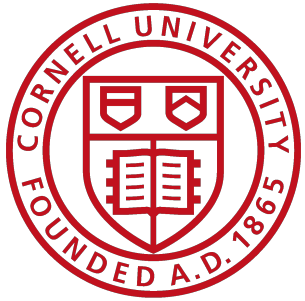
A near real-time forecasting system of US crop yields

Ariel Ortiz-Bobea (et al. soon)

TWEEDS Workshop
March 29, 2019, Portland, OR



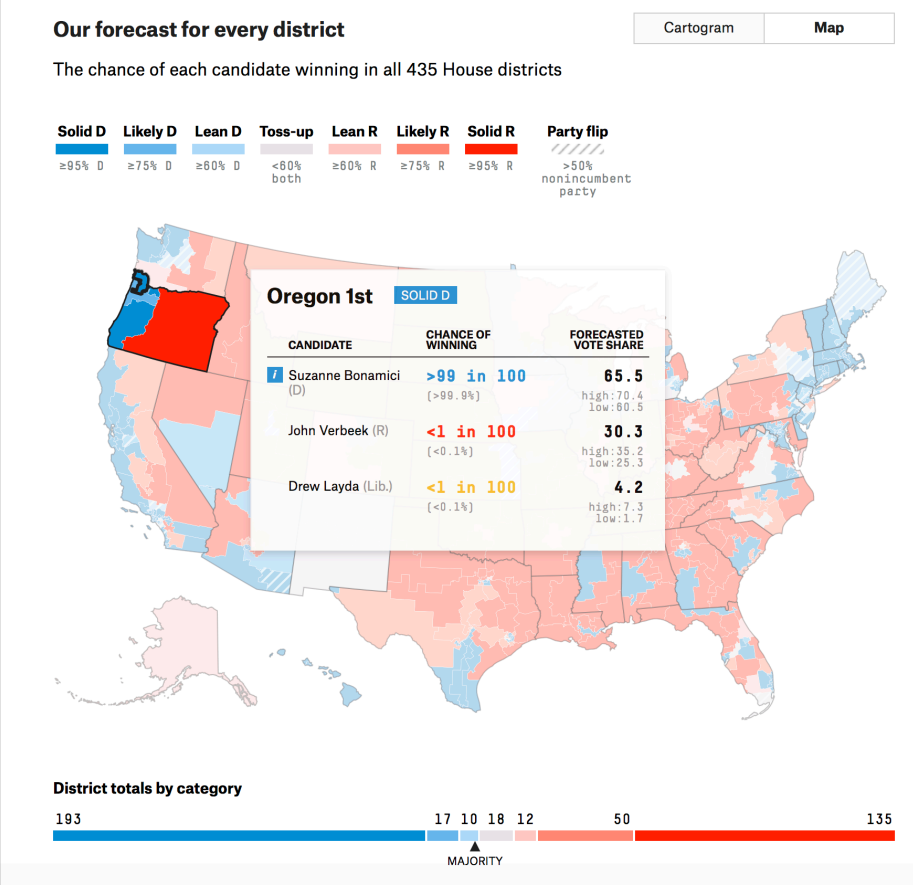
Motivation

- Academic work on ag-climate has largely focused on future (or "recent") impacts of climate change
- Fewer insights related to *ongoing* climatic phenomena
- Focus on such events has a different (broader) audience
- Is focus on the present the gateway for effectively raising awareness about the future?

Floodwaters surround corn sitting under a collapsed grain bin in this aerial photograph over Thurman, Iowa, on 3/23/2019. Bloomberg photo by Daniel Acker.

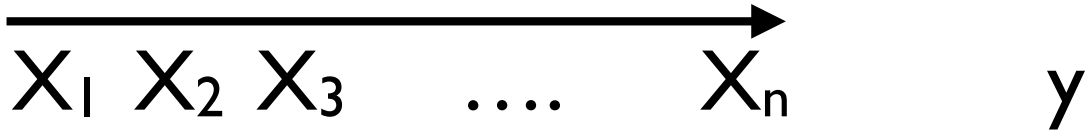
A "538" for ag production

- A portal for ag production forecasts
- Features: daily forecasts, publicly available data, county/state/ national level forecasts, insights on drivers, scalable
- Preliminary work to support proposal for Cornell's Digital Ag initiative
- Team: economics, statistics, computer science, climate science, remote sensing



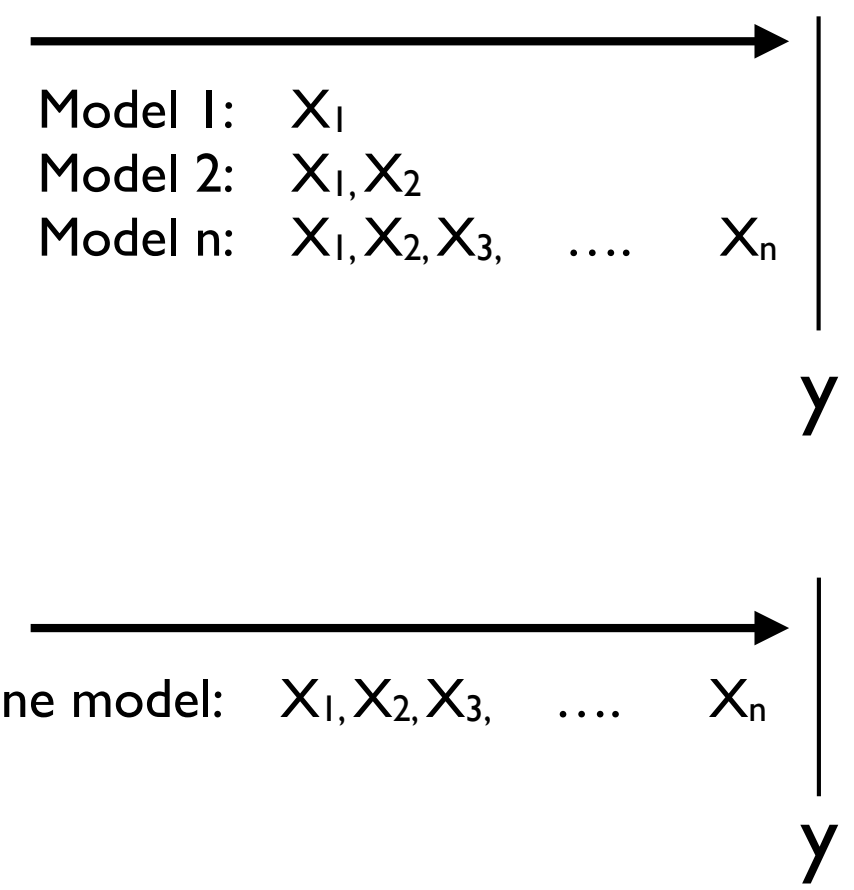
The forecasting problem

- Publicly available data (county level corn yields)
- Low frequency dependent variable (annual)
- 2000+ counties, since 1980s
- Yield level known at the end of the season
- High and mixed temporal frequency of covariates (observed/forecasted daily weather, soil moisture, bi-weekly/monthly remote sensing observations)
- Regional heterogeneity



