STATISTICS
TEXAS A&M UNIVERSITY

Handout 15

Statistical Analysis with the GLIMMIX Procedure

Applications Using the GLIMMIX Procedure

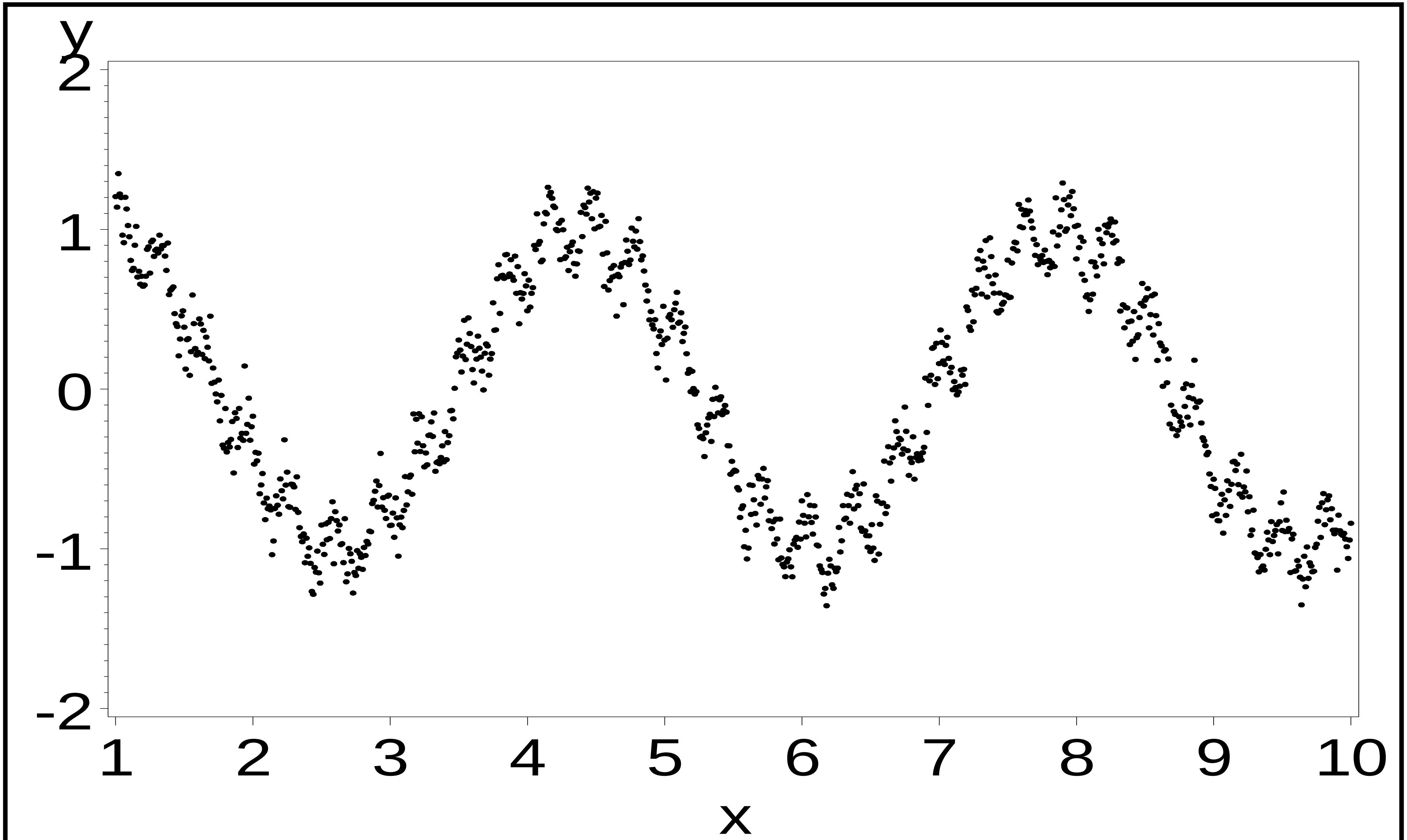
Introduction to Radial Smoothing

Introduction to Radial Smoothing

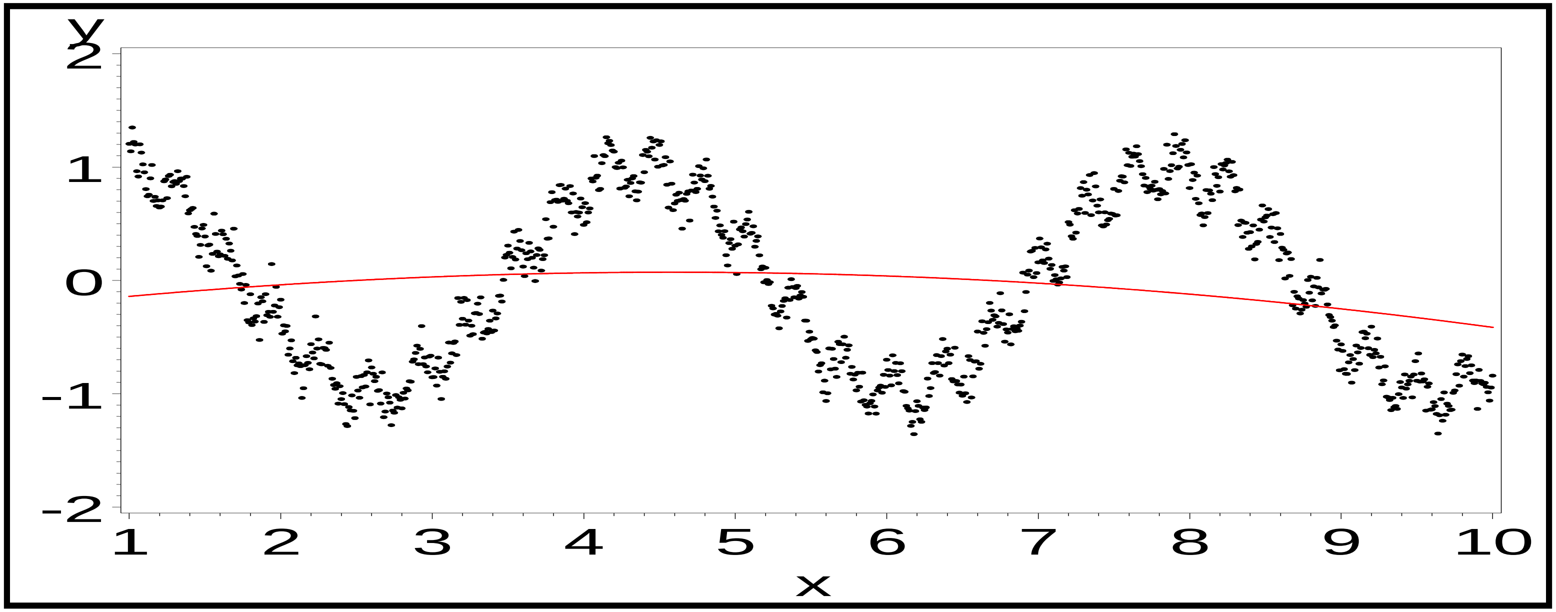
Objective

- Define low-rank radial smoothing.
- Use the GLIMMIX procedure to fit a model with radial smoothing.

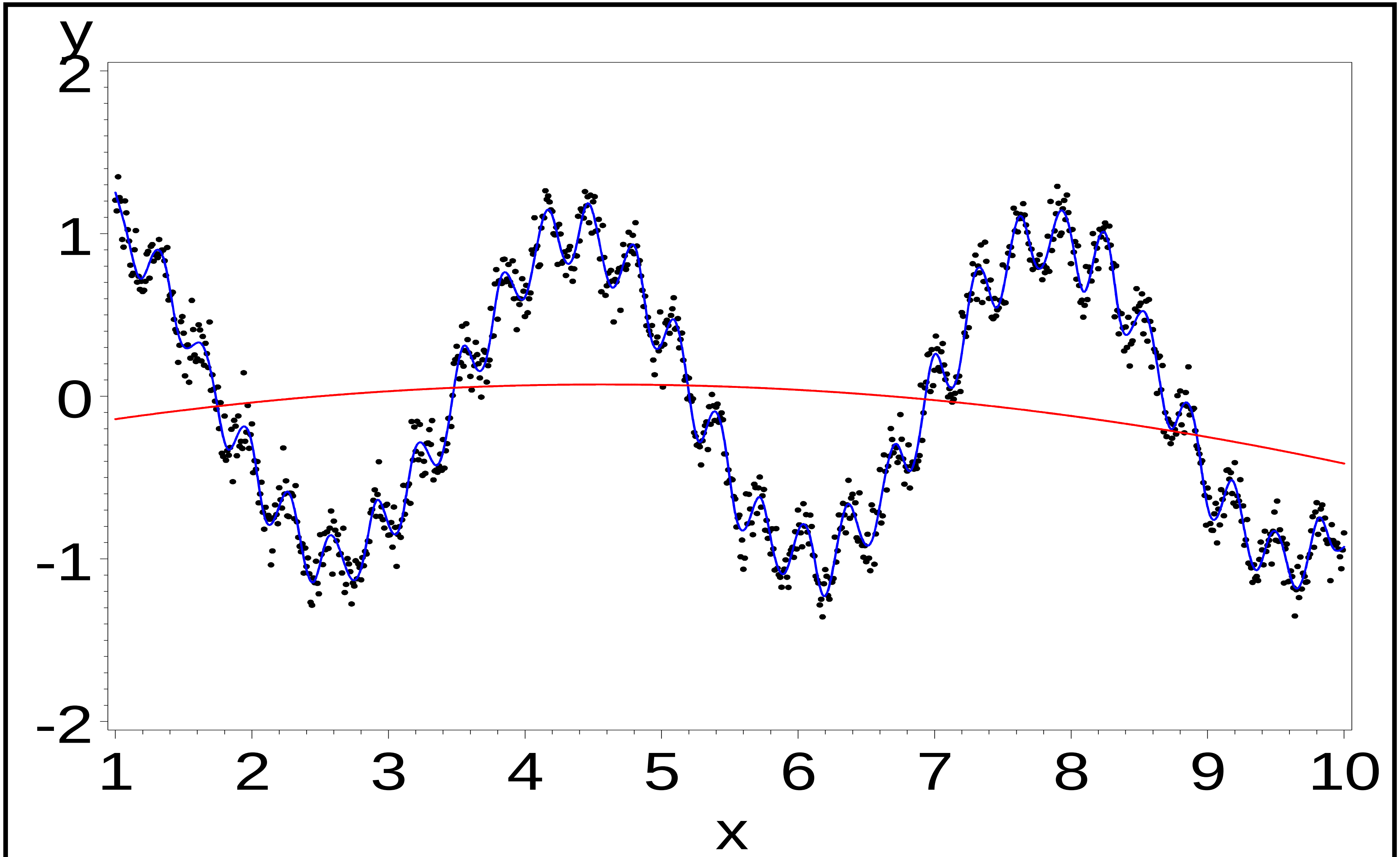
Motivating Example: The Data



Motivating Example: The Quadratic Model



Motivating Example: Radial Smoothing



Example Code

```
proc glimmix;
   model y=x x*x / solution;
   random x / type=rsmooth
              knotmethod=equal(100);
   output out=out pred(blup)   = pred1
                  pred(noblup) = pred2;
run;
```

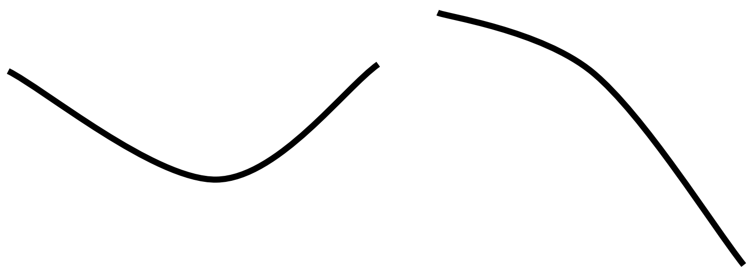
The radial smoother implemented with the TYPE=RSMOOTH option in the RANDOM statement is an approximate low-rank thin-plate spline as described in Ruppert, Wand, and Carroll (2003, Chapter 13.4–13.5).

Classical Splines

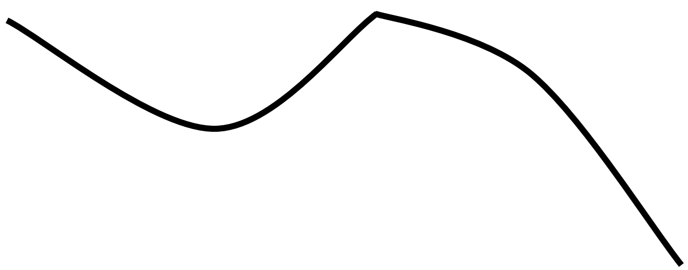
– *Splines* are piecewise polynomial segments of degree *n* joined at knots with varying continuity and smoothness constraints.

– The *knots* are the abscissas of these joint points.

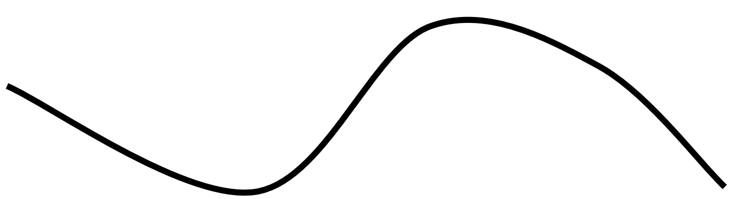
Quadratic Splines with One Knot and Varying Continuity Restrictions



Two piecewise polynomials



A continuous spline without a continuous first derivative



A continuous spline with a continuous first derivative

Model with the Linear Spline

Spline coefficients

Spline functions

Spline knots

$$y_i = \beta_0 + \beta_1 x_i + \sum_{j=1}^{K} \gamma_j f(x_i; t_j) + \varepsilon_i$$

i=1 to n

Linear spline: $f(x_i; t_j) = \begin{cases} x_i - t_j & x_i > t_j \\ 0 & otherwise \end{cases}$

Y_i : the i[th] observation of the dependent variable y.

x_i: the i[th] observation of the independent variable x.

