Model Output Statistics (MOS) - Objective Interpretation of NWP Model Output

University of Maryland – April 4, 2012

Mark S. Antolik
Meteorological Development Laboratory
Statistical Modeling Branch
NOAA/National Weather Service
Silver Spring, MD

(301) 713-0023 ext. 110
email: mark.antolik@noaa.gov

MOS Operational System “Fun Facts”

With apologies to David Letterman, of course!

- 9 million regression equations
- 75 million forecasts per day
- 1200 products sent daily

- 400,000 lines of code – mostly FORTRAN
- 180 min. supercomputer time daily

- All developed and maintained by ~ ~~12~~ 8 MDL / SMB meteorologists!

OUTLINE

1. Why objective statistical guidance?
2. What is MOS?
   - Definition and characteristics
   - The "traditional" MOS product suite (GFS, NAM)
   - Other additions to the lineup
3. Simple regression examples / REEP
4. Development strategy - MOS in the "real world"
5. Verification
6. Dealing with NWP model changes
7. Where we're going – GMOS and the future

WHY STATISTICAL GUIDANCE?

- **Add value to direct NWP model output**
  - **Objectively interpret model**
    - **- remove systematic biases**
    - **- quantify uncertainty**
  - **Predict what the model does not**
  - **Produce site-specific forecasts**
    - **(i.e. a "downscaling" technique)**
- **Assist forecasters**
  - **"First Guess" for expected local conditions**
  - **"Built-in" model/climo memory for new staff**

What is MOS?

MODEL OUTPUT STATISTICS (MOS)

Relates observed weather elements **(PREDICTANDS)** to appropriate variables **(PREDICTORS)** via a statistical approach.

Predictors are obtained from:

1. Numerical Weather Prediction (NWP) Model Forecasts
2. Prior Surface Weather Observations
3. Geoclimatic Information

Current Statistical Method:

MULTIPLE LINEAR REGRESSION

(Forward Selection)

MODEL OUTPUT STATISTICS (MOS)

Properties

- Mathematically simple, yet powerful
- Need historical record of observations at forecast points
  (Hopefully a long, stable one!)
- Equations are applied to future run of similar forecast model

MODEL OUTPUT STATISTICS (MOS)

Properties (cont.)

- Non-linearity can be modeled by using NWP variables and transformations

- Probability forecasts possible from a single run of NWP model

- Other statistical methods can be used
  e.g. Polynomial or logistic regression;
  Neural networks

MODEL OUTPUT STATISTICS (MOS)

- **ADVANTAGES**
  - Recognition of model predictability
  - Removal of some systematic model bias
  - Optimal predictor selection
  - Reliable probabilities
  - Specific element and site forecasts

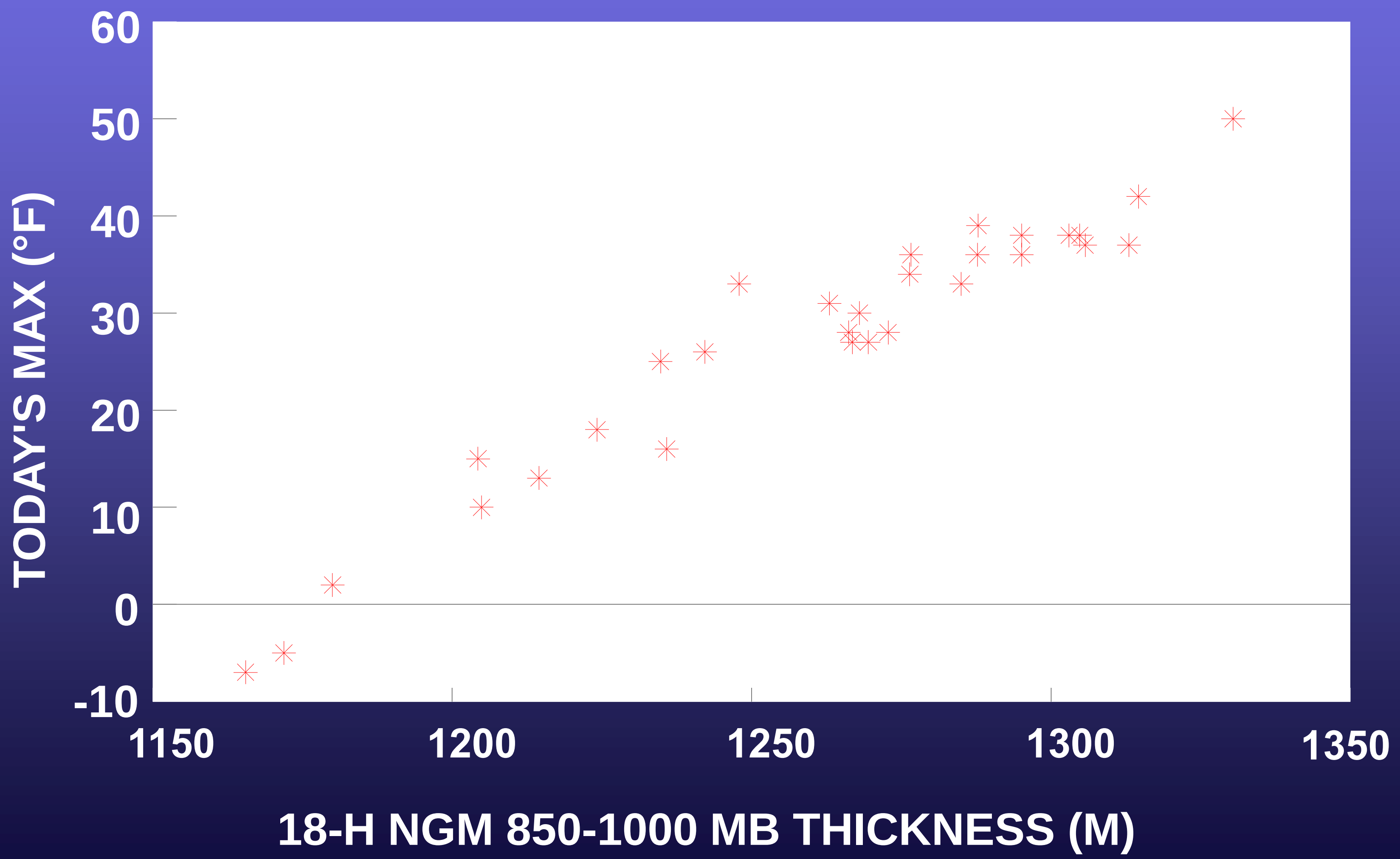
- **DISADVANTAGES**
  - Short samples
  - Changing NWP models
  - Availability & quality of observations