Modeling choice

- Estimate separate independent models as time progresses? (**serial model approach**)
  - Pro: traditional framework; probably better forecasts
  - Cons: difficult to interpret
- Estimate a single model for all time periods? (**single model approach**)
  - A forecast requires imputing unrealized values of covariates
  - Pros: easier to interpret
  - Cons: less accurate forecasts

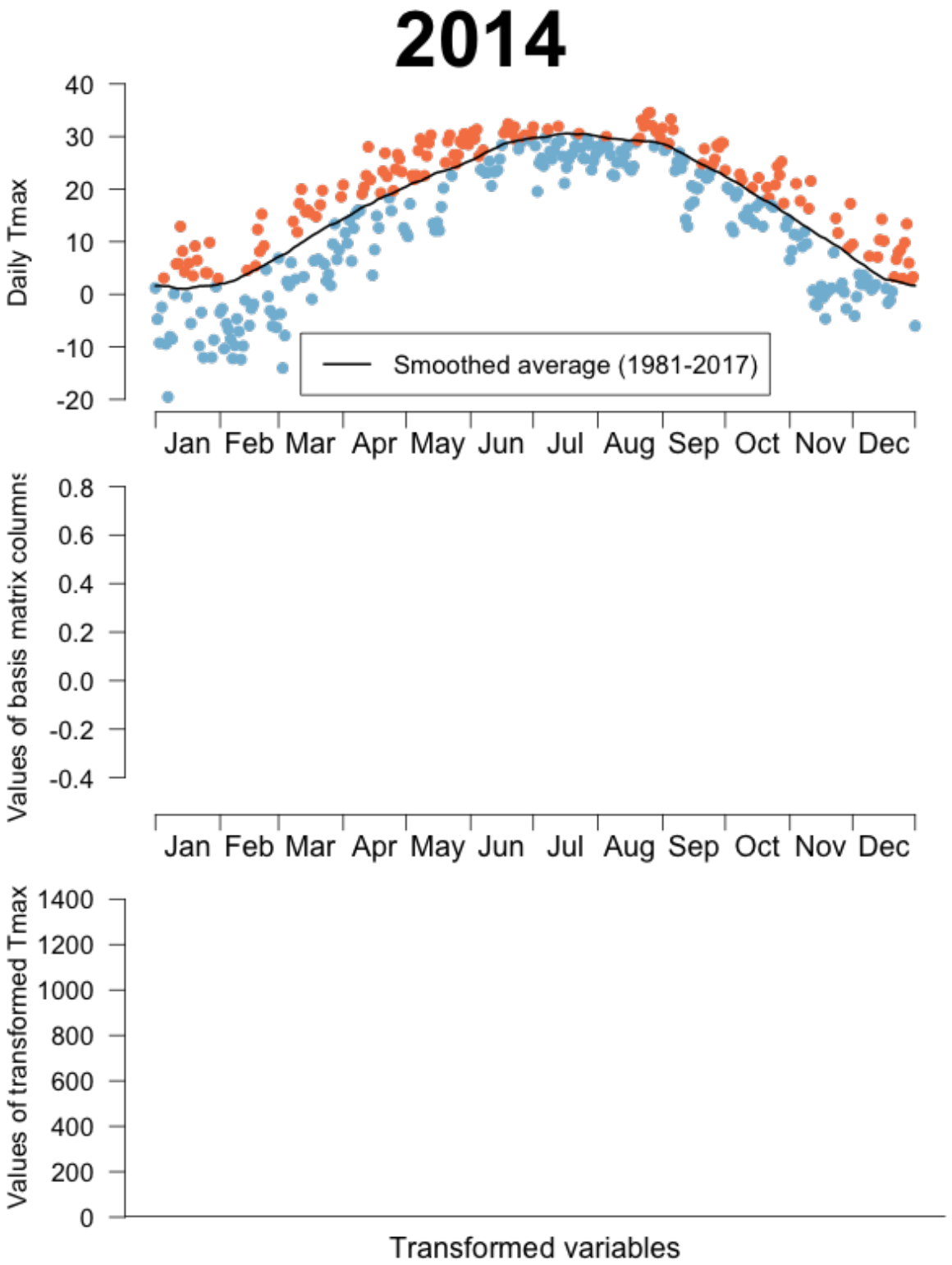