β_0: the intercept estimate

β_1: the slope estimate

γ_j: the spline coefficients

K: the total number of knots

t_j: the knots.

Spline and Mixed Model

Spline coefficients

Spline functions

Spline knots

$$y_i = \beta_0 + \beta_1 x_i + \sum_{j=1}^{K} \gamma_j f(x_i; t_j) + \varepsilon_i$$

$$Y = X\beta + Z\gamma + e$$

Mixed Model Smoothing

- In classical spline models, the spline coefficients are fixed effects.
- In mixed-model smoothing (PROC GLIMMIX), the solutions for the spline coefficients are solutions for the random effects.
- Random effects must be continuous variables.

Advantages of Random Spline Coefficient

- The REML/ML estimate of σ_r^2 can be easily obtained.
- Only one parameter σ_r^2 needs to be estimated. The spline coefficients are EBLUPs.
- Inferences (standard errors, prediction intervals, and so on) on the fixed effects reflect that the spline coefficients are EBLUPs.
- The significance of the smooth component can be tested (likelihood ratio test for σ_r^2).
- The smoothing component selection is *demystified*.

Radial Smoothing

– In classical spline models,
  – spline functions are typically low-degree polynomial functions
  – the fit for any given point is obtained by interpolation.

– In mixed model smoothing (PROC GLIMMIX), spline functions are computed using a low-rank radial smoother and the results are contained in the **Z** matrix. The **Z** matrix
  – is a $n \times k$ matrix, where n is the number of observations and k is the number of knots
  – incorporates information about spacing between the data and knots, and spacing between the knots
  – is built only once prior to model fitting, and does not depend on parameters.

Question

Which of the following is true about (penalized) spline and mixed model radial smoothing?

a. They both use nonparametric smoothing techniques.
b. In classical splines, the spline coefficients are fixed. In mixed-model radial smoothing, the spline coefficients are random.
c. In classical splines, the spline function is a low-degree polynomial. In mixed-model smoothing, the spline function is a radial smoother.
d. In mixed-model radial smoothing, the results of the radial smoother are contained in the Z matrix.
e. All of the above
f. None of the above

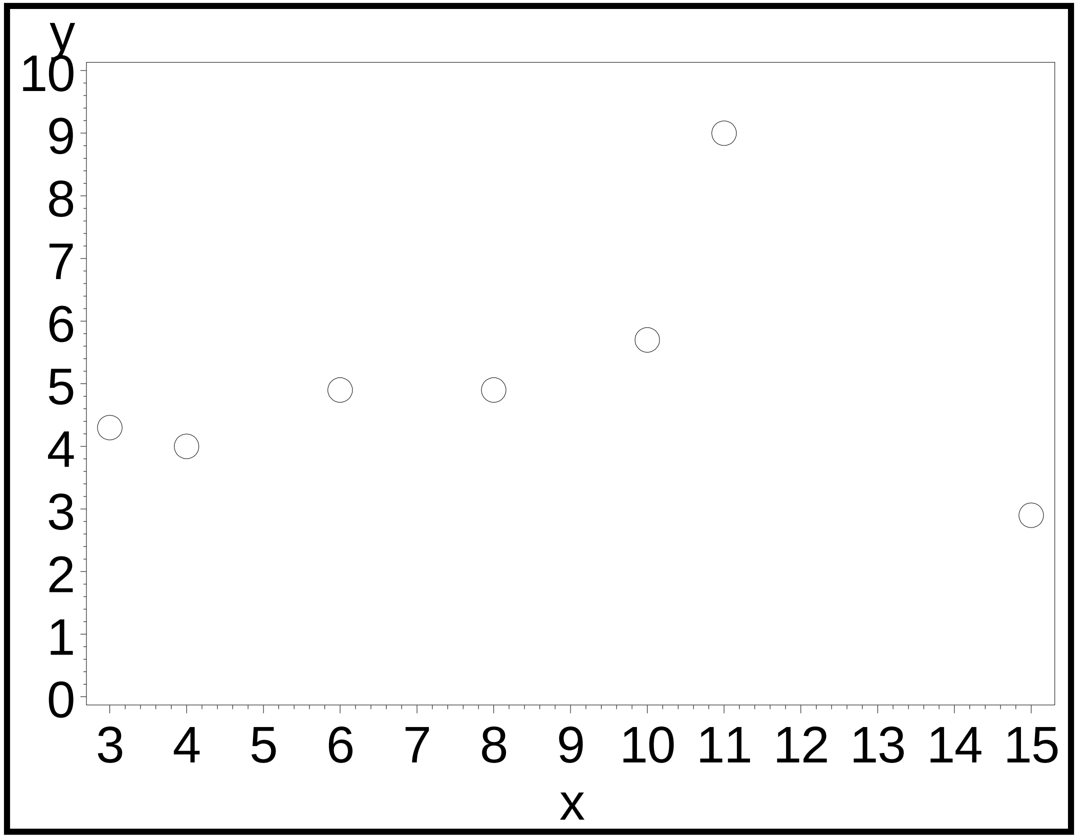
OneRadialsmoothing Example

data **one**: data **knots**:

x	y
3	4.3
4	4.0
6	4.9
8	4.9
10	5.7
11	9.0
15	2.9

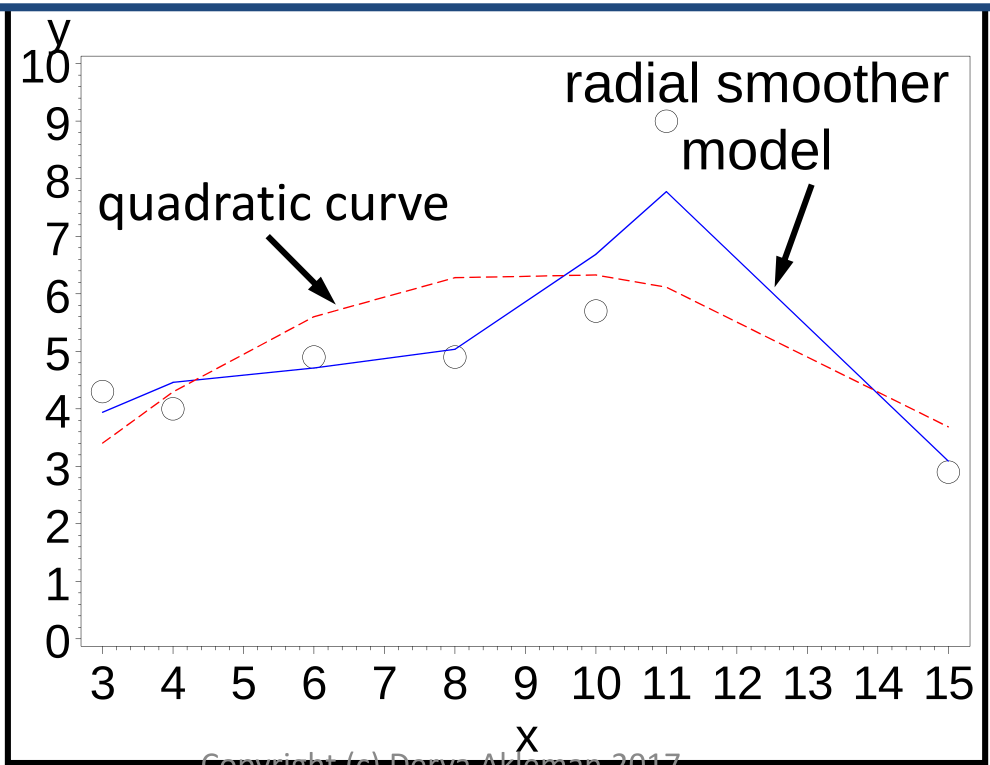
x
3
9
15

n=7, k=3



Radial Smoother Model

```
proc glimmix data=one;
   model y=x x*x / solution;
   random x / solution type=rsmooth
                   knotmethod=data(knots);
   output out=out pred=pred;
run;
```



Solutions for Fixed and Random Effects

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	-26.7684	40.2162	2	-0.67	0.5741
x	2.9749	7.6373	2.416	0.39	0.7286
x*x	0.03587	0.4239	2.416	0.08	0.9391

Solution for Random Effects

Effect	Estimate	Std Err Pred	DF	t Value	Pr > |t|
RSmooth(x)	0.4773	0.5091	2	0.94	0.4474
RSmooth(x)	0.3880	0.3396	2	1.14	0.3715
RSmooth(x)	-0.5882	0.5226	2	-1.13	0.3773

What Is Going On?

```
proc glimmix data=one outdesign=two;
   model y=x x*x / solution;
   random x / solution type=rsmooth
                  knotmethod=data(knots);
run;
```

is equivalent to

```
proc glimmix data=two;
   model y=x x*x / solution;
   random _z1 _z2 _z3 / type=toep(1) solution;
run;
```

OUTDESIGN= Data Set

x	y	_X1	_X2	_X3	_Z1	_Z2	_Z3
3	4.3	1	3	9	41.3823	4.1195	-0.1869
4	4.0	1	4	16	31.0833	8.7159	-0.9115
6	4.9	1	6	36	16.1679	10.8138	-0.7196
8	4.9	1	8	64	7.1040	8.7922	1.8598
10	5.7	1	10	100	1.8598	8.7922	7.1040
11	9.0	1	11	121	0.2567	9.9174	11.0339
15	2.9	1	15	225	-0.1869	4.1195	41.3823

The Predicted Value

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}$$

$$= \begin{bmatrix} 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 6 & 36 \\ 1 & 8 & 64 \\ 1 & 10 & 100 \\ 1 & 11 & 121 \\ 1 & 15 & 225 \end{bmatrix} \begin{bmatrix} -26.7684 \\ 2.9749 \\ 0.03587 \end{bmatrix} + \begin{bmatrix} 41.38 & 4.12 & -0.19 \\ 31.08 & 8.72 & -0.91 \\ 16.17 & 10.81 & -0.72 \\ 7.10 & 8.79 & 1.86 \\ 1.86 & 8.79 & 7.10 \\ 0.25 & 9.92 & 11.03 \\ -0.19 & 4.12 & 41.38 \end{bmatrix} \begin{bmatrix} 0.4773 \\ 0.3880 \\ -0.5882 \end{bmatrix}$$

How are the Zs Computed?

If n_r denotes the number of continuous random effects in the RANDOM statement, then the covariance structure of the random effects γ is determined as follows:

Suppose that $\mathbf{x}_i$ denotes the vector of random effects for the i^{th} observation. Let $\boldsymbol{\tau}_k$ denote the $(n_r \times 1)$ vector of knot coordinates. $k = 1, \ldots, K$, and K are the total number of knots. The Euclidean distance between the k^{th} knot and the t^{th} knot is computed as follows:

$$d_{kt} = \| \boldsymbol{\tau}_k - \boldsymbol{\tau}_t \| = \sqrt{\sum_{j=1}^{n_r} (\tau_{jk} - \tau_{jt})^2}$$

The distance between knots and the random effects is computed as shown below:

$$h_{ik} = \| \mathbf{x}_i - \boldsymbol{\tau}_k \| = \sqrt{\sum_{j=1}^{n_r} (x_{ij} - \tau_{jk})^2}$$

How are the Zs Computed?

The **Z** matrix for the GLMM is constructed as follows:

$$\mathbf{Z} = \tilde{\mathbf{Z}}\Omega^{-1/2}$$

where the $(n \times K)$ matrix $\tilde{\mathbf{Z}}$ has the following typical element:

$$\left[\tilde{\mathbf{Z}}\right]_{ik} = \begin{cases} h_{ik}^{p} & n_r \ odd \\ h_{ik}^{p}\log\{h_{ik}\} & n_r \ even \end{cases}$$

and the $(K \times K)$ matrix $\mathbf{\Omega}$ has the following typical element:

$$[\Omega]_{kp} = \begin{cases} d_{kt}^{p} & n_r \ odd \\ d_{kt}^{p}\log\{d_{kt}\} & n_r \ even \end{cases}$$

The exponent in these expressions equals $p = 2m - n_r$, where the optional value m corresponds to the derivative penalized in the thin-plate spline. A larger value of m yields a smoother fit. The GLIMMIX procedure requires $p > 0$ and chooses by default $m = 2$, if $n_r < 3$ and $m = (n_r + 1)/2$ otherwise.

The components of γ are assumed to have equal variance σ_r^2. The *smoothing parameter* λ of the low-rank spline is related to the variance components in the model $\lambda^2 = f(\phi, \sigma_r^2)$. See Ruppert, Wand, and Carroll (2003) for details. If the conditional distribution does not provide a scale parameter ϕ, you can add a single R-side residual parameter.

How Are the Zs Computed?

$$\mathbf{x} = \begin{bmatrix} 3 & 4 & 6 & 8 & 10 & 11 & 15 \end{bmatrix}$$

$$\tau = \begin{bmatrix} 3 & 9 & 15 \end{bmatrix}$$

$$d_{kt} = \sqrt{(\tau_k - \tau_t)^2} = |\tau_k - \tau_t| \Rightarrow \mathbf{d} = \begin{bmatrix} 0 & 6 & 12 \\ 6 & 0 & 6 \\ 12 & 6 & 0 \end{bmatrix}$$

$$h_{ik} = \sqrt{(x_i - \tau_k)^2} = |x_i - \tau_k| \Rightarrow \mathbf{h} = \begin{bmatrix} 0 & 6 & 12 \\ 1 & 5 & 11 \\ 3 & 3 & 9 \\ 5 & 1 & 7 \\ 7 & 1 & 5 \\ 8 & 2 & 6 \\ 12 & 6 & 0 \end{bmatrix}$$

How Are the Zs Computed?

$$\left[\tilde{\mathbf{Z}}\right]_{ik} = h_{ik}^3 \quad \Longrightarrow \quad \tilde{\mathbf{Z}} = \begin{bmatrix} 0 & 216 & 1728 \\ 1 & 125 & 1331 \\ 27 & 27 & 729 \\ 125 & 1 & 343 \\ 343 & 1 & 125 \\ 512 & 8 & 216 \\ 1728 & 216 & 0 \end{bmatrix}$$

$$\left[\Omega\right]_{kp} = d_{kp}^3 \quad \Longrightarrow \quad \boldsymbol{\Omega} = \begin{bmatrix} 0 & 216 & 1728 \\ 216 & 0 & 216 \\ 1728 & 216 & 0 \end{bmatrix}$$

How Are the Zs Computed?

$$\mathbf{Z} = \tilde{\mathbf{Z}}\boldsymbol{\Omega}^{-\frac{1}{2}} \Longrightarrow \mathbf{Z} = \begin{bmatrix} 41.38 & 4.12 & -0.19 \\ 31.08 & 8.72 & -0.91 \\ 16.17 & 10.81 & -0.72 \\ 7.10 & 8.79 & 1.86 \\ 1.86 & 8.79 & 7.10 \\ 0.25 & 9.92 & 11.03 \\ -0.19 & 4.12 & 41.38 \end{bmatrix}_{7\times 3}$$

Summary

- When you specify TYPE=RSMOOTH in the RANDOM statement in PROC GLIMMIX, the **Z** matrix is computed based on the low-rank radial smoothing algorithm. That is, the number of columns in the **Z** matrix is the same as the number of knots, and the elements in the **Z** matrix are computed based on the spacing between the data and knots and spacing between the knots.