Application of Linear Regression to MOS Development

MOS LINEAR REGRESSION

JANUARY 1 - JANUARY 30, 1994 0000 UTC

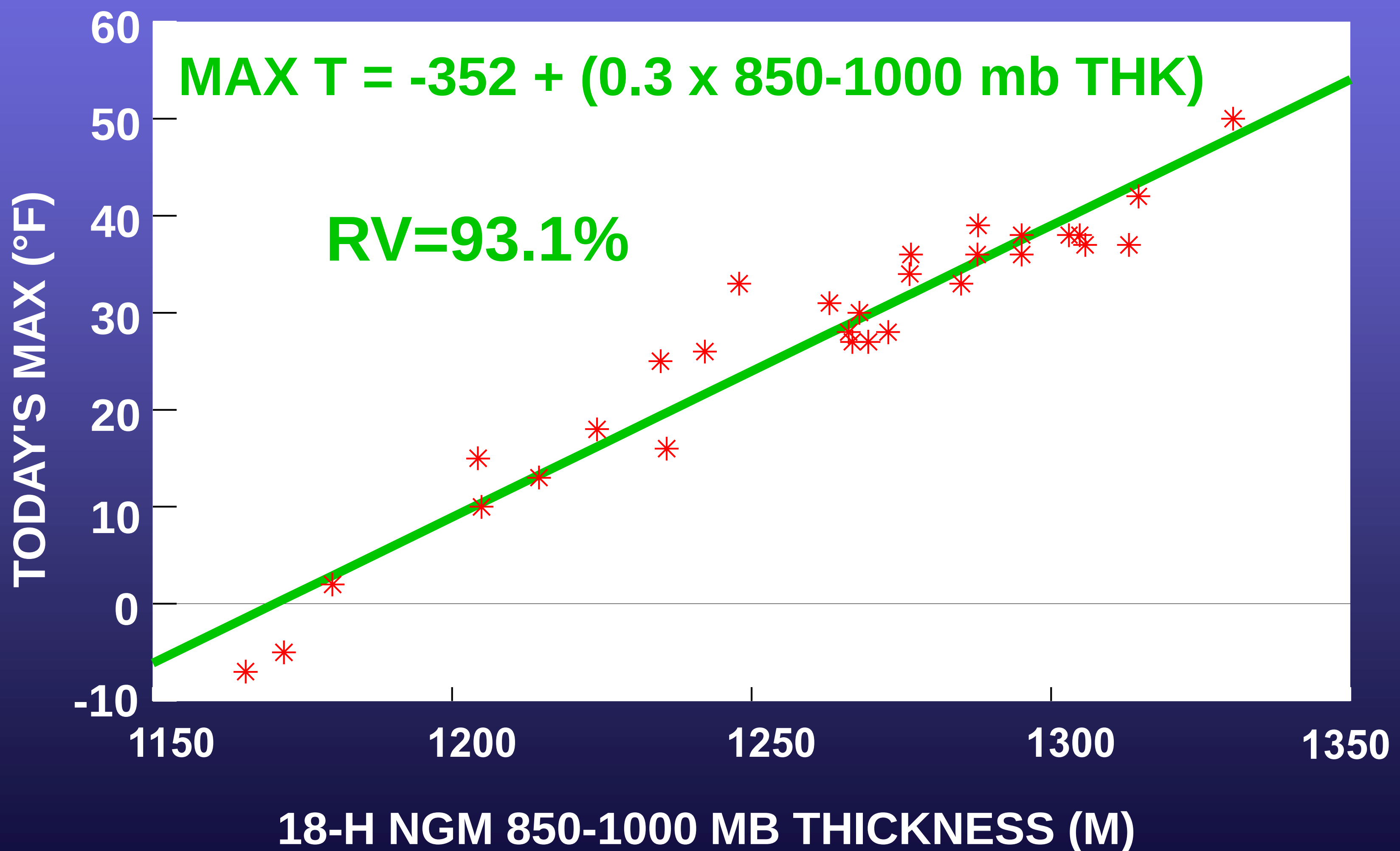
KCMH



MOS LINEAR REGRESSION

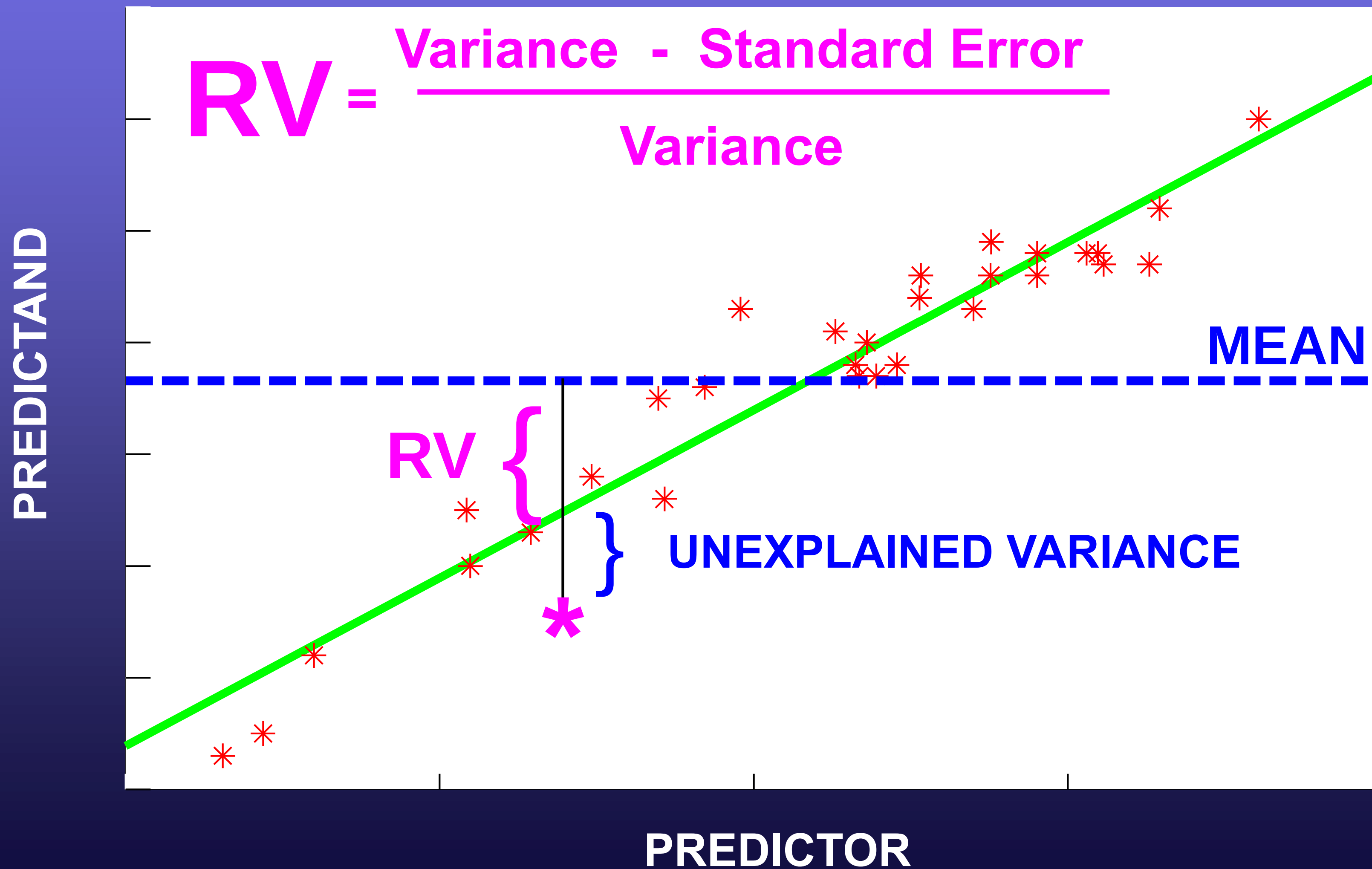