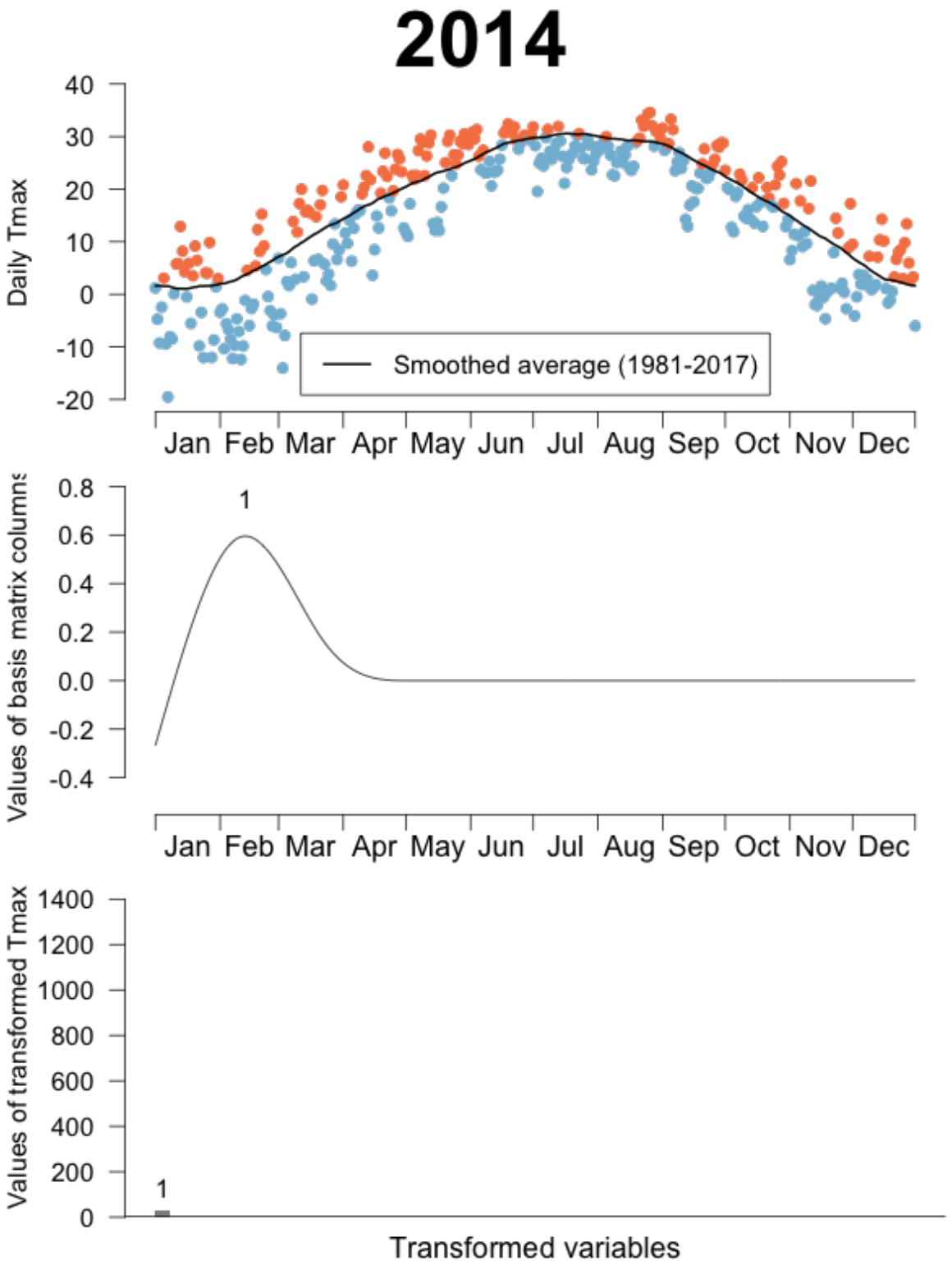
The single model approach

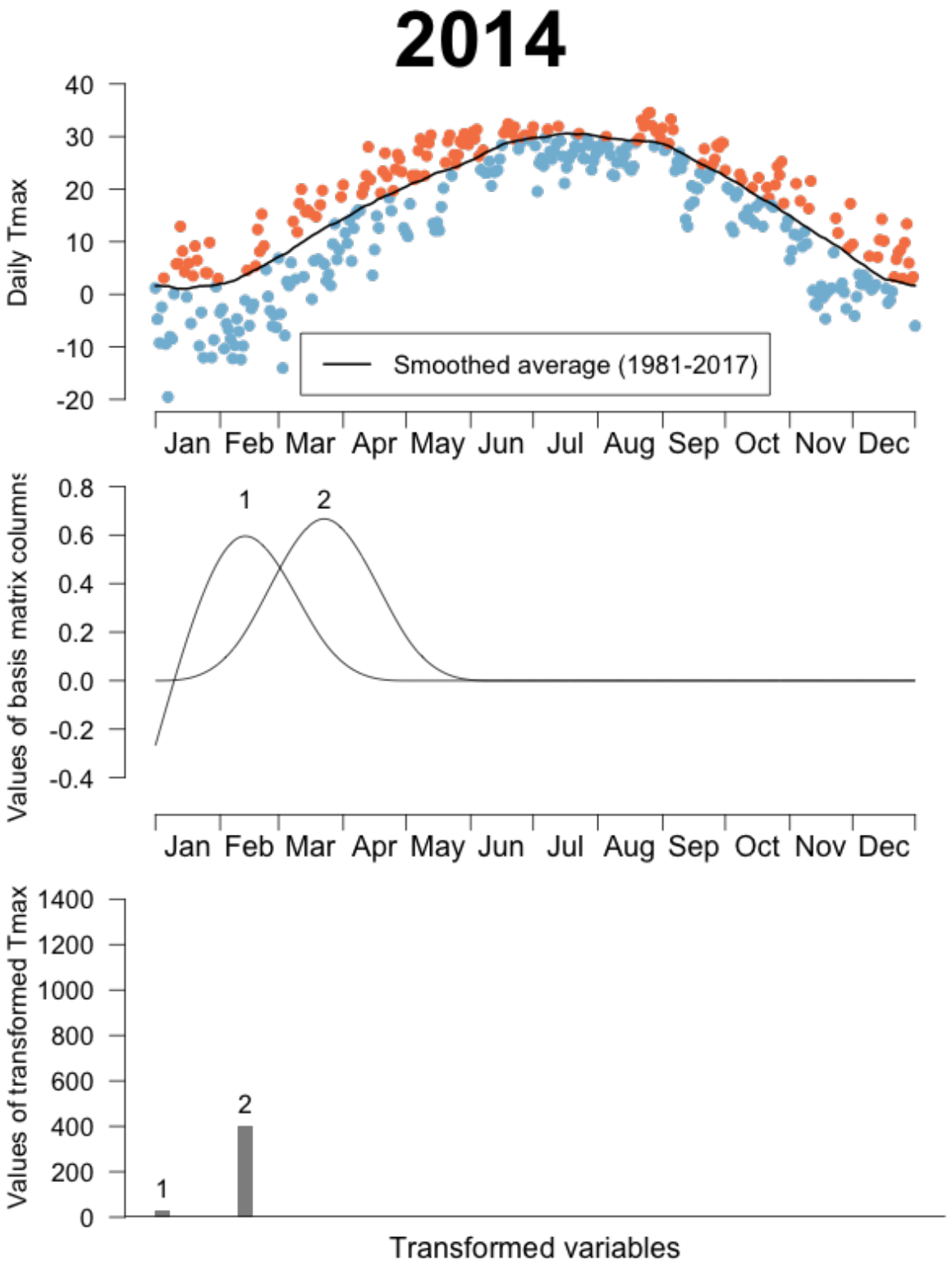
Spline panel model

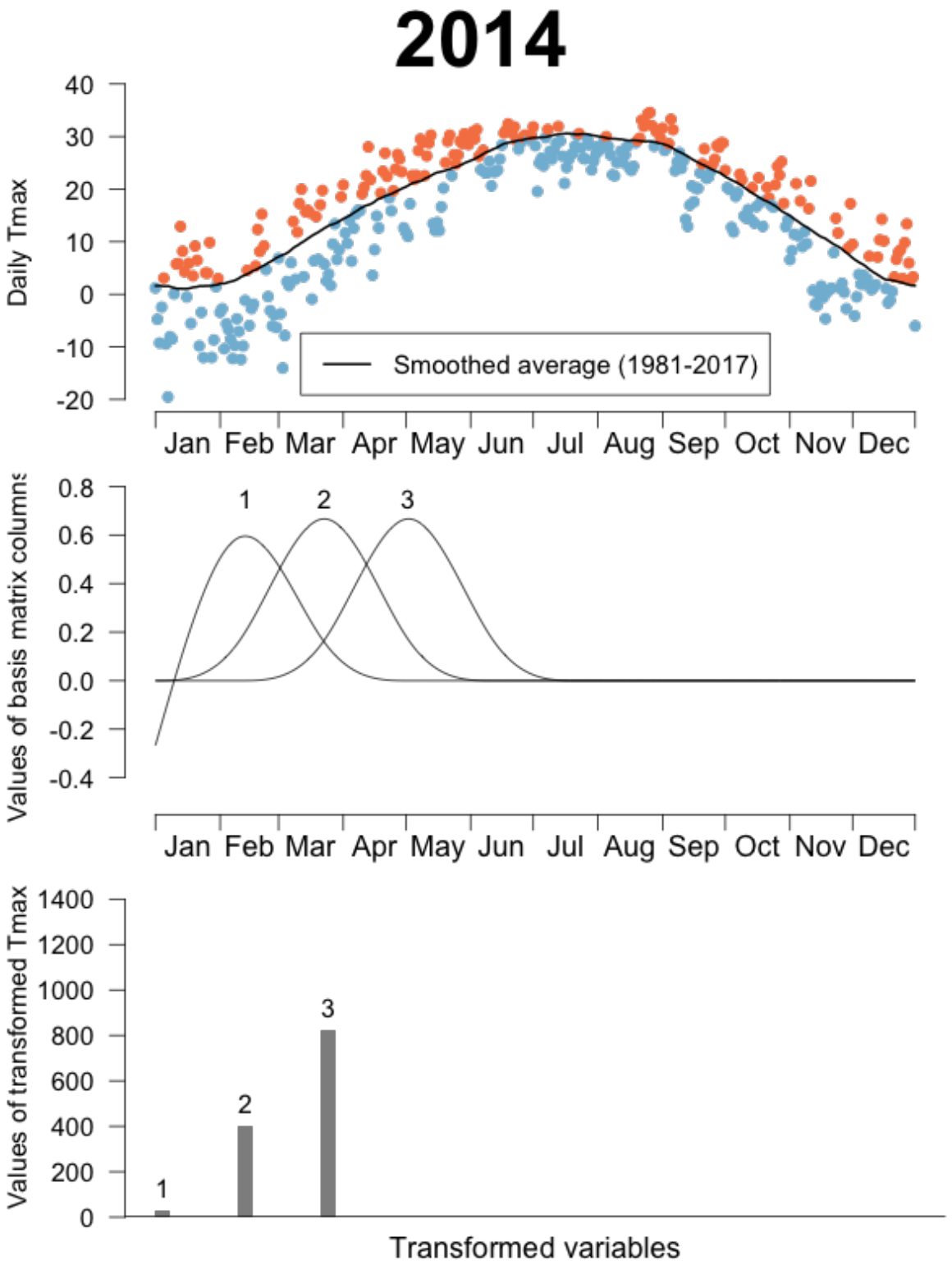
- Daily weather data transformed with a natural cubic spline
- Neighboring days have similar effect on yield
- Estimate single model to obtain coefficients
- Forecast = prediction with hybrid matrix of covariates composed of realized weather and future (imputed) weather