- The spline coefficients (γ) are the random effects and the EBLUPs can be displayed using the SOLUTION option in the RANDOM statement in PROC GLIMMIX.

Low-Rank Radial Smoother

– Low-rank means the number of knots is (considerably) smaller than the number of observations.

– Radial smoothers play a similar role as spline functions in classical splines, with the exception that they are computed using the information about spacing between the data and knots, and spacing between the knots.

What About the G Matrix?

```
random x / type=rsmooth;
```

- The **G** matrix ($k\times k$) has the variance component structure, whose parameter estimates can be quick to obtain.

$$\mathbf{G} = \begin{bmatrix} \sigma_r^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_r^2 \end{bmatrix}_{k\times k} = \sigma_r^2 \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}_{k\times k}$$

- σ_r^2 measures the variations among the spline coefficients and the magnitude can depend on the number of knots.
- The degree of smoothness is related to the variance component σ_r^2. As this value increases, the fitted curve becomes less smooth.
- The magnitude of the variance component σ_r^2 depends upon the metric of the random effects. Rescaling of the random effects can help the optimization.

Summary of Radial Smoother

-The elements in the **Z** matrix

- can be reproduced using SAS/IML
- were computed based on spacing between the knots and the spacing between the data (x) and the knots and some matrix operations.

-The BLUP is computed through $\mathbf{Z}\gamma$, where γ is

- the spline coefficients
- considered random effects
- affected by the **Z** matrix.

-The variance component σ_r^2

- measures the variations among the spline coefficients
- can change its magnitude as the number of knots changes.

Question

The low-rank radial smoother is flexible, but resource intensive for large data sets because the covariance structure is very complex.

- True
- False

Applications of Radial Smoothing

– Periodic data

– Longitudinal data

– Spatial data

– Many continuous independent variables

STATISTICS
TEXAS A&M UNIVERSITY

A Longitudinal Example (aids.sas)

The human deficiency virus (HIV) causes AIDS by attacking an immune cell named the CD4+ cell, which facilitates the body's ability to fight infection. An uninfected person has approximately 1100 cells per milliliter of blood. Because CD4+ cells decrease in number from the time of infection, a person's CD4+ cell count can be used to monitor disease progression. A subset of Multicenter AIDS Cohort Study was obtained for 369 infected men to examine CD4+ cell counts over time.

Id: subject identification number

Time: time in years since seroconversion (time when HIV becomes detectable)

Age: in years relative to arbitrary origin

Cigarettes: packs of cigarettes smoked per day

Drug: recreational drug use (1: yes, 0:no)

Partners: number of partners relative to arbitrary origin

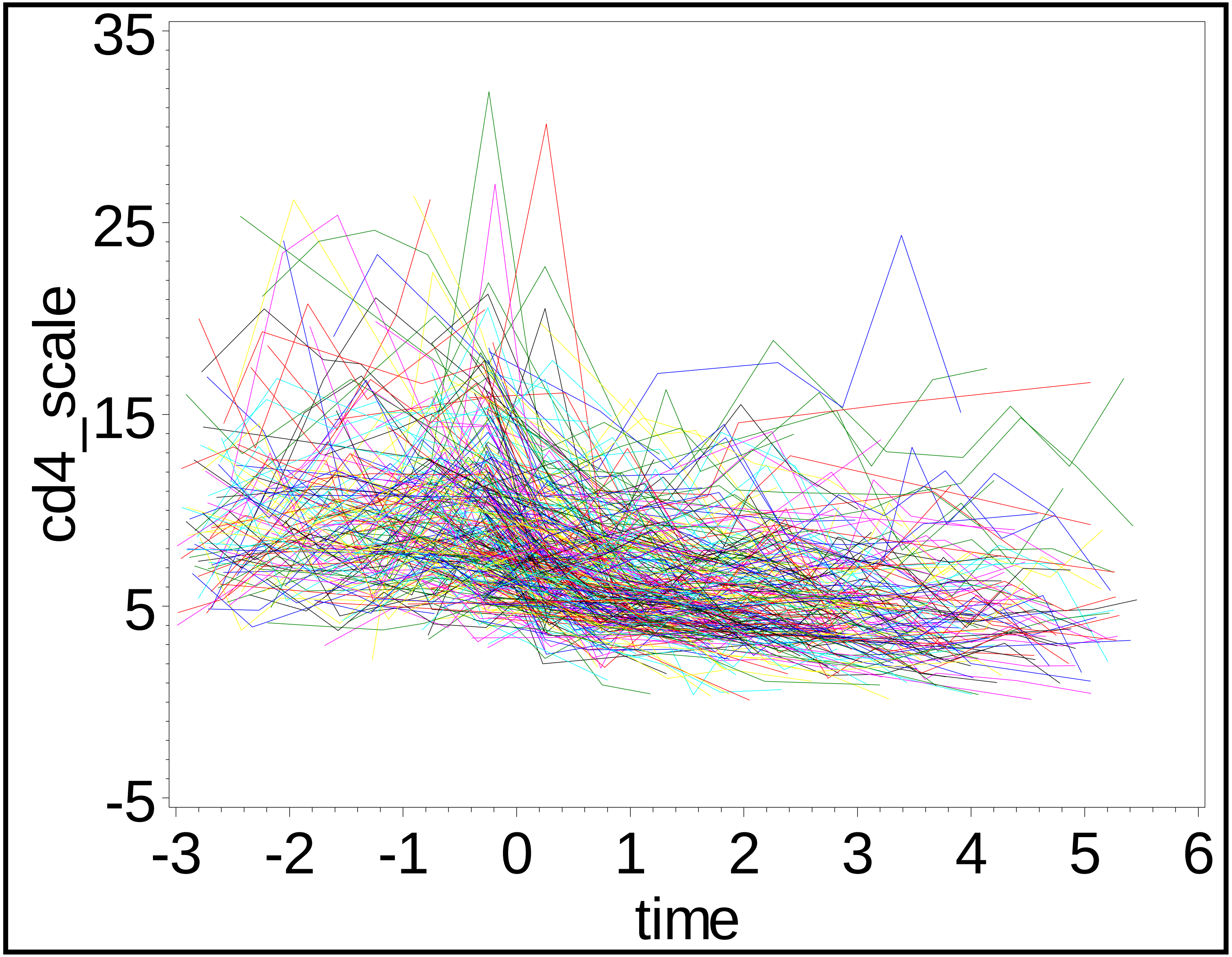
Depression: CES-D (a depression scale)

cd4_scale: scaled CD4+ cell count (the count divided by 100)

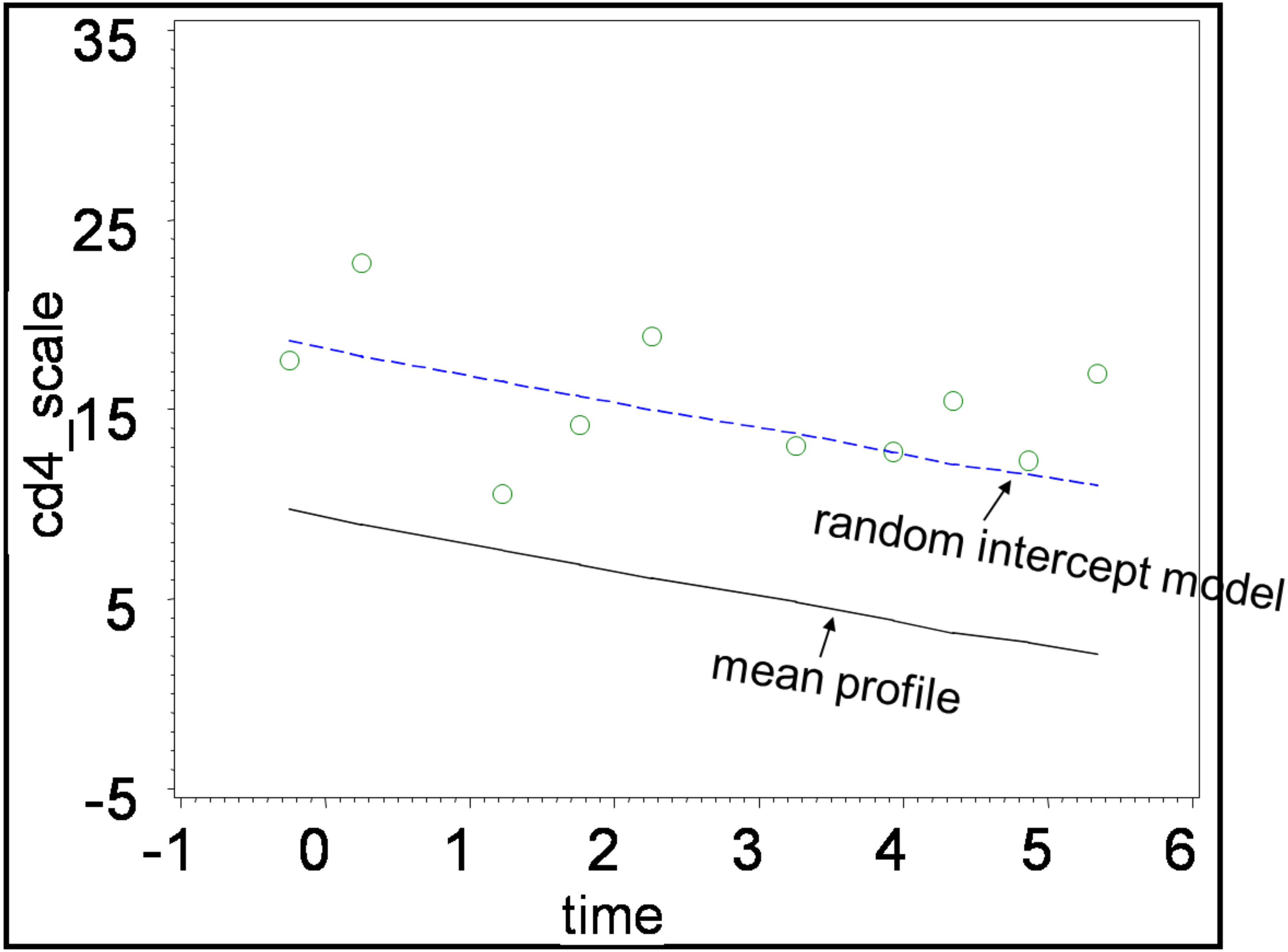
A Longitudinal Example

id	time	age	cigarettes	drug	partners	depression	cd4_scale
10002	-0.74196	6.57	0	0	5	8	5.48
10002	-0.24641	6.57	0	1	5	2	8.93
10002	0.24367	6.57	0	1	5	-1	6.57
10005	-2.72964	6.95	0	1	5	4	4.64
10005	-2.25051	6.95	0	1	5	-4	8.45
10005	-0.22177	6.95	0	1	5	-5	7.52
10005	0.22177	6.95	0	1	5	2	4.59
10005	0.77481	6.95	0	1	5	-3	1.81
10005	1.25667	6.95	0	1	5	-7	4.34
10029	-1.24025	2.64	0	1	5	18	8.46
10029	-0.74196	2.64	0	1	5	18	11.02
10029	-0.25188	2.64	0	1	5	38	8.01
10029	0.25188	2.64	0	1	5	7	8.24
10029	0.76934	2.64	0	1	5	15	8.66
10029	1.41273	2.64	0	1	5	21	7.04
10029	1.80698	2.64	0	1	5	25	7.57
...							