JANUARY 1 - JANUARY 30, 1994 0000 UTC

KCMH



REDUCTION OF VARIANCE

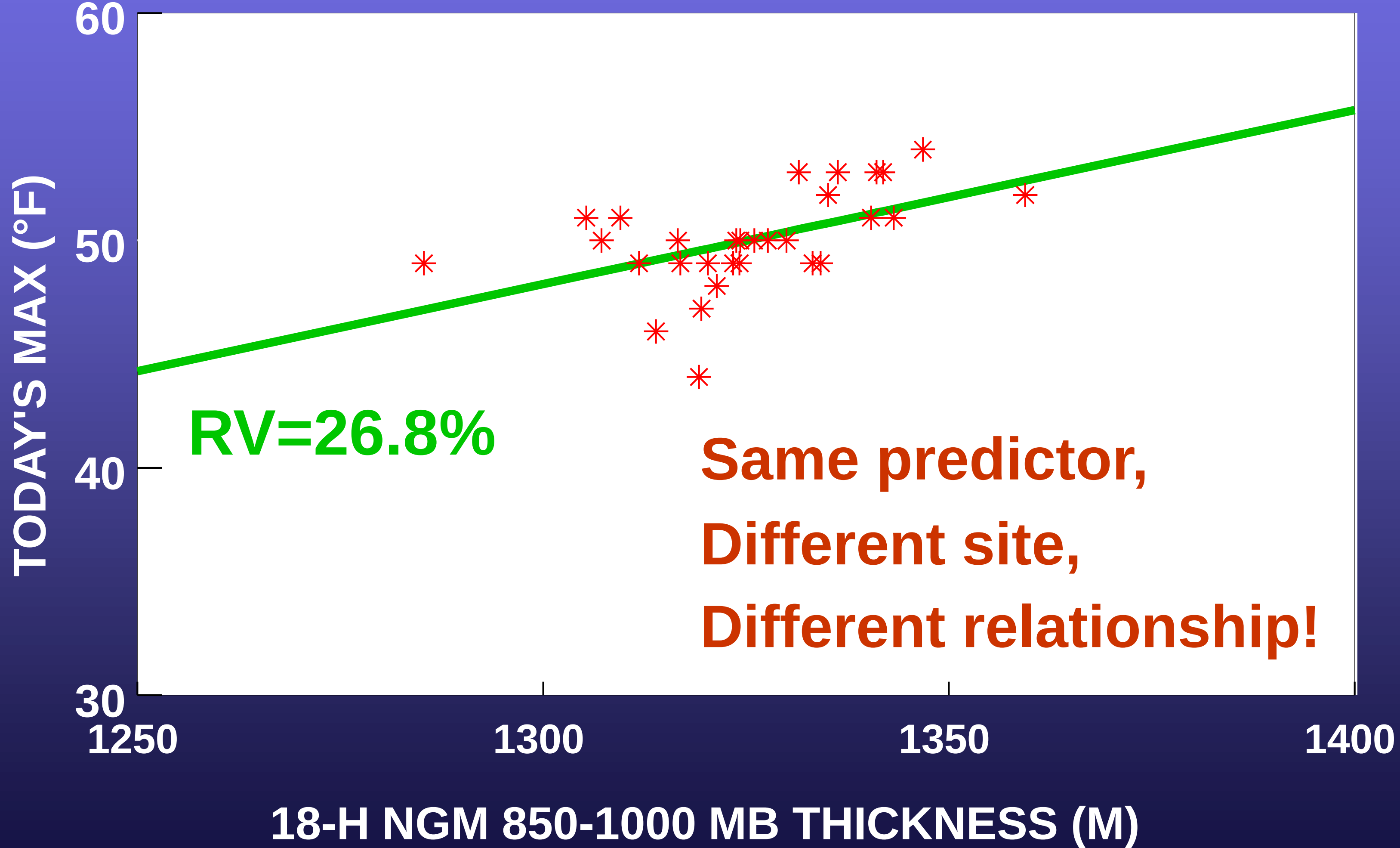
A measure of the “goodness” of fit and
Predictor / Predictand correlation