- Forecasts have 2 sources of uncertainty:
  - Model (year-block bootstrap)
  - Weather (bootstrap of weather from past years)
- Future implementation:
  - Tensor splines for effects varying over level and time
  - Higher-order singular value decomposition of tensor spline
  - Weather forecasts + representation of unrealized weather

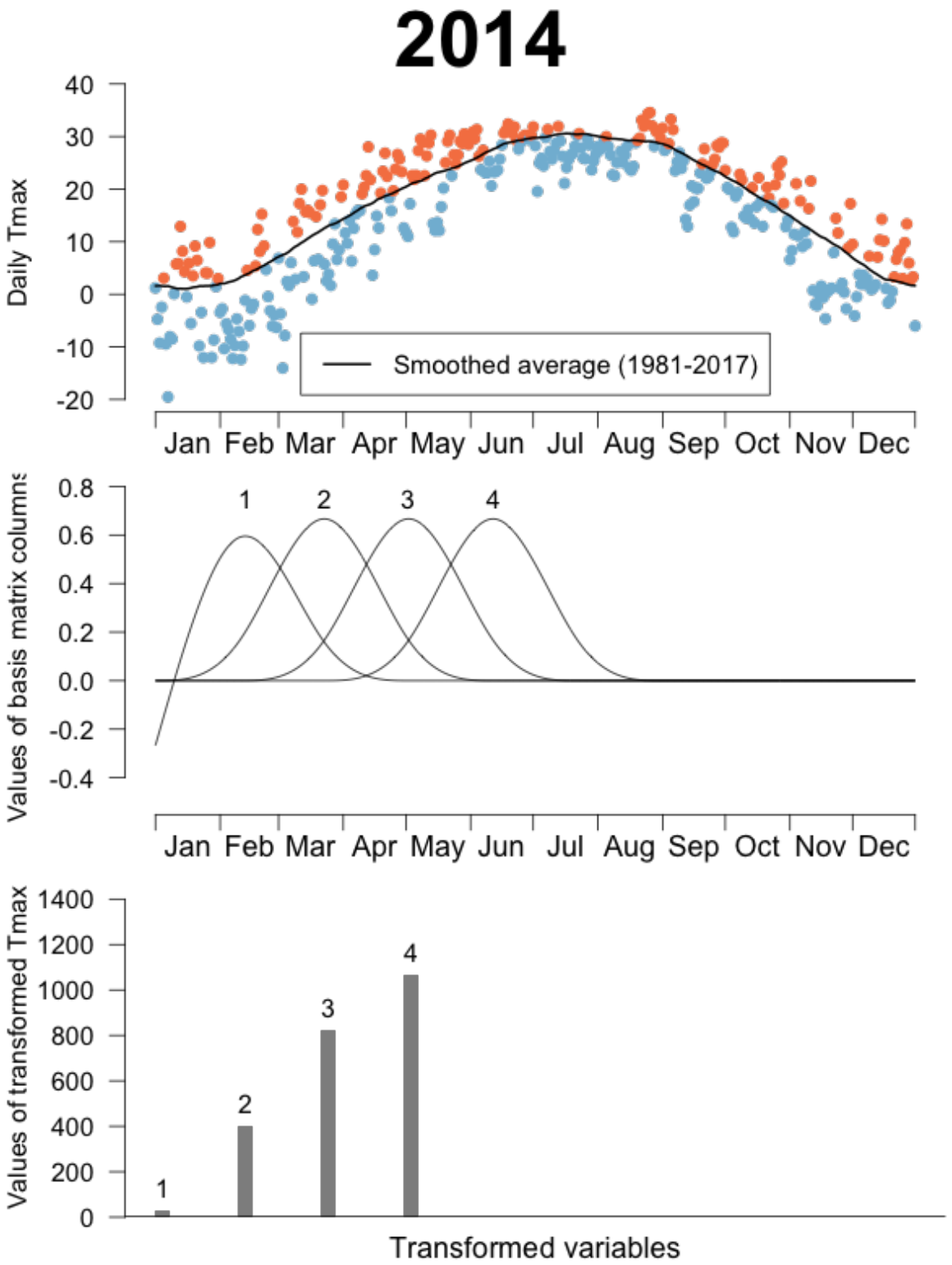
Transforming daily weather with a spline