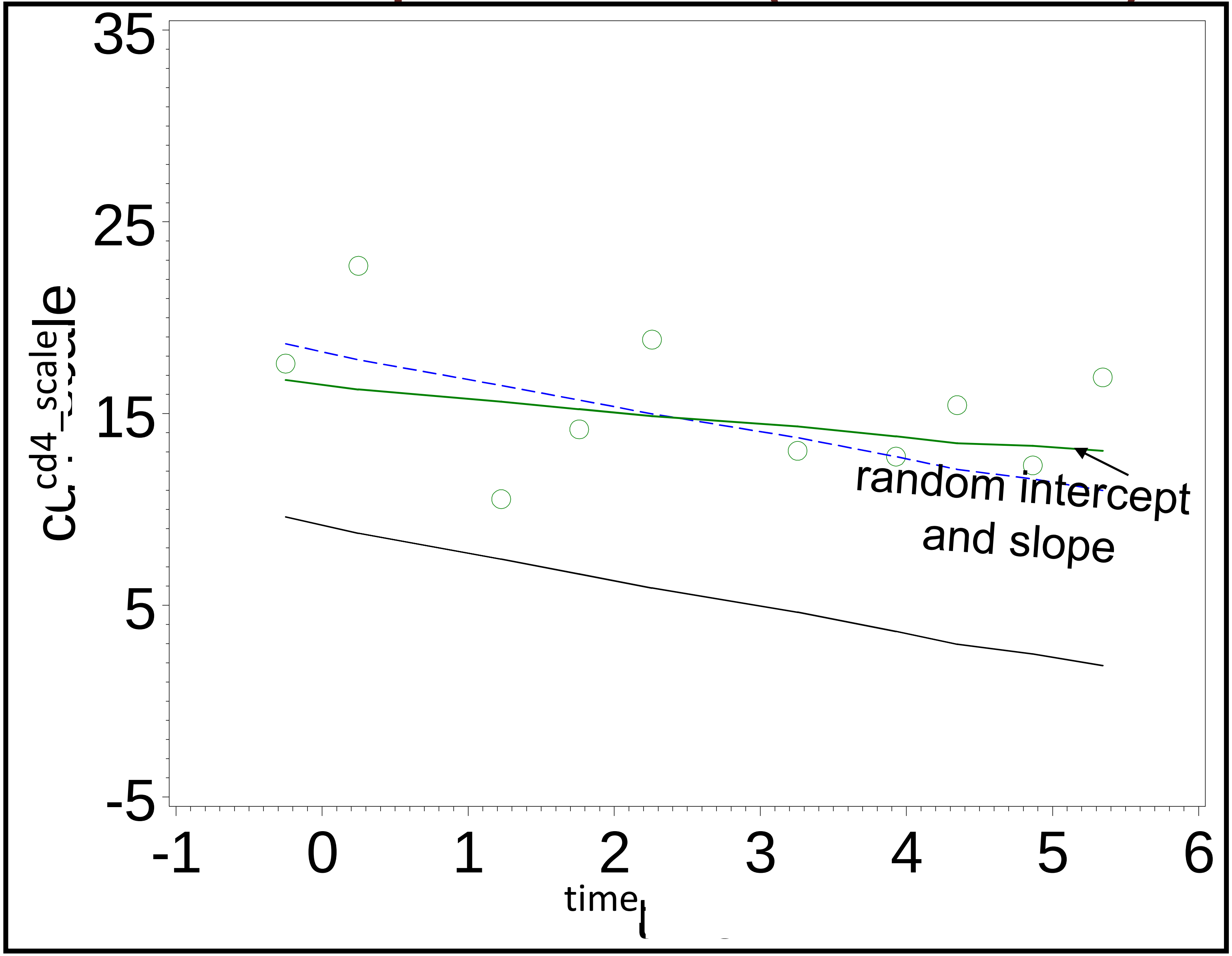
Individual Profile Plot



Random Intercept Model (id=10131)

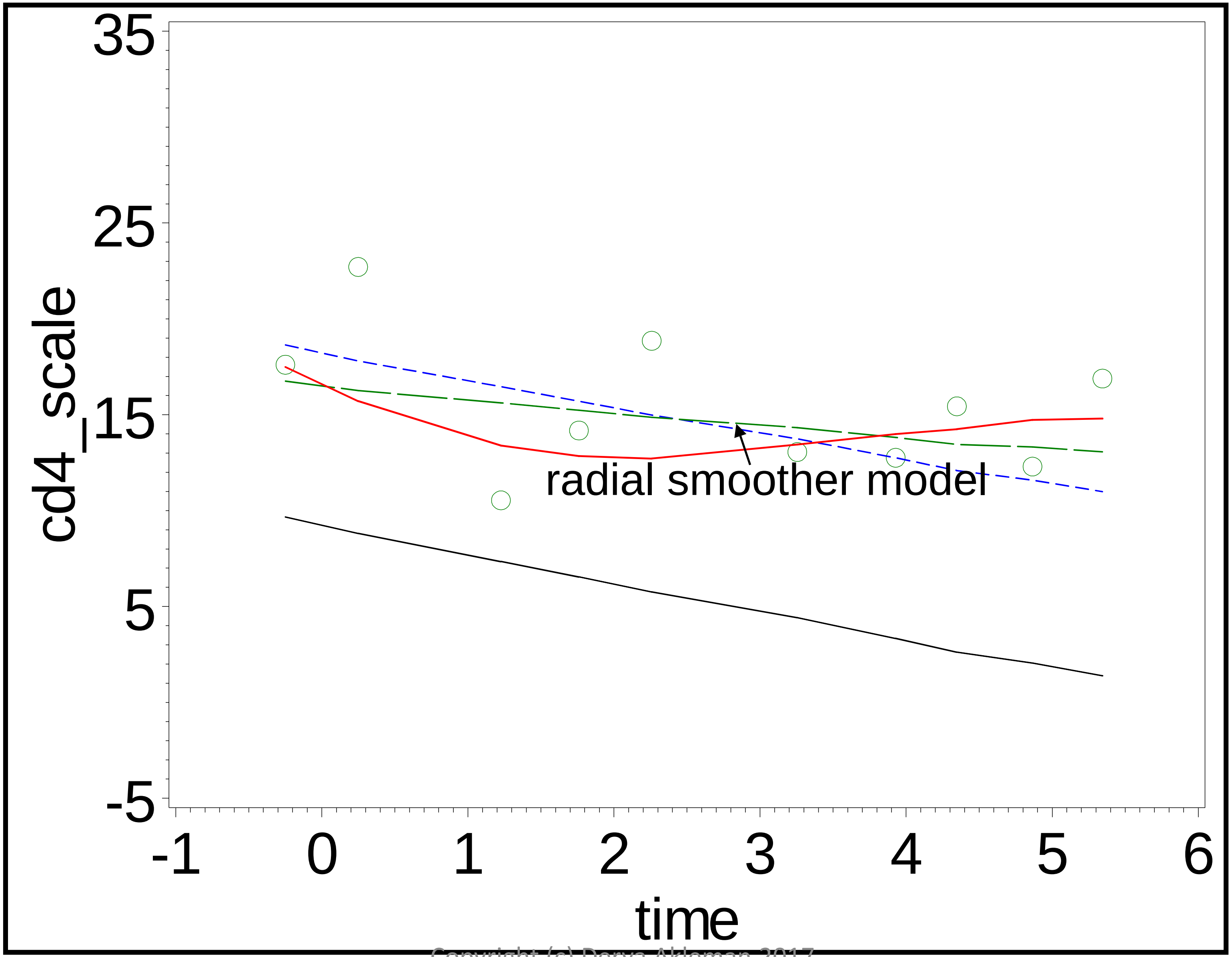


Random Intercept and Slope Model (id=10131)



STATISTICS
TEXAS A&M UNIVERSITY

Radial Smoother Model (id=10131)



Fitting a Model Using the Random coefficients and also the Radial Smoother

This demonstration illustrates concepts discussed previously.

(aidsExample.sas)

Question

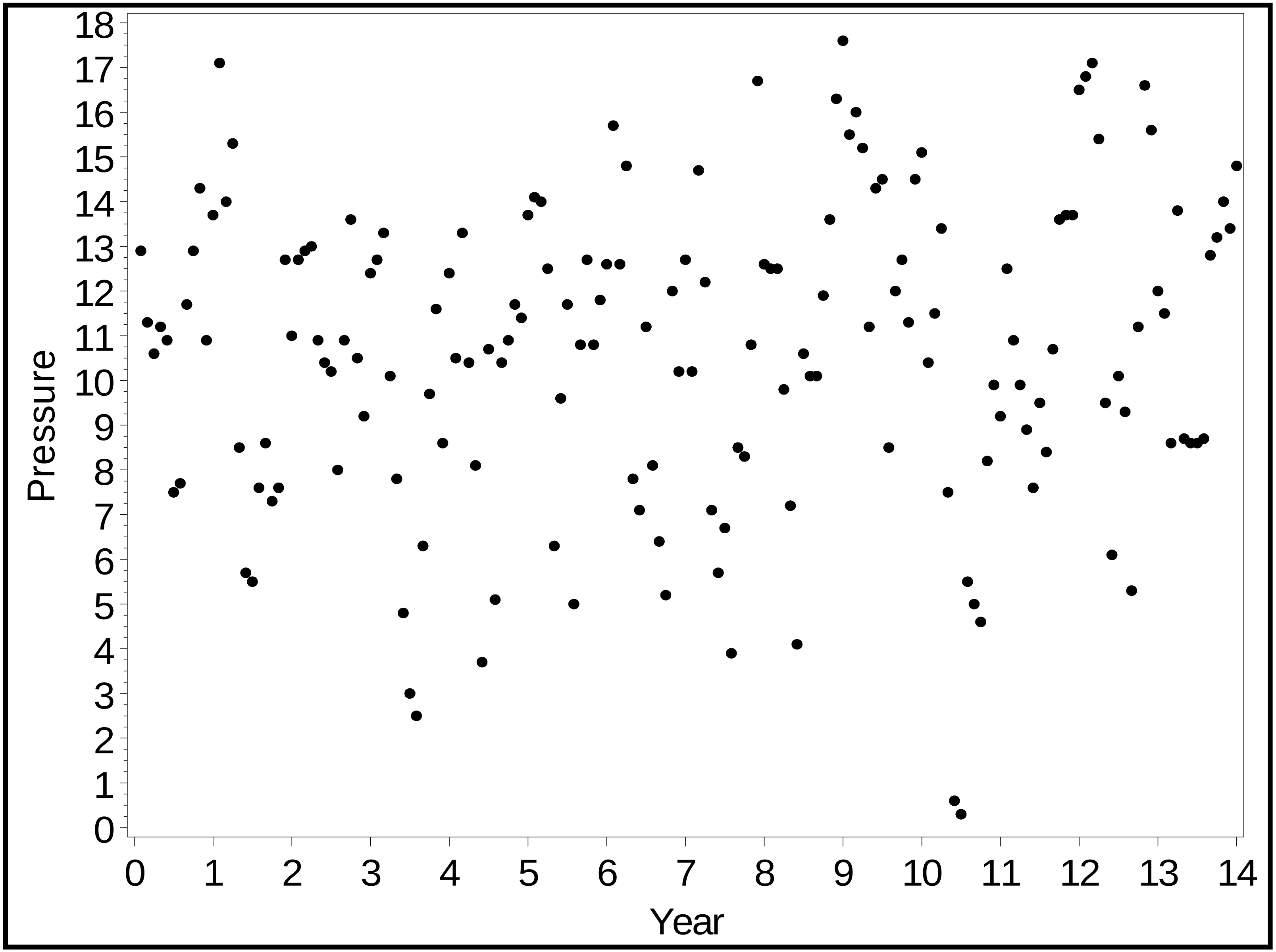
For longitudinal data, the radial smoother model is similar to random coefficient models, except that the nonparametric radial smoother rather than a parametric linear model is used to capture the trajectories.

- Yes
- No

ENSO Example

The dataset consists of monthly averaged atmospheric pressure differences between Easter Island and Darwin, Australia, for a period of 168 months. The scientists want to stucy these differences over time.

Pressure versus Year



Fitting a Model Using the Radial Smoother

This demonstration illustrates concepts discussed previously. The description of the variables in the dataset:

Pressure: monthly averaged atmospheric pressure differences between two locations

Month: month since the beginning of the study

Year: years since the beginning of the study

Reason in radial smoothing:it is not obvious how to define a parametric relationship between pressure and month.

ensoexample

Question

In a radial smoothing model, the variance component estimates the variance among the spline coefficients and should not affect the smoothness of the fitted curve.

- ❍ True
- ❍ False

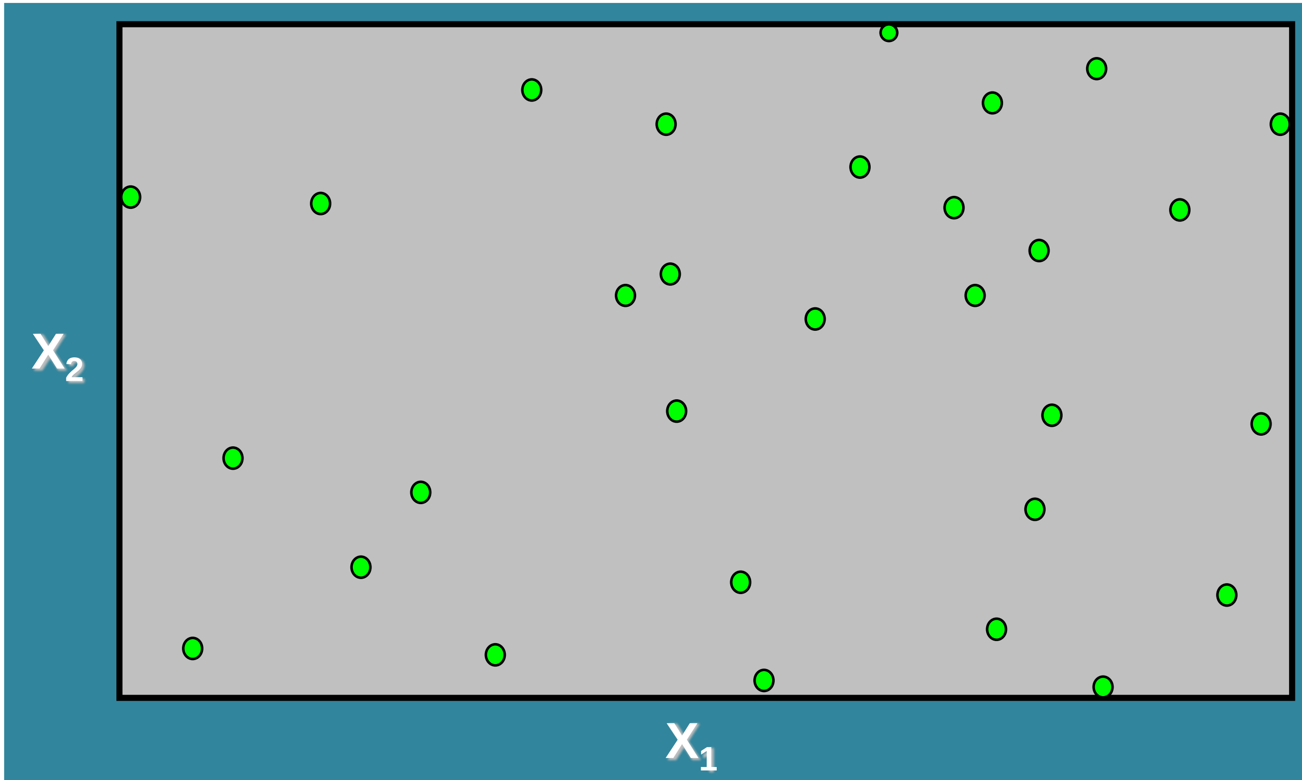
Knot Selection

- Mixed models smoothing corresponds to a penalized spline fit, which is much less sensitive to the number of knots than the unpenalized spline fit.
- Recommendations for knot selection include the following:
  - one knot per four unique observations
  - no more than 20-40 knots for longitudinal problems
  - between 20-150 knots for spatial problems
  - choosing the knots according to sample quantiles
  - for irregular spaces, being certain that the most knots selected are within the range of the data