MOS LINEAR REGRESSION

JANUARY 1 - JANUARY 30, 1994 0000 UTC

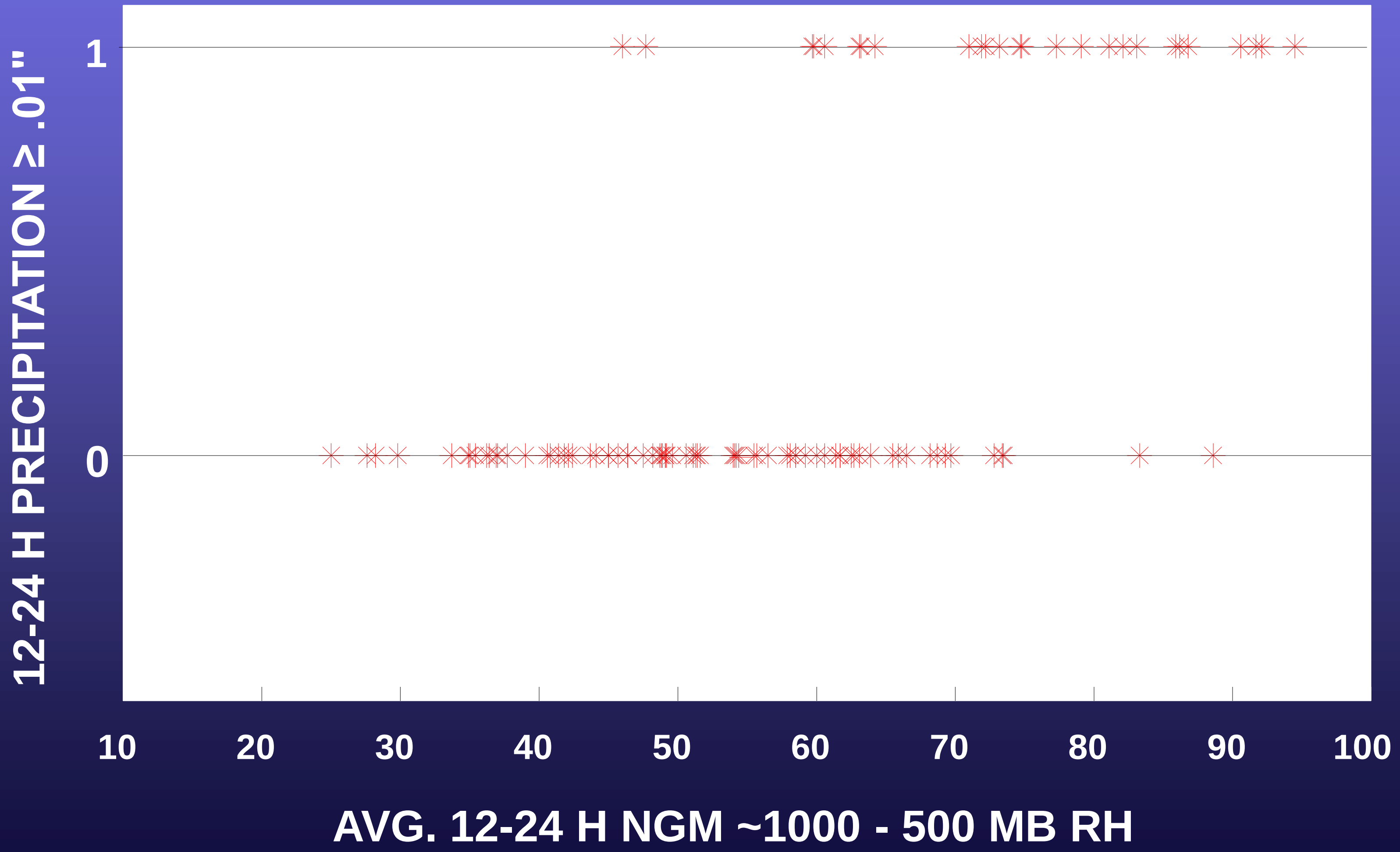
KUIL



MOS LINEAR REGRESSION

DECEMBER 1 1993 - MARCH 5 1994   0000 UTC