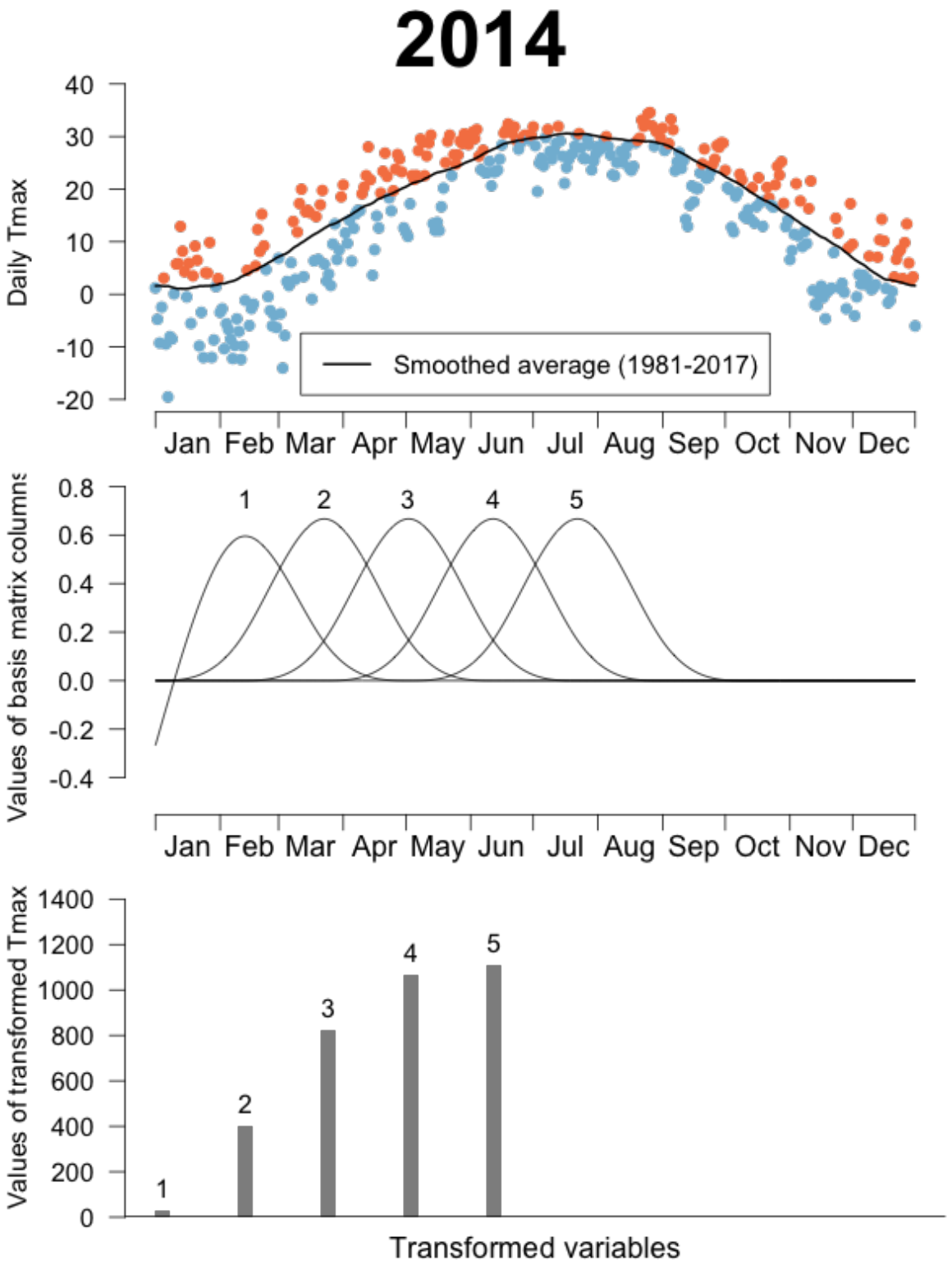

2014
Daily Tmax
40
30
20
10
0
-10
-20
Smoothed average (1981-2017)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of basis matrix columns
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of transformed Tmax
1400
1200
1000
800
600
400
200
0
Transformed variables