Knot Selection Methods

- *k-d* tree (default)
  - Partition recursively the space enclosing the values of the random effects by splitting cells at the median of the data in the cell.
  - Space knots according to the density of the data.
- Equal
  - Space knots evenly to get a regular grid of knots.
- Data
  - The knot coordinates are given in a separate data set.

k-d Tree Construction

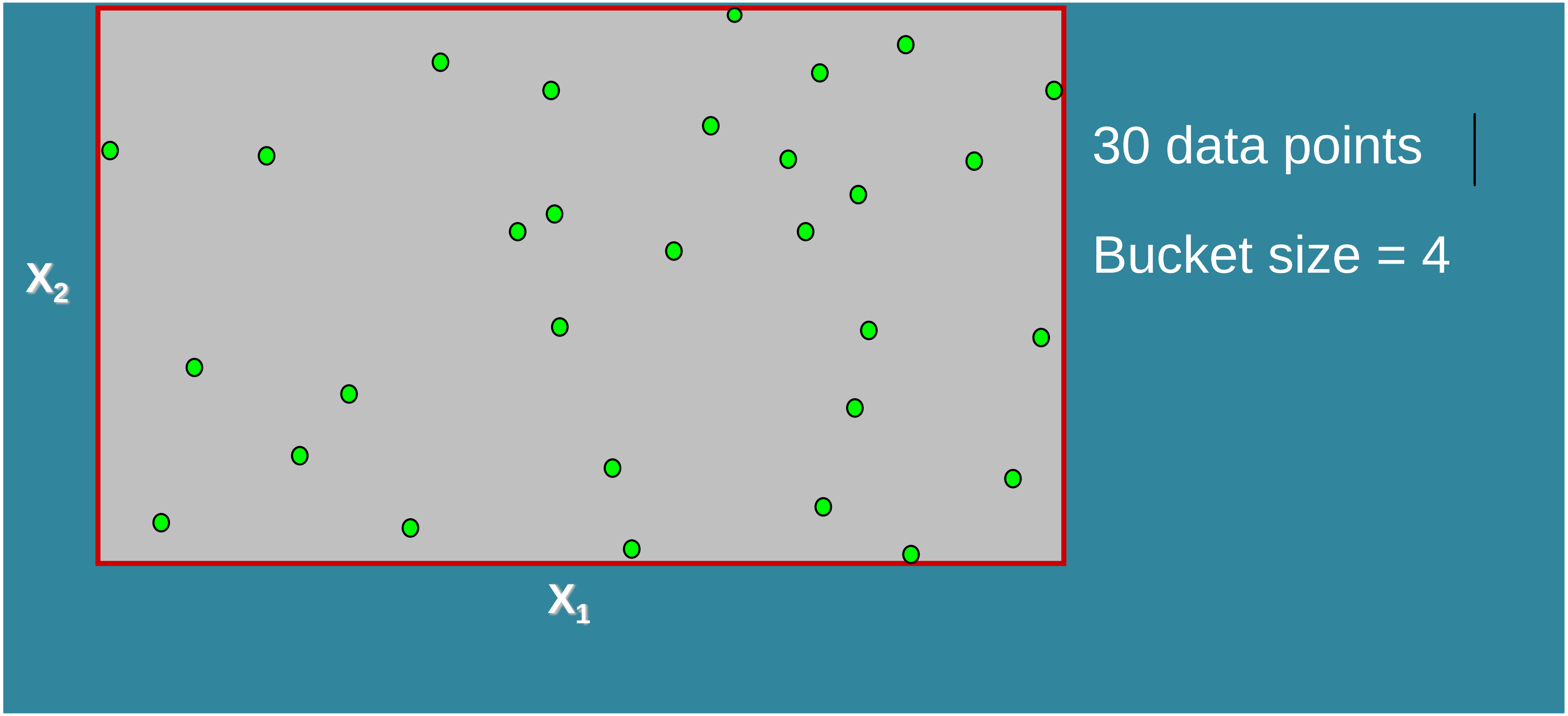


30 data points

Bucket size = 4

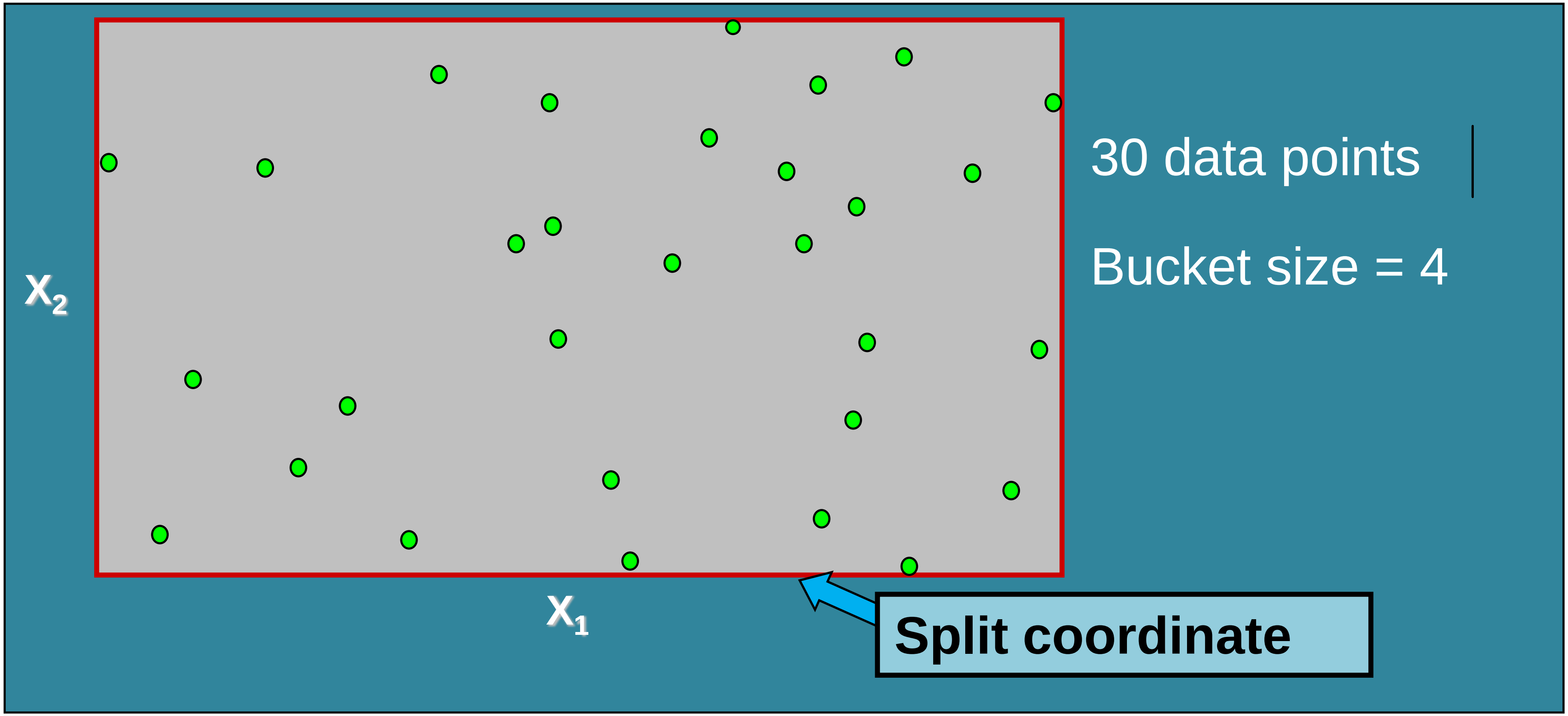
Step 1: Select the bucket size. The default is 4 for one random effect and at most 0.1*n for more than one random effect.

k-d Tree Construction



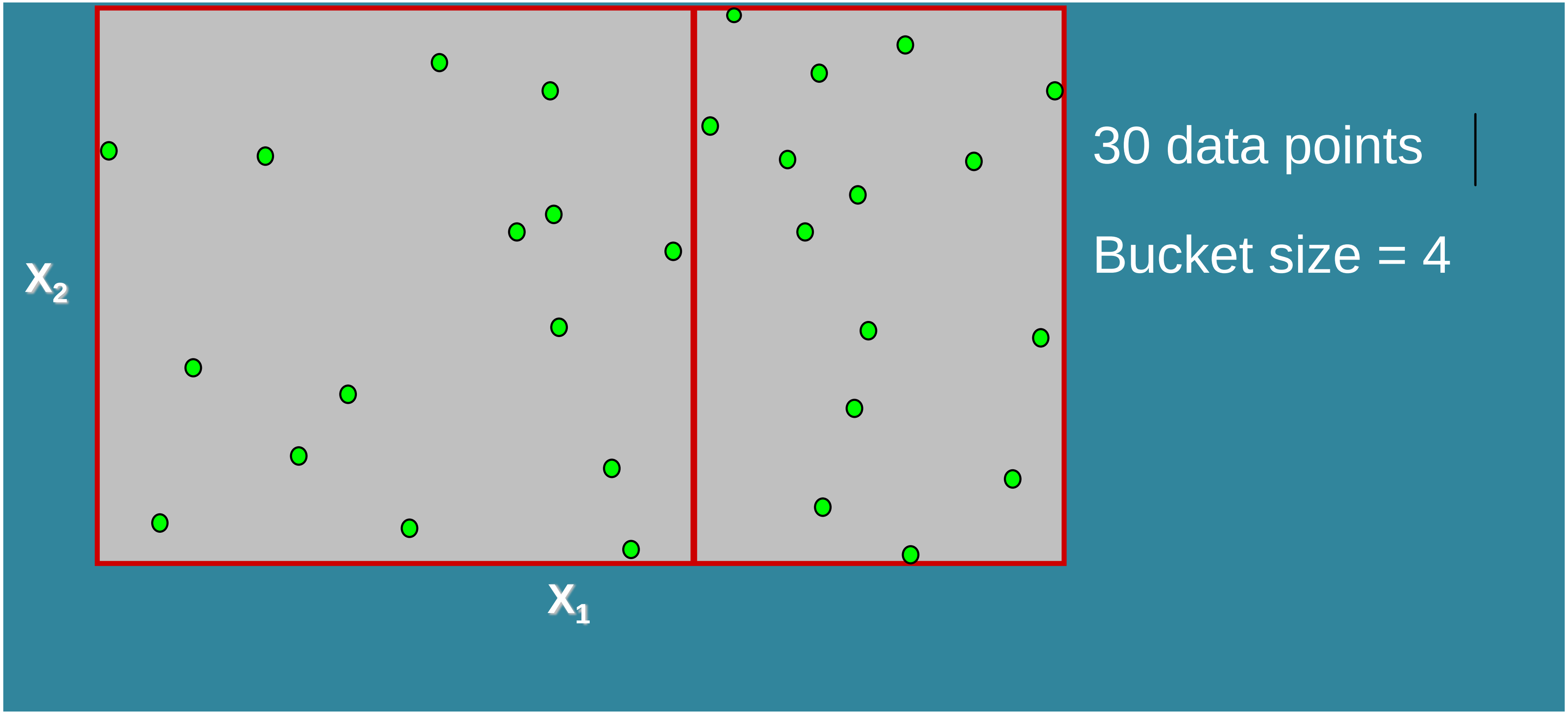
Step 2: Start with the smallest cell enclosing the predictor data. In the case of one independent variable, the cell is a line. For 2 independent variables, it is rectangle.

k-d Tree Construction



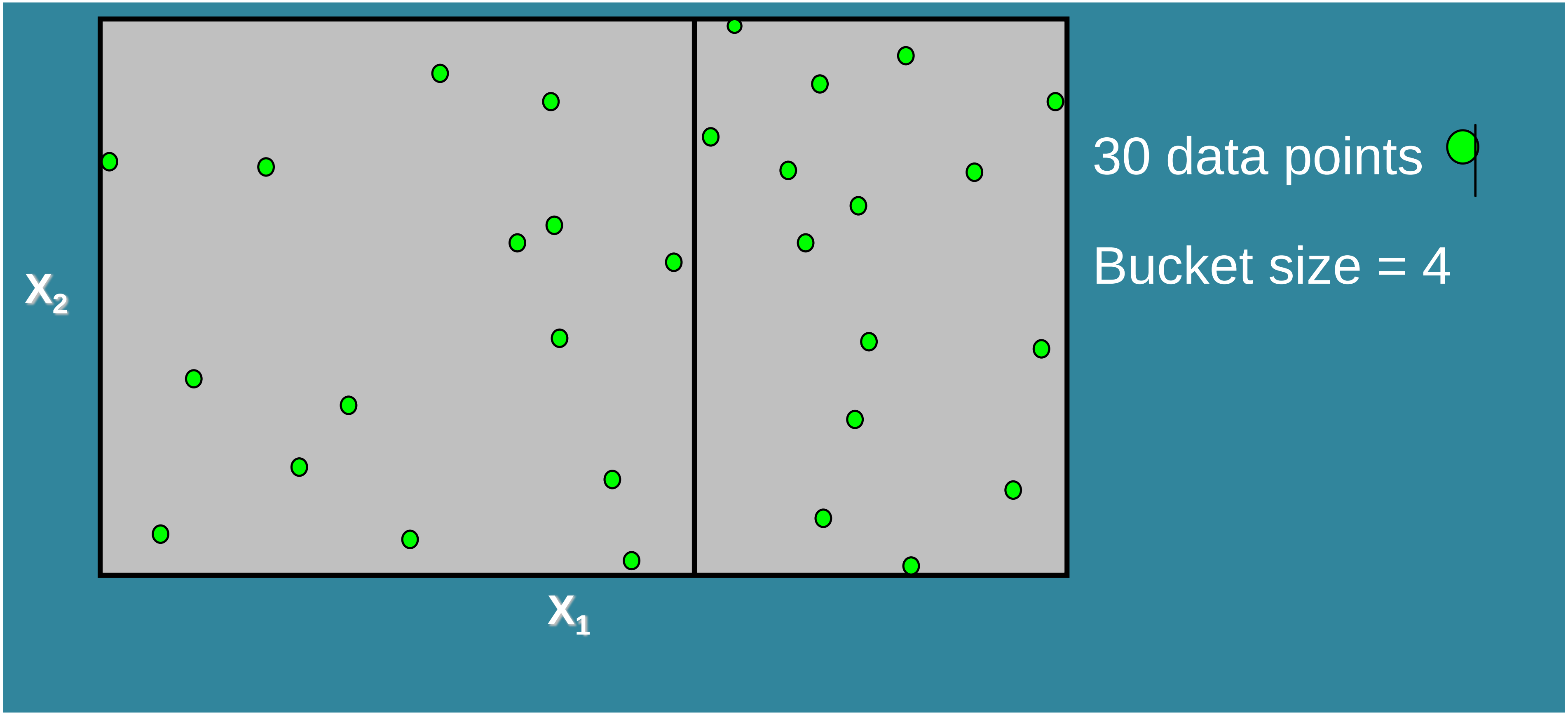
Step 3: Select the direction of the longest cell edge as the split coordinate.