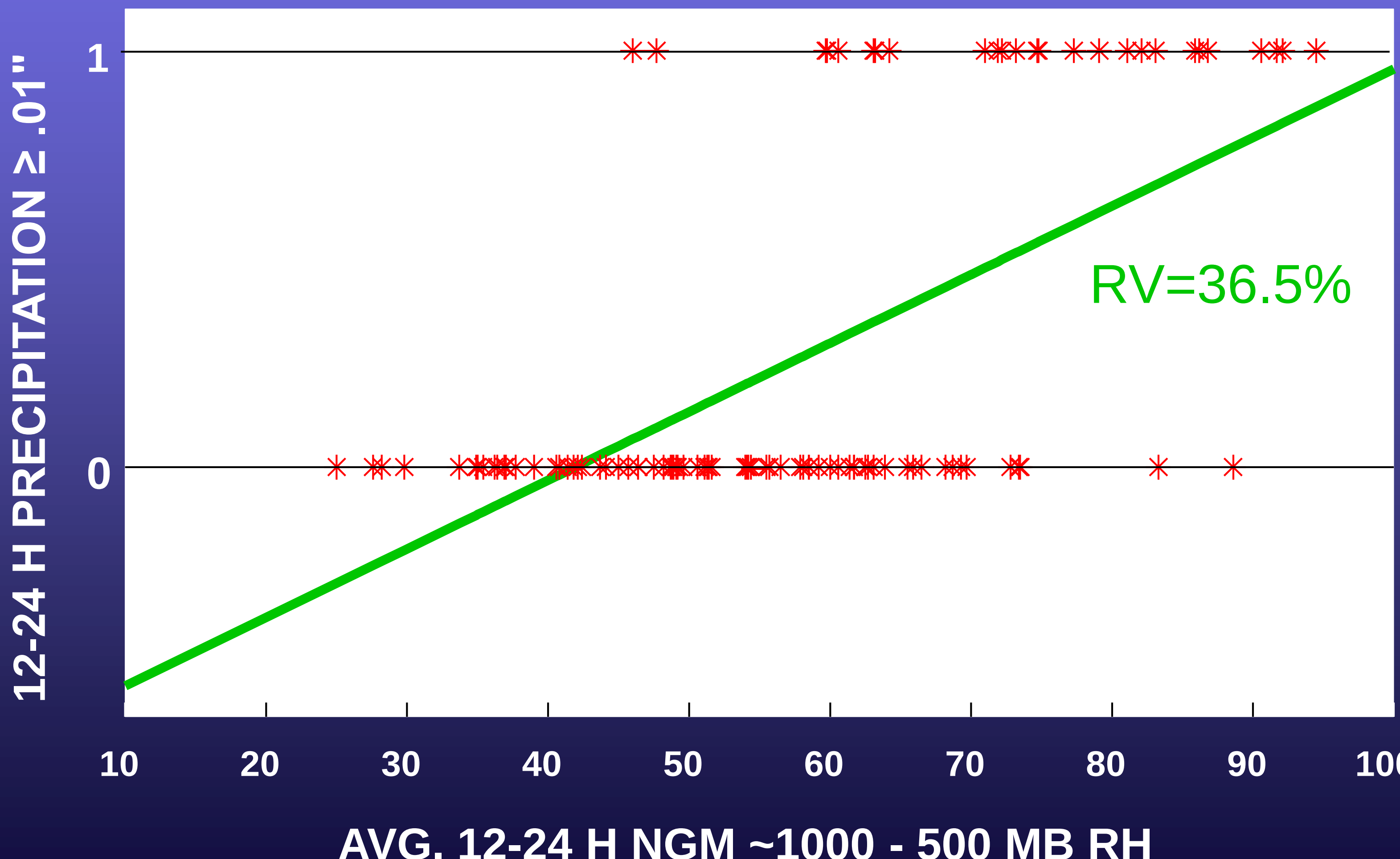
KCMH



MOS LINEAR REGRESSION

DECEMBER 1 1993 - MARCH 5 1994 0000 UTC

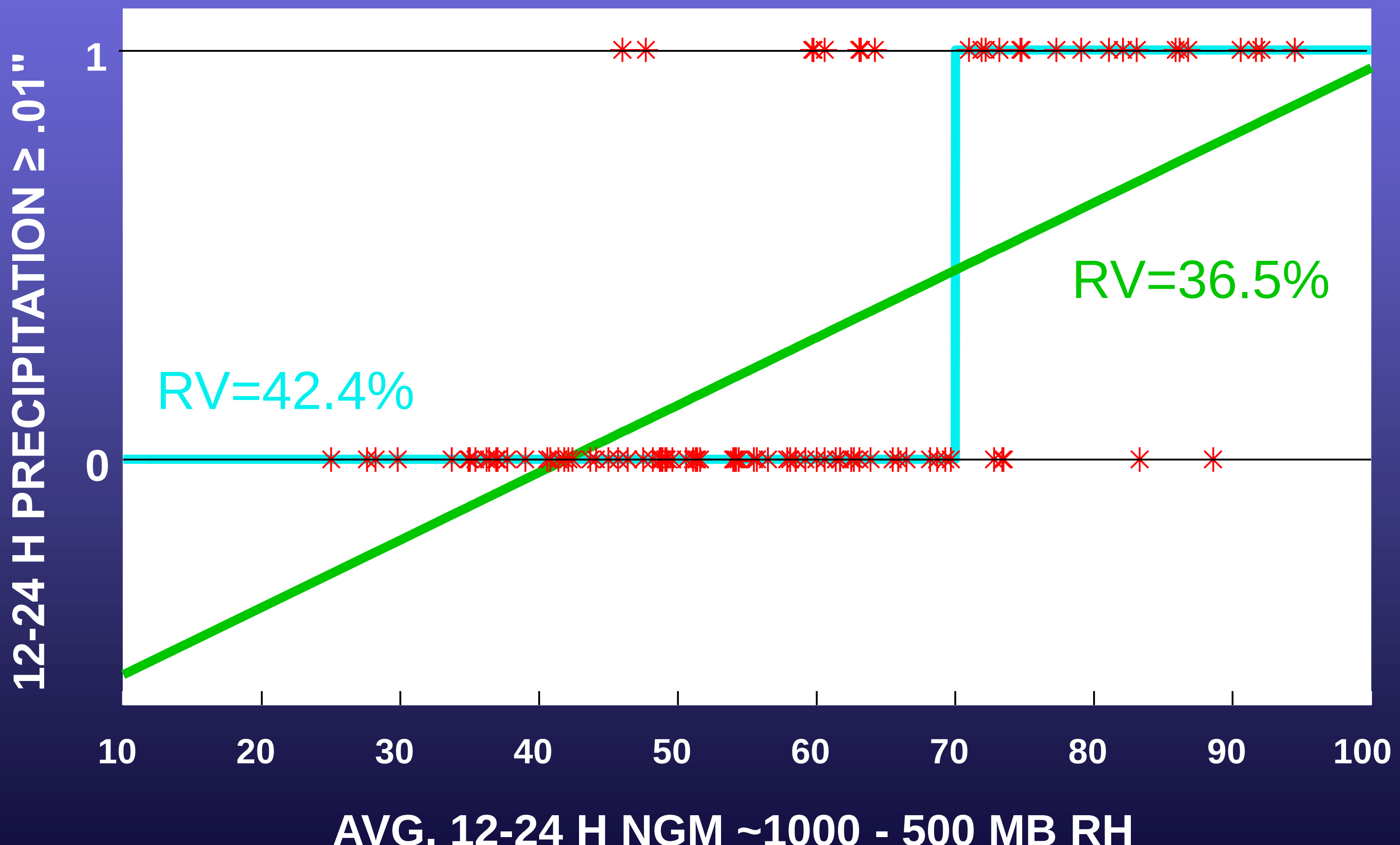
KCMH



MOS LINEAR REGRESSION

