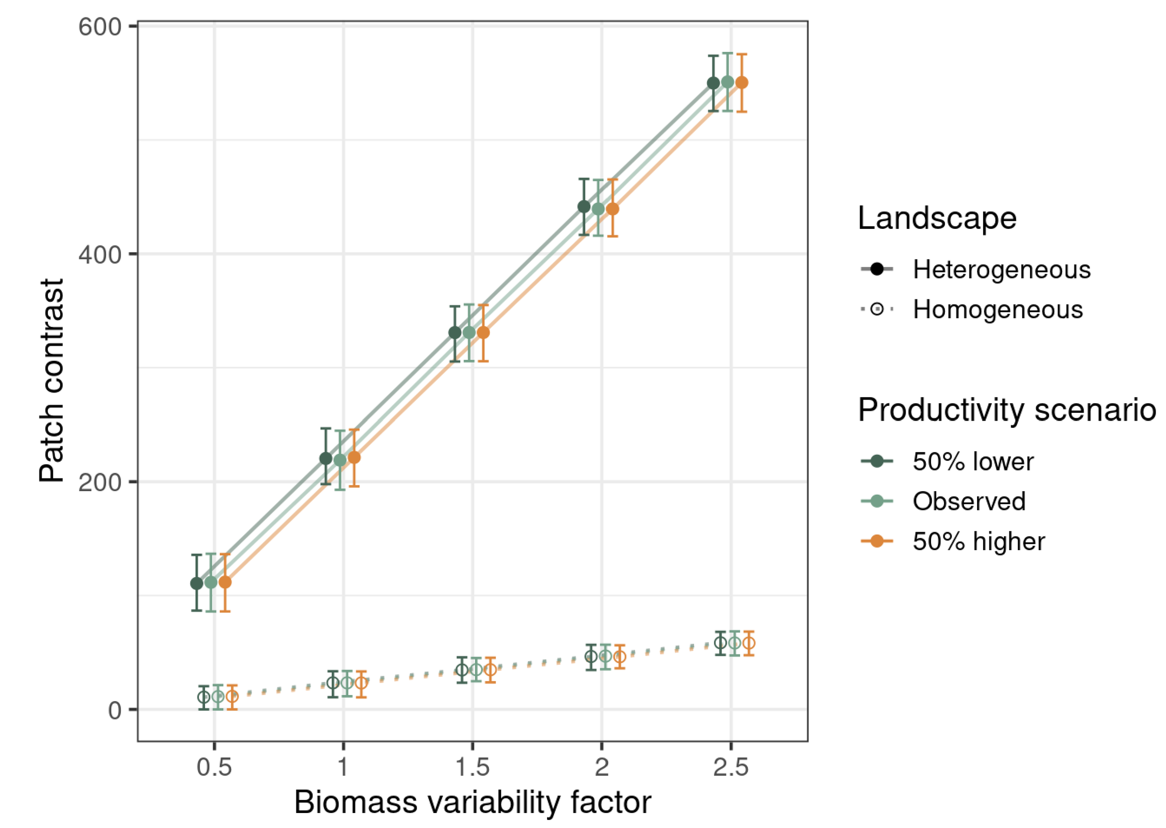



Fig. 6: The proportion variance attributable to the patch term in a random-effect regression—*Patch Contrast*—within heterogeneous and homogeneous landscapes from Fig. 3.

## Avoid CV when means vary!

- CV is only useful to compare variability when mean values are constant.
- CV is not sensitive to spatial heterogeneity.
- Variance partitioning is robust for data with spatial structure—always a good idea when sampling heterogeneous environments.
- Variance partitioning can be accomplished with `proc GLIMMIX` in SAS, `lme4::(g)lmer` in **R**. The **R** package `rptR` also performs variance decomposition.