k-d Tree Construction



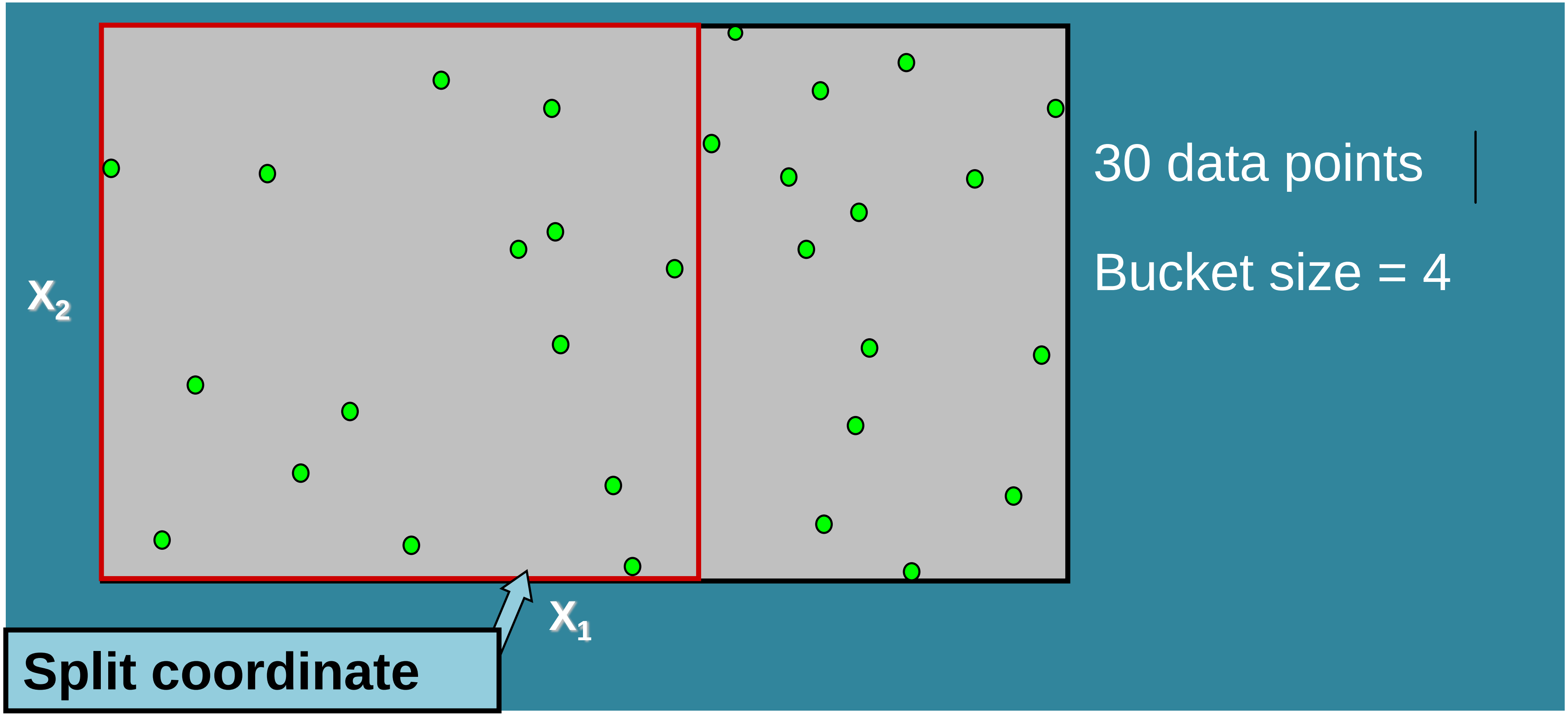
Step 4: Divide the cell in two at the median of the split coordinate of the data in the cell.

k-d Tree Construction



Step 5: Repeat the process for each child cell until all cells have, at most, the number of points specified in the bucket size (four points for this example).

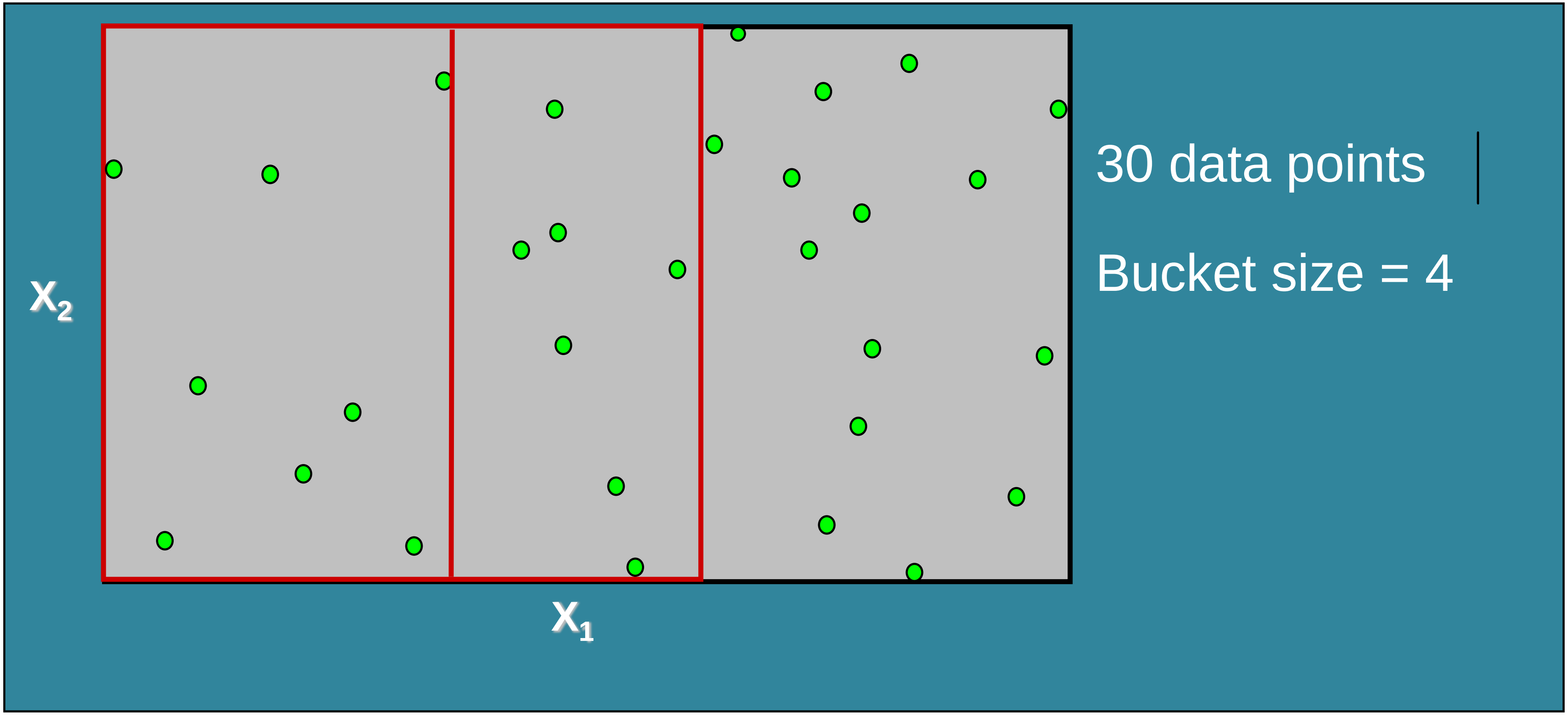
k-d Tree Construction



Step 5: Repeat the process for each child cell until all cells have, at most, four points.

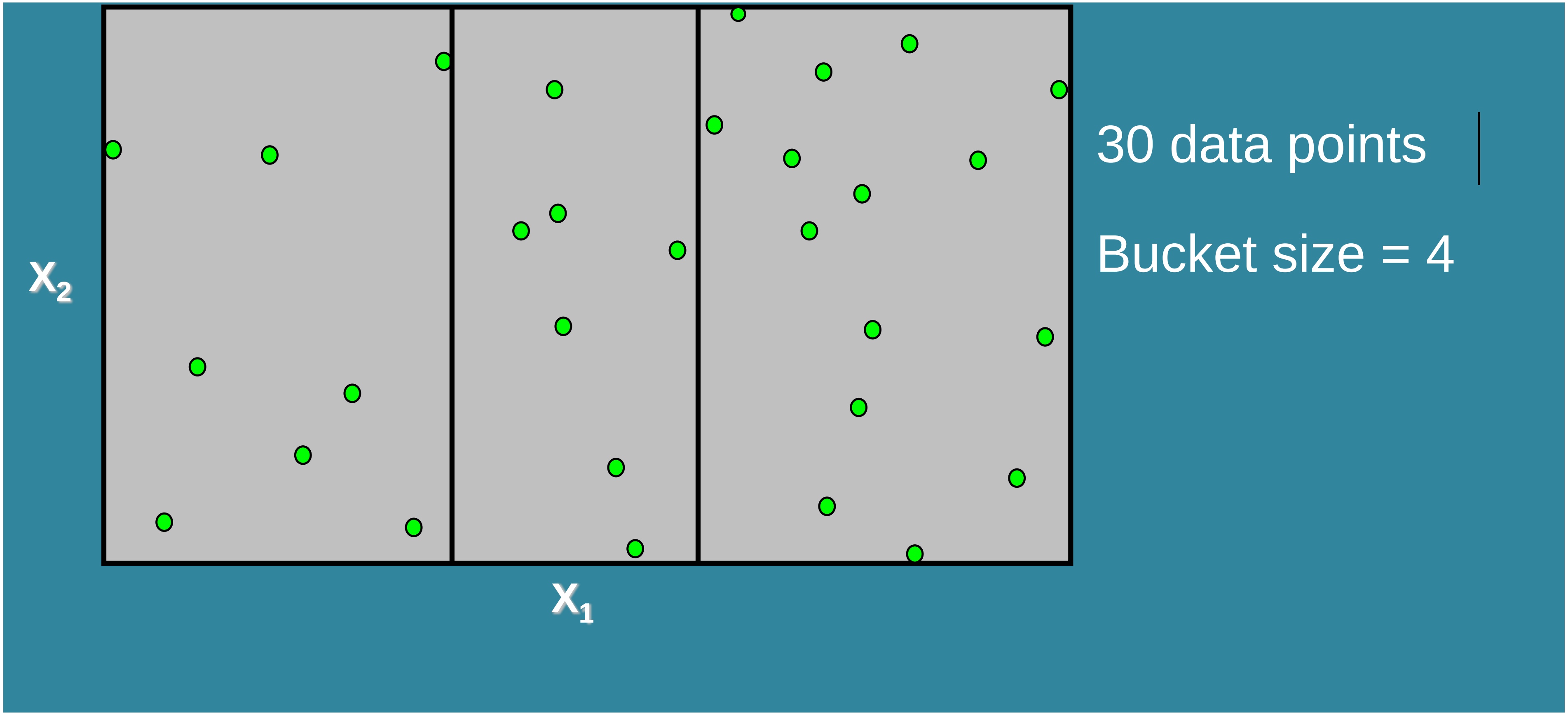
STATISTICS
TEXAS A&M UNIVERSITY

k-d Tree Construction



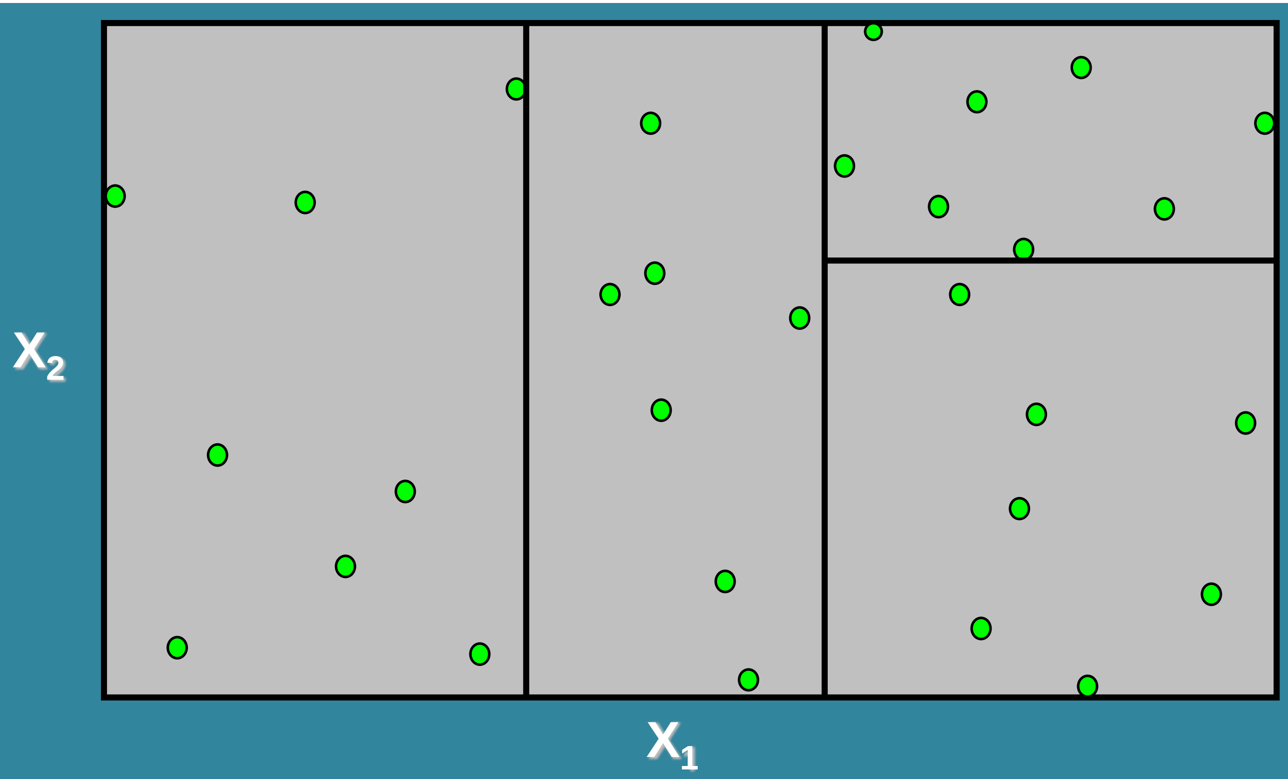
Step 5: Repeat the process for each child cell until all cells have, at most, four points.