DECEMBER 1 1993 - MARCH 5 1994 0000 UTC

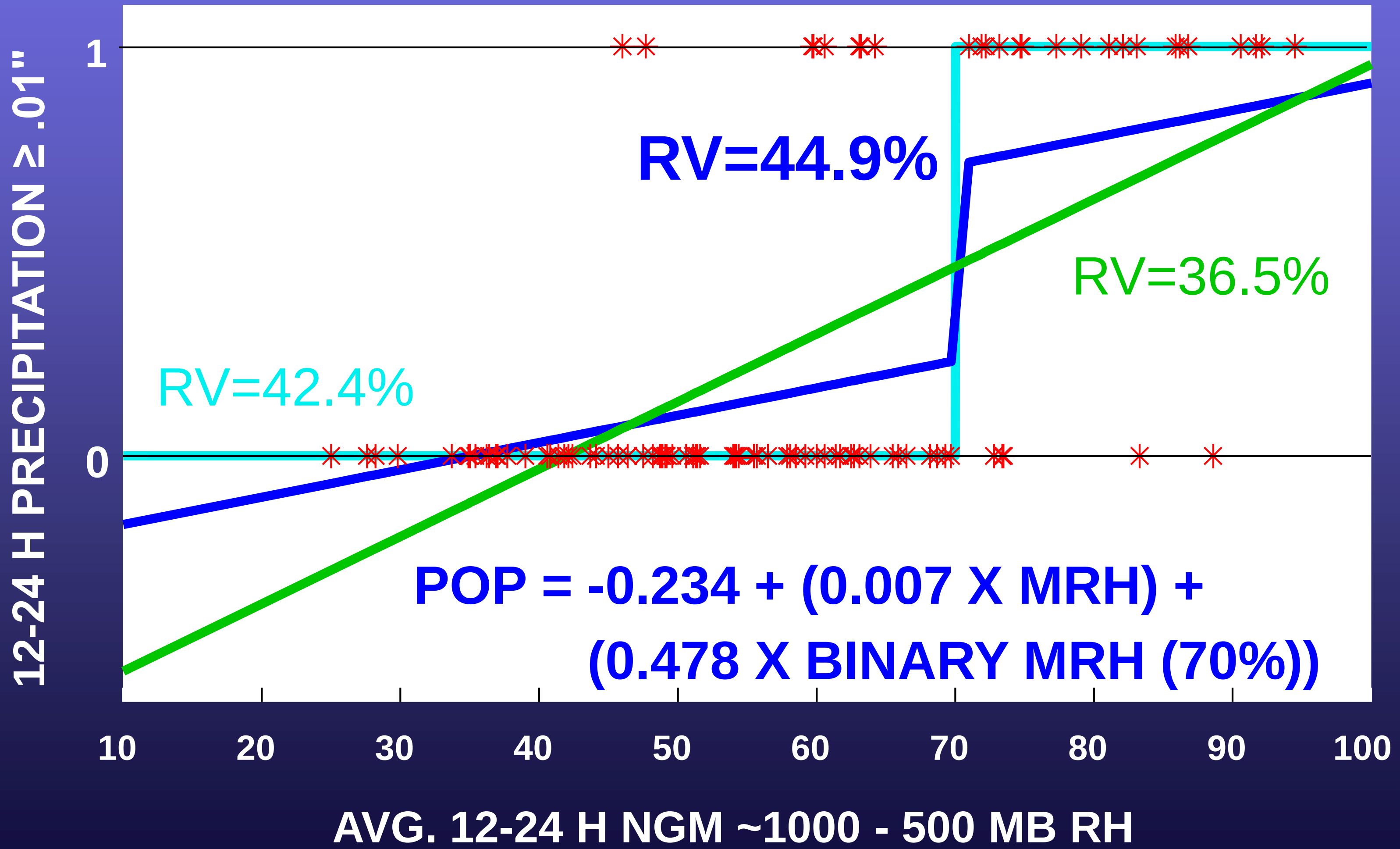
KCMH



MOS LINEAR REGRESSION

DECEMBER 1 1993 - MARCH 5 1994 0000 UTC

KCMH



EXAMPLE REGRESSION EQUATIONS

$$Y = a + bX$$