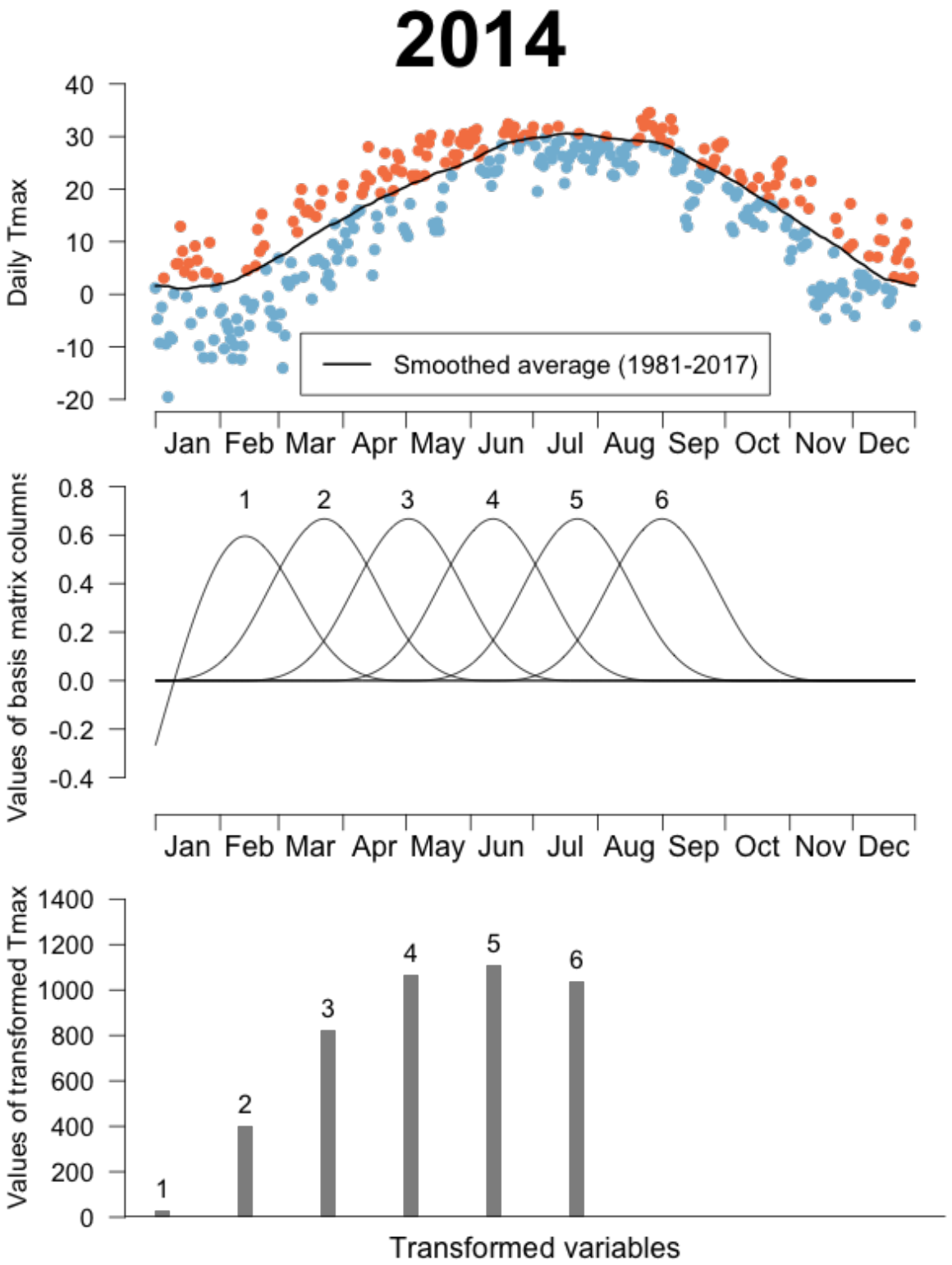

2014
Daily Tmax
40
30
20
10
0
-10
-20
Smoothed average (1981-2017)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of basis matrix columns
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
1
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of transformed Tmax
1400
1200
1000
800
600
400
200
0
1
Transformed variables

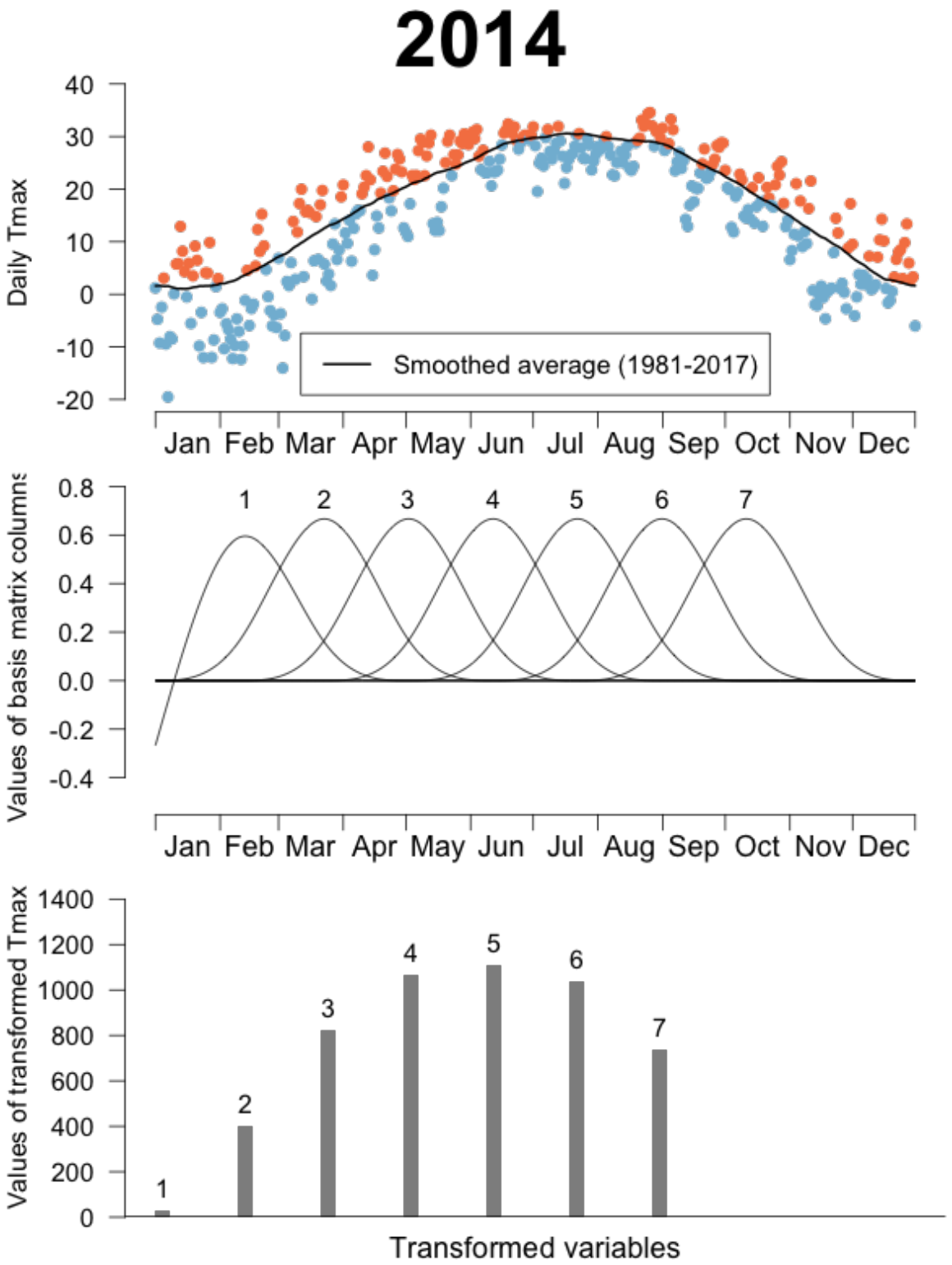

2014
Daily Tmax
40
30
20
10
0
-10
-20
Smoothed average (1981-2017)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of basis matrix columns
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
1
2
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of transformed Tmax
1400
1200
1000
800
600
400
200
0
1
2
Transformed variables