k-d Tree Construction



Step 5: Repeat the process for each child cell until all cells have, at most, four points.

k-d Tree Construction

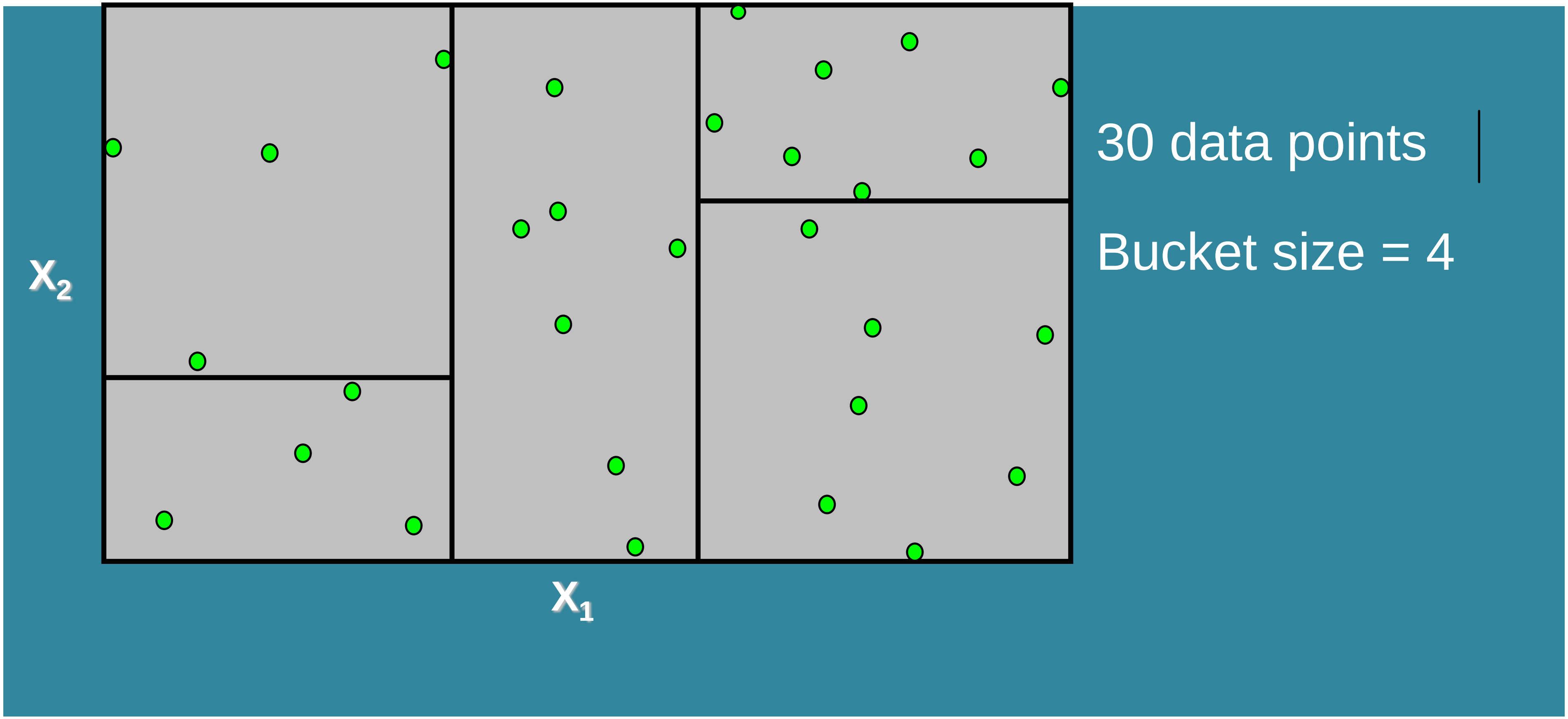


30 data points

Bucket size = 4

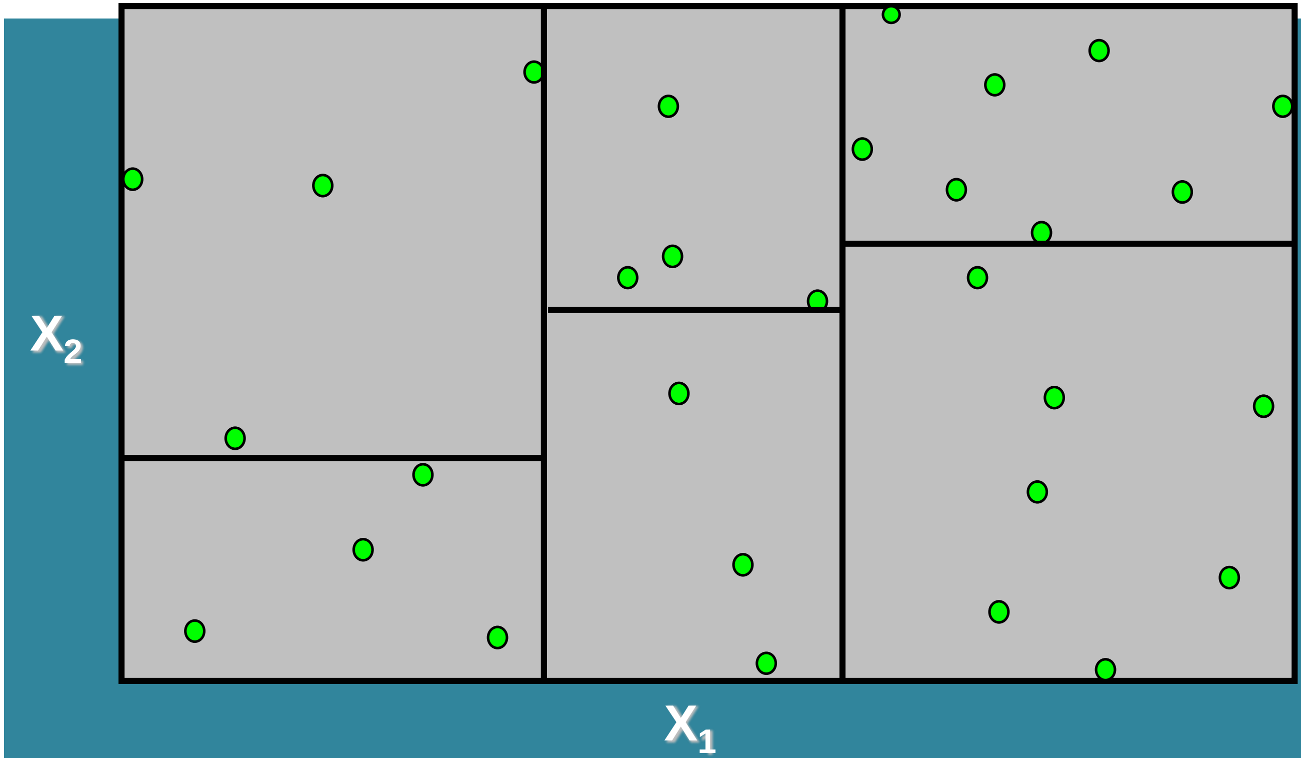
Step 5: Repeat the process for each child cell until all cells have, at most, four points.

k-d Tree Construction



Step 5: Repeat the process for each child cell until all cells have, at most, four points.

k-d Tree Construction

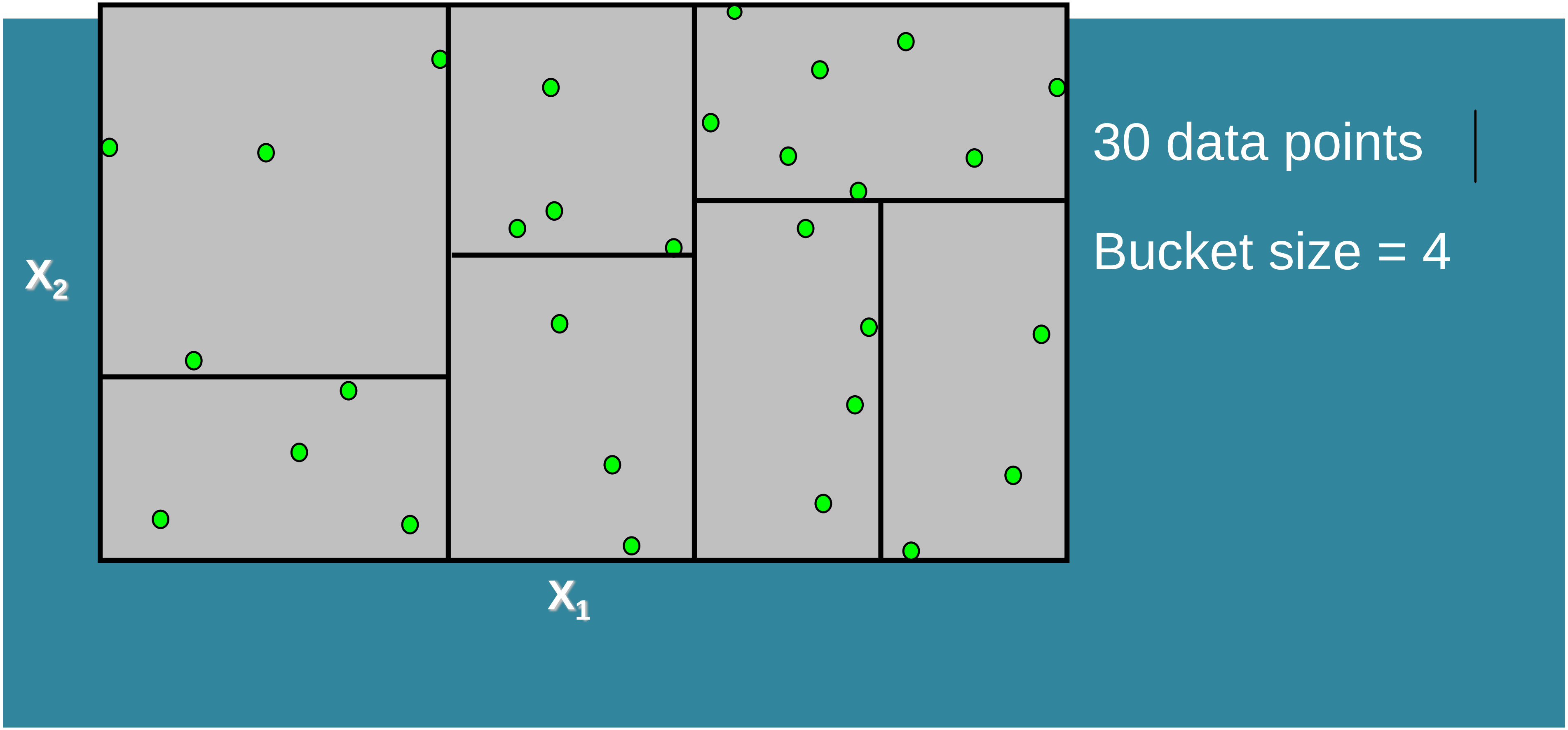


30 data points

Bucket size = 4

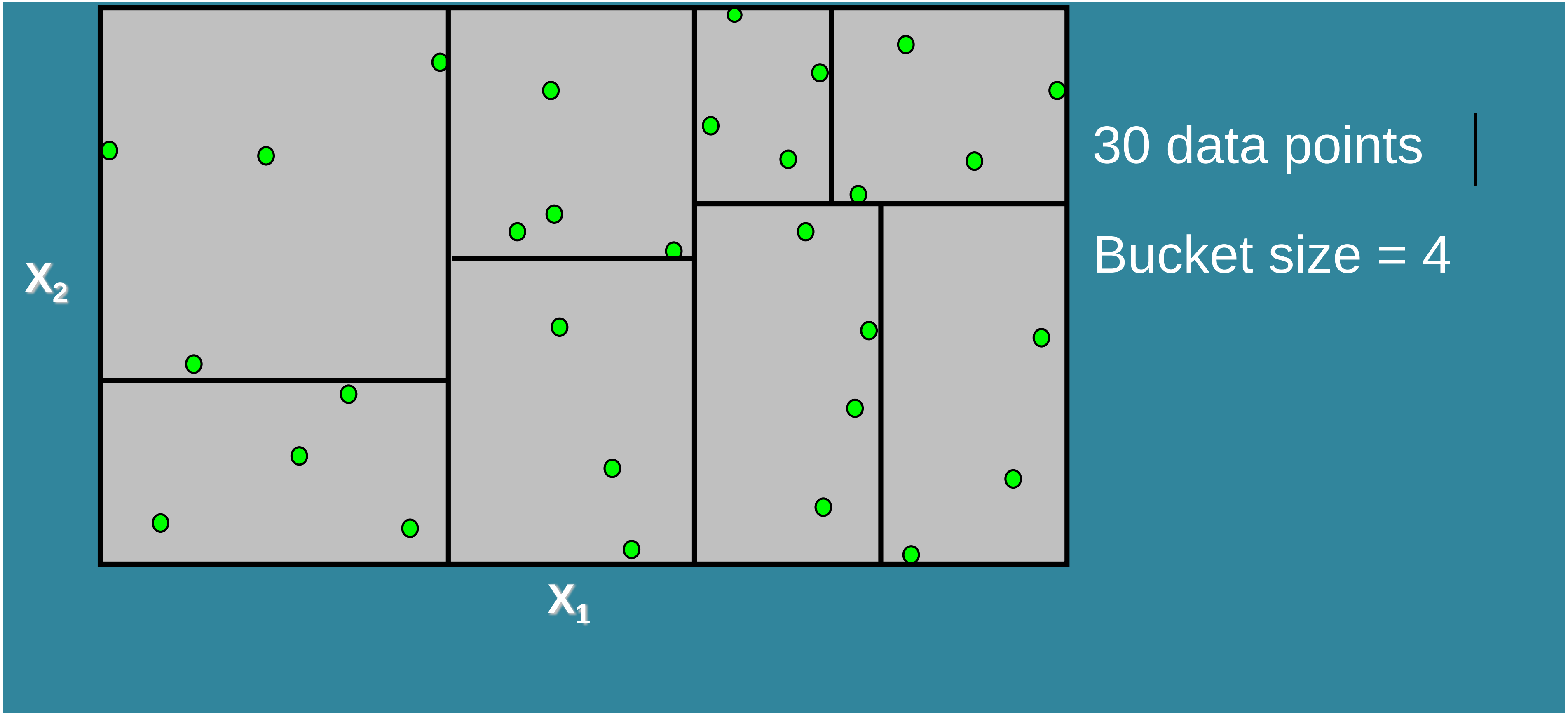
Step 5: Repeat the process for each child cell until all cells have, at most, four points.