CMH MAX TEMPERATURE EQUATION

MAX T = -352 + (0.3 x 850 -1000 mb THICKNESS)

CMH PROBABILITY OF PRECIPITATION EQUATION

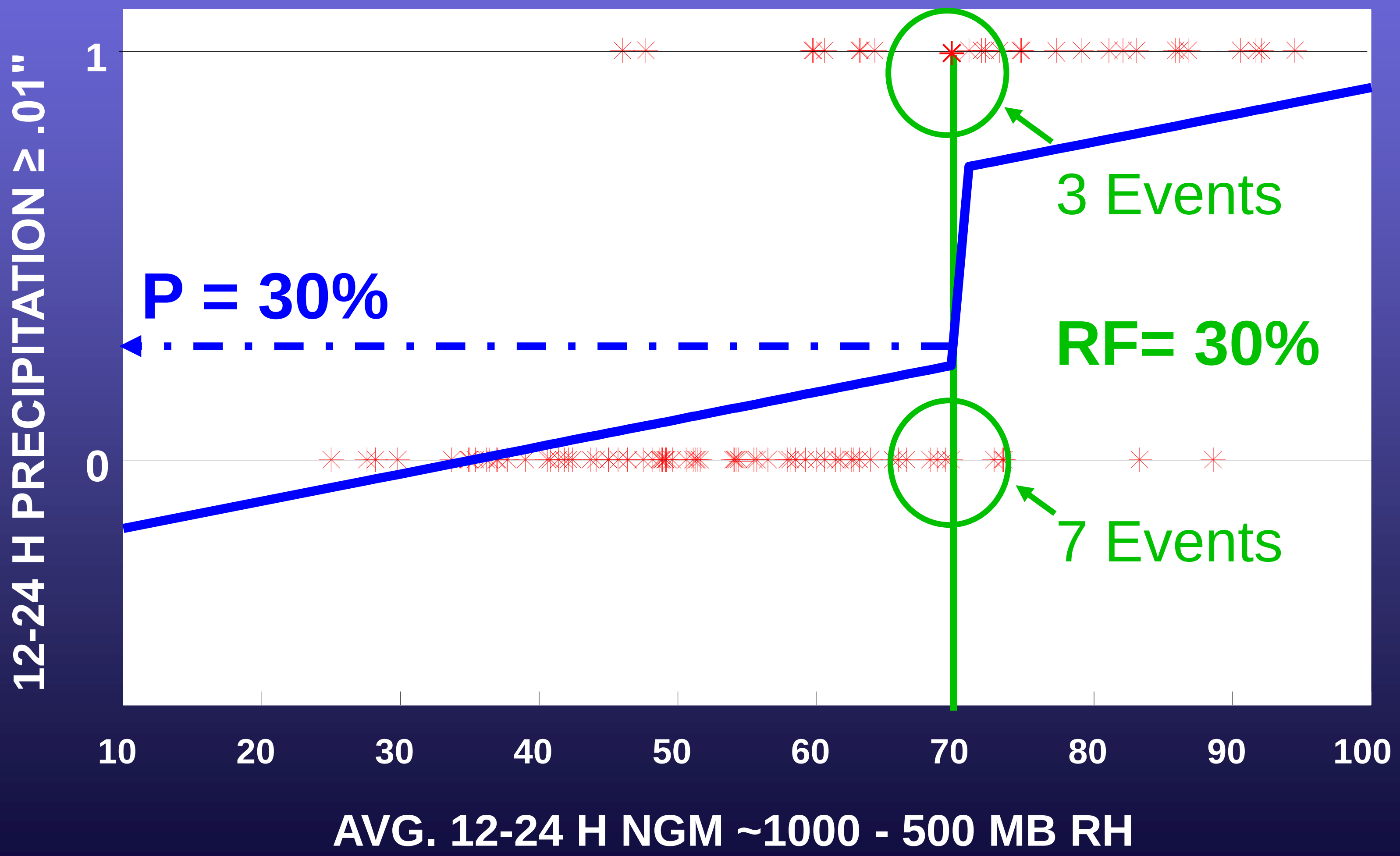
POP = -0.234 + (0.007 x MEAN RH)