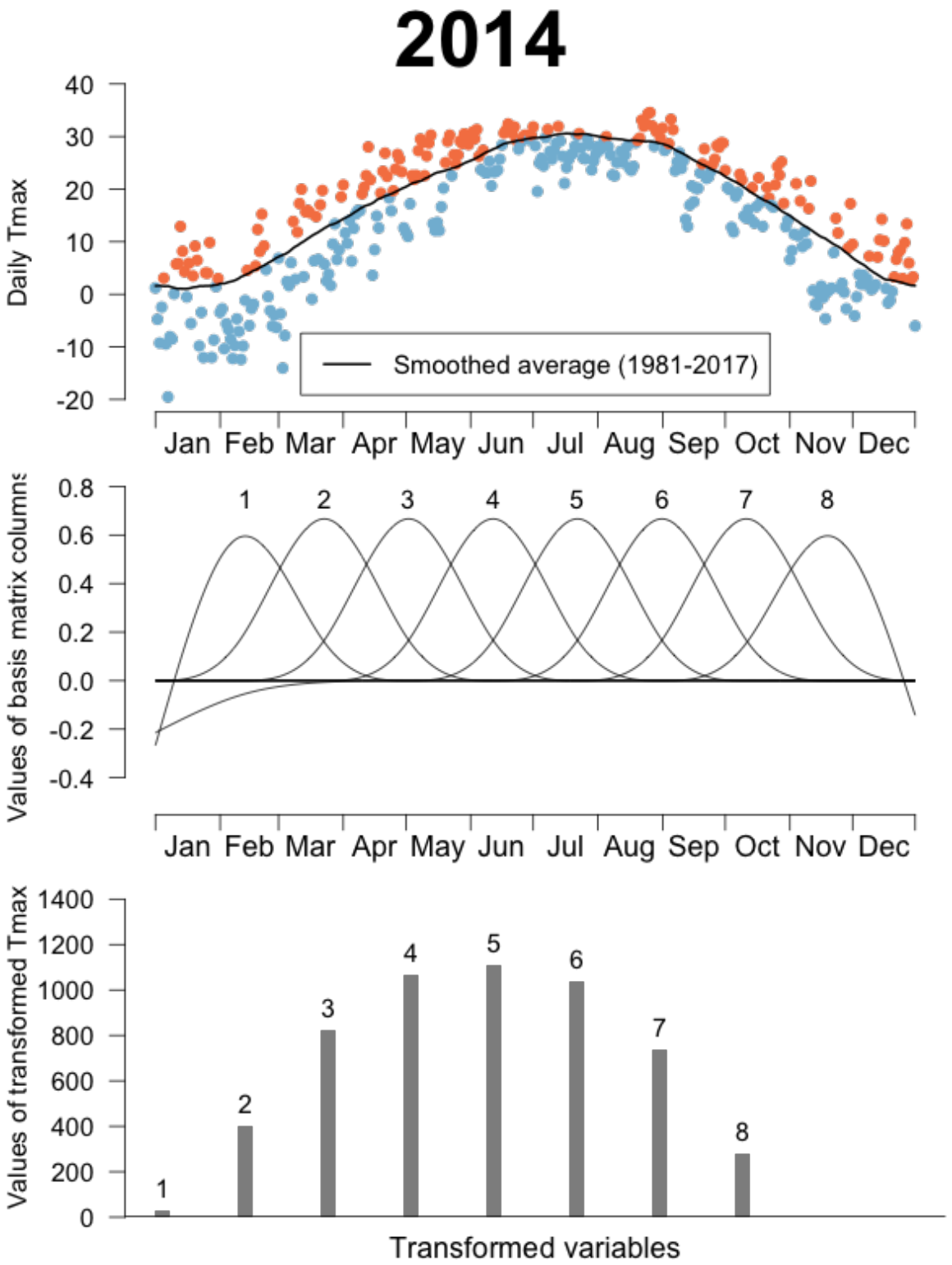

2014
Daily Tmax
40
30
20
10
0
-10
-20
Smoothed average (1981-2017)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of basis matrix columns
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
1
2
3
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of transformed Tmax
1400
1200
1000
800
600
400
200
0
1
2
3
Transformed variables