k-d Tree Construction



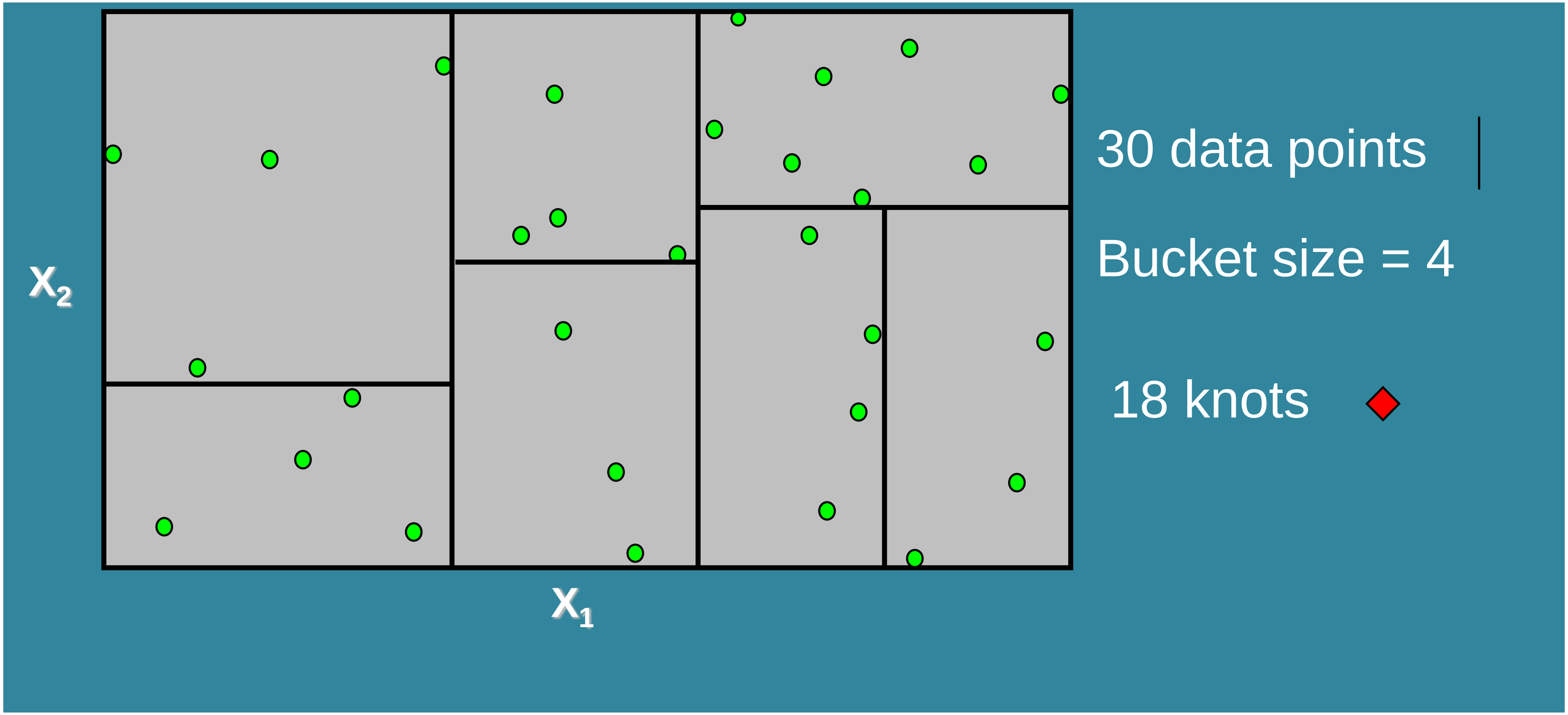
Step 5: Repeat the process for each child cell until all cells have, at, most four points.

k-d Tree Construction



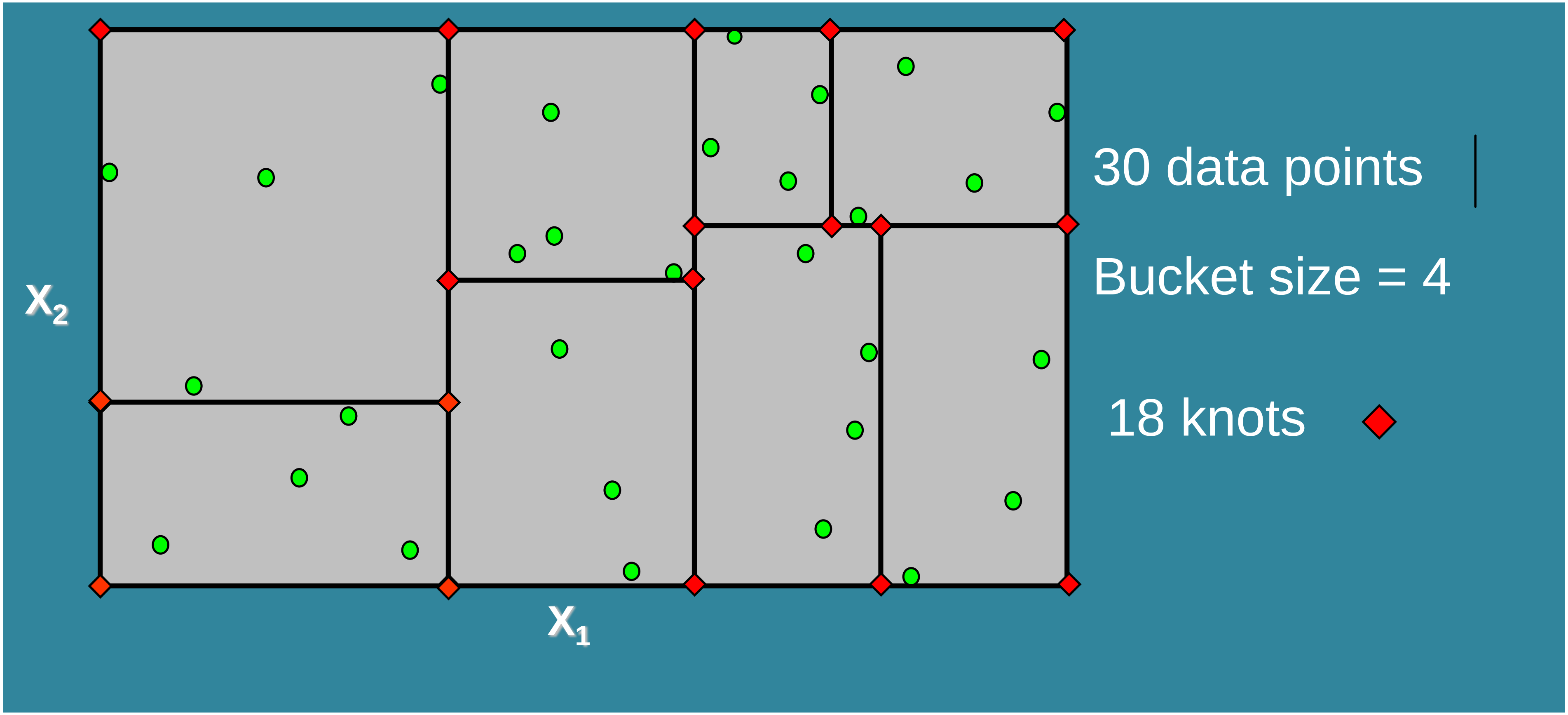
Step 5: Repeat the process for each child cell until all cells have, at most, four points.

k-d Tree Construction



Step 5: Repeat the process for each child cell until all cells have, at most, four points.

k-d Tree Construction



Step 6: The corners of the cells of the final *k-d* tree are used as the knots (fit points).

Question

Even though mixed model smoothing can be less sensitive to knot selections than some other smoothing methods, knot selection is important.

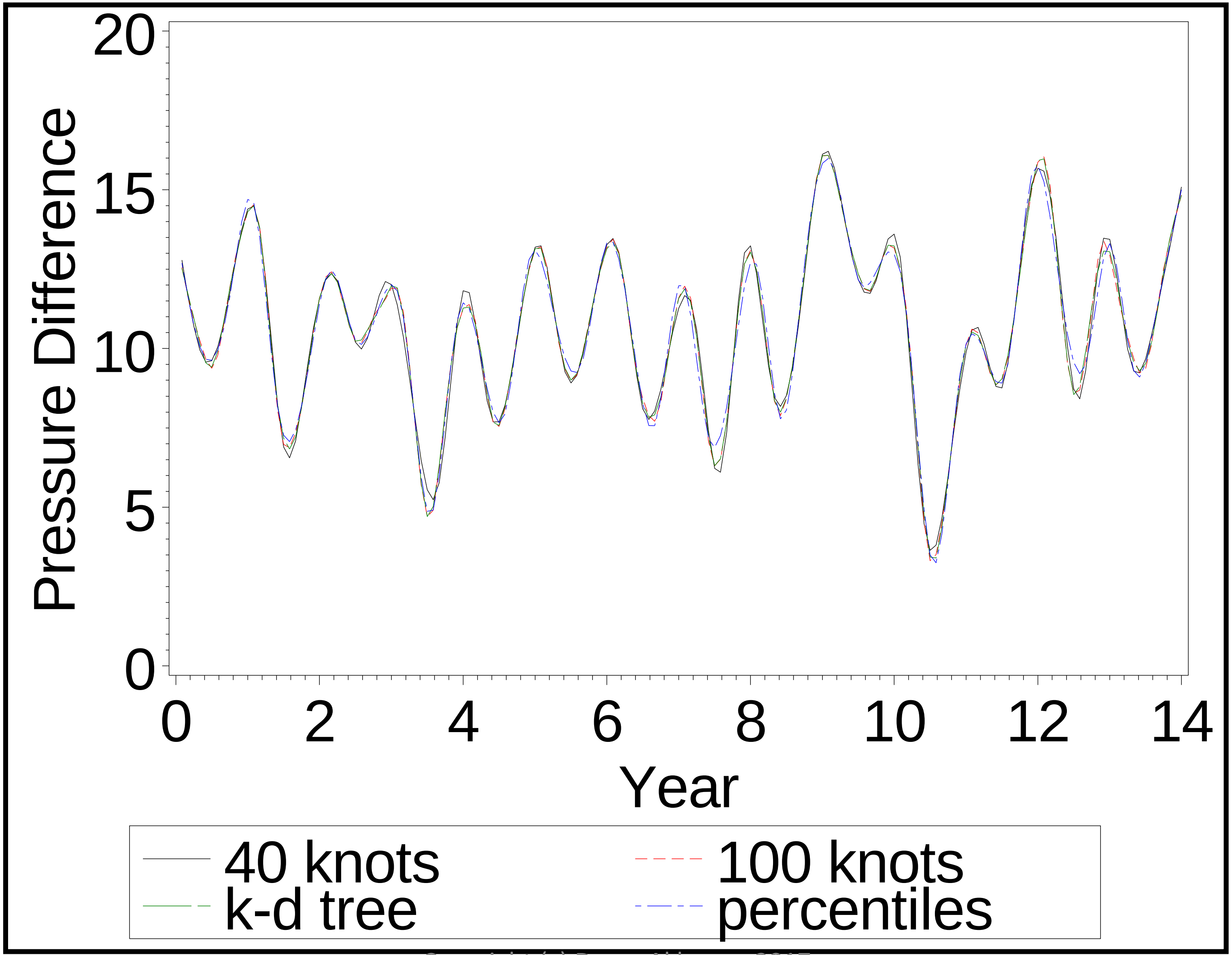
- True
- False

Fitting a Radial Smoother Model Using Different Knots

This demonstration illustrates concepts discussed previously.

ensoexample

Model Fit Using Different Knots



Summary of Radial Smoother Demonstrations

- The variance of the spline coefficients affects the EBLUPs of the spline coefficients and also the smoothness of the fitted curve for a given number of knots.
- The fit of the radial smoother model might or might not vary much for a different number of knots and/or different locations of the knots.
  - The number of the knots seems to have a greater impact on model performance and model fit than the location of the knots.
  - As long as the number of knots is sufficient, the models are very similar.
  - It might be safer to have more knots than necessary as long as it is numerically feasible.

Question

Which of the following is **false**?

a. Radial smoothing might be more flexible than spatial covariance structures.
b. Radial smoothing is useful in modeling the variations that might not be modeled adequately using other structures.
c. Radial smoothing is unstable, and is not recommended.

Question

The radial smoother is different from the spline function because of which condition?

a. With the radial smoother, the spline coefficients are random effects.
b. With the radial smoother, the degree of smoothness is a function of a variance component, and is therefore demystified.
c. With the radial smoother, the radial-basis spline function is used.
d. All of the above
e. None of the above