+ (0.478 x BINARY MEAN RH CUTOFF AT 70%)*

*(IF MRH ≥ 70% BINARY MRH = 1; else BINARY MRH = 0)

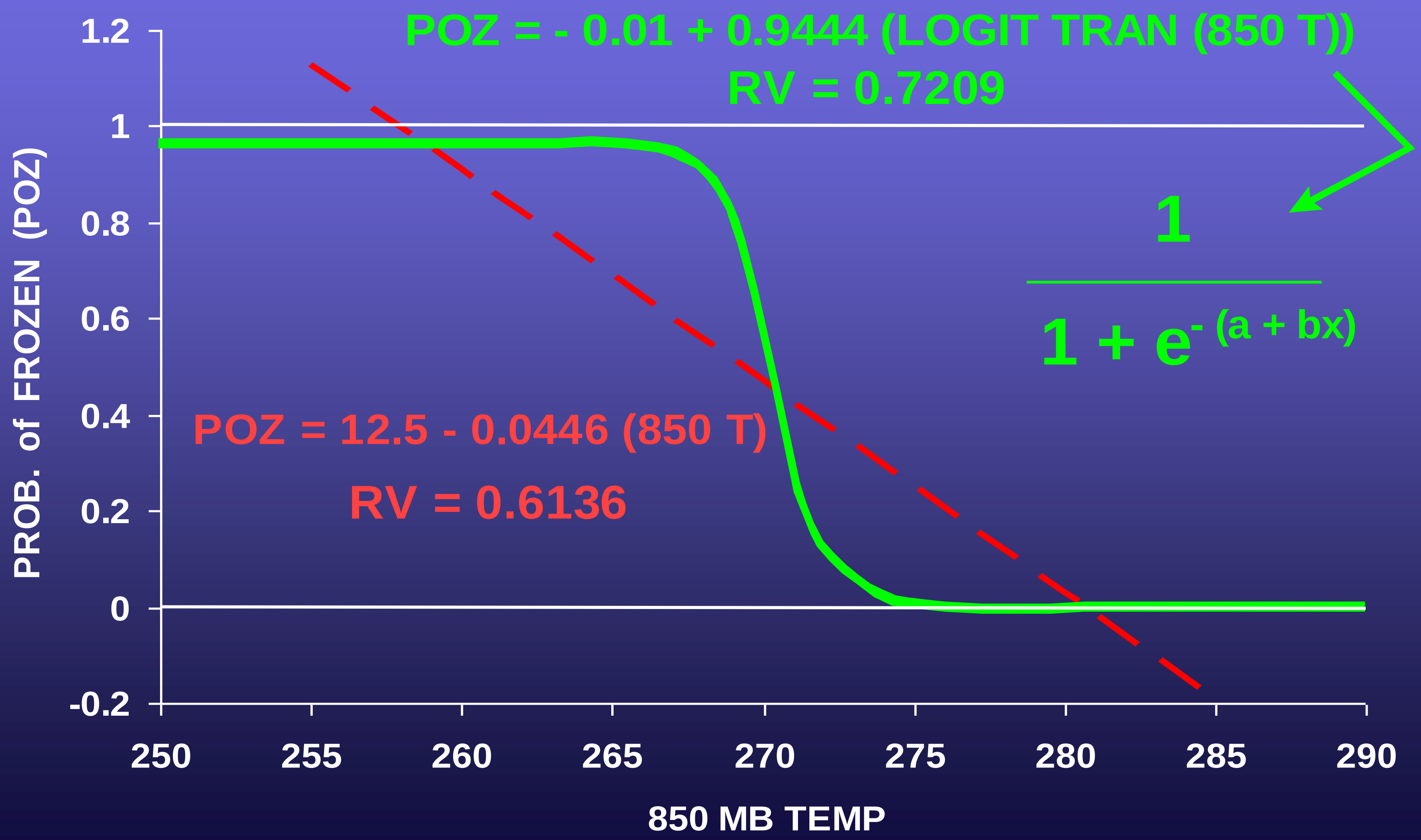
If the predictand is BINARY, MOS regression equations produce estimates of event PROBABILITIES...

KCMH



Logit Transform Example

KPIA (Peoria, IL)   0000 UTC ; 18-h projection



DEVELOPMENTAL CONSIDERATIONS

(cont.)

- **Terms in Equations; Selection Criteria**

“REAL” REGRESSION EQUATIONS

MOS regression equations are MULTIVARIATE , of form:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \ldots + a_N X_N$$

Where,

the "a's" represent COEFFICIENTS

the "X's" represent PREDICTOR variables

The maximum number of terms, N, can be ***QUITE*** large:

For GFS QPF, N = 15 For GFS VIS, N = 20

The **FORWARD SELECTION** procedure determines the predictors and the order in which they appear.

FORWARD SELECTION

- METHOD OF PREDICTOR SELECTION ACCORDING TO CORRELATION WITH PREDICTAND
- “BEST” OR STATISTICALLY MOST IMPORTANT PREDICTORS CHOSEN FIRST

- **FIRST** predictor selected accounts for greatest reduction of variance (RV)
- Subsequent predictors chosen that give greatest RV in conjunction with predictors already selected
- **STOP** selection when desired maximum number of terms is reached or new predictors provide less than a user-specified minimum RV

DEVELOPMENTAL CONSIDERATIONS

(cont.)

- Terms in Equations; Selection Criteria

- Dependent Data
  - Sample Size, Stability, Representativeness
    - AVOID OVERFIT !!
  - Stratification - Seasons
  - Pooling – Regions