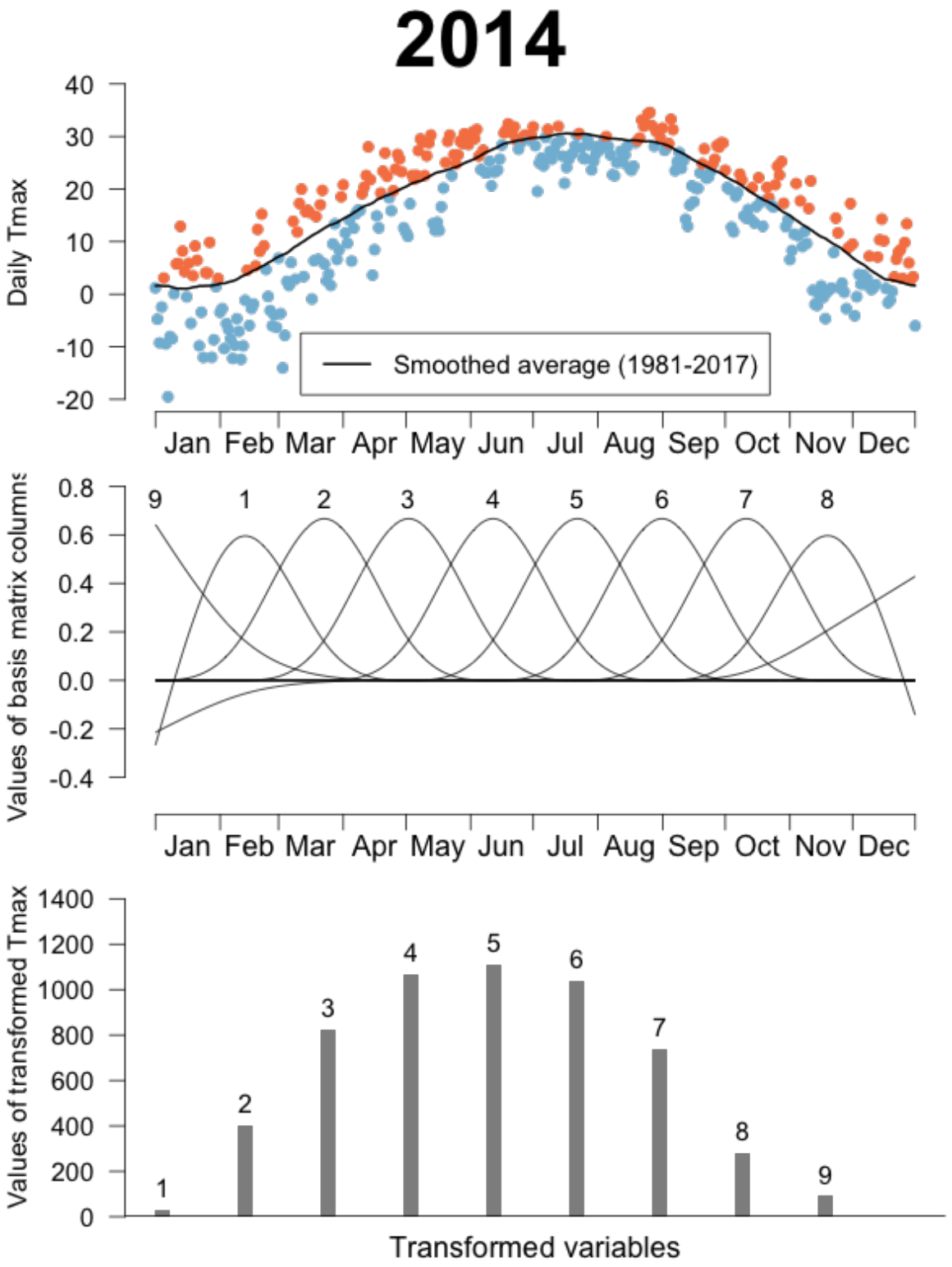

2014
Daily Tmax
40
30
20
10
0
-10
-20
Smoothed average (1981-2017)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of basis matrix columns
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
1
2
3
4
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of transformed Tmax
1400
1200
1000
800
600
400
200
0
1
2
3
4
Transformed variables

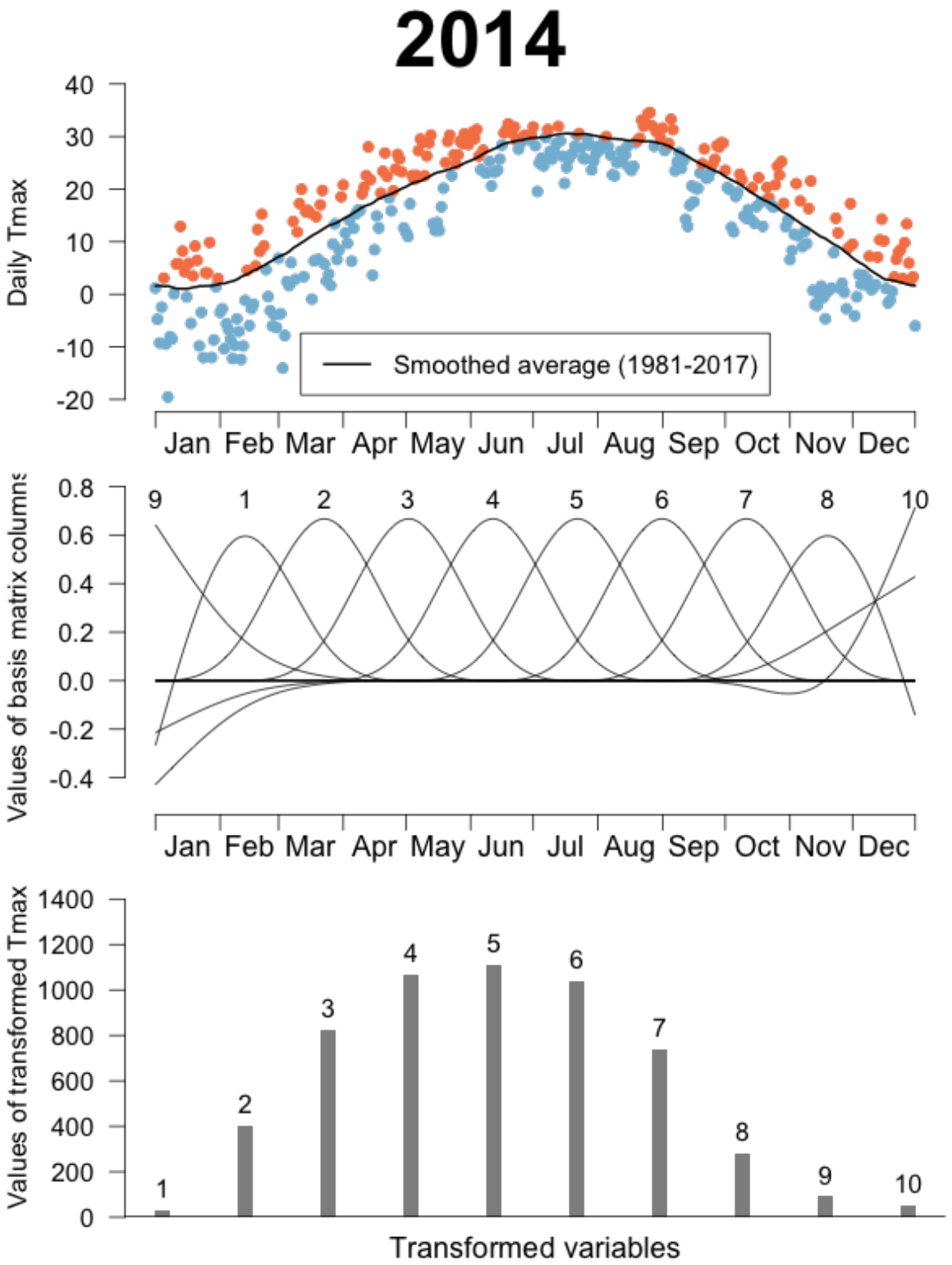

2014
Daily Tmax
40
30
20
10
0
-10
-20
Smoothed average (1981-2017)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of basis matrix columns
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
1 2 3 4 5
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of transformed Tmax
1400
1200
1000
800
600
400
200
0
1 2 3 4 5
Transformed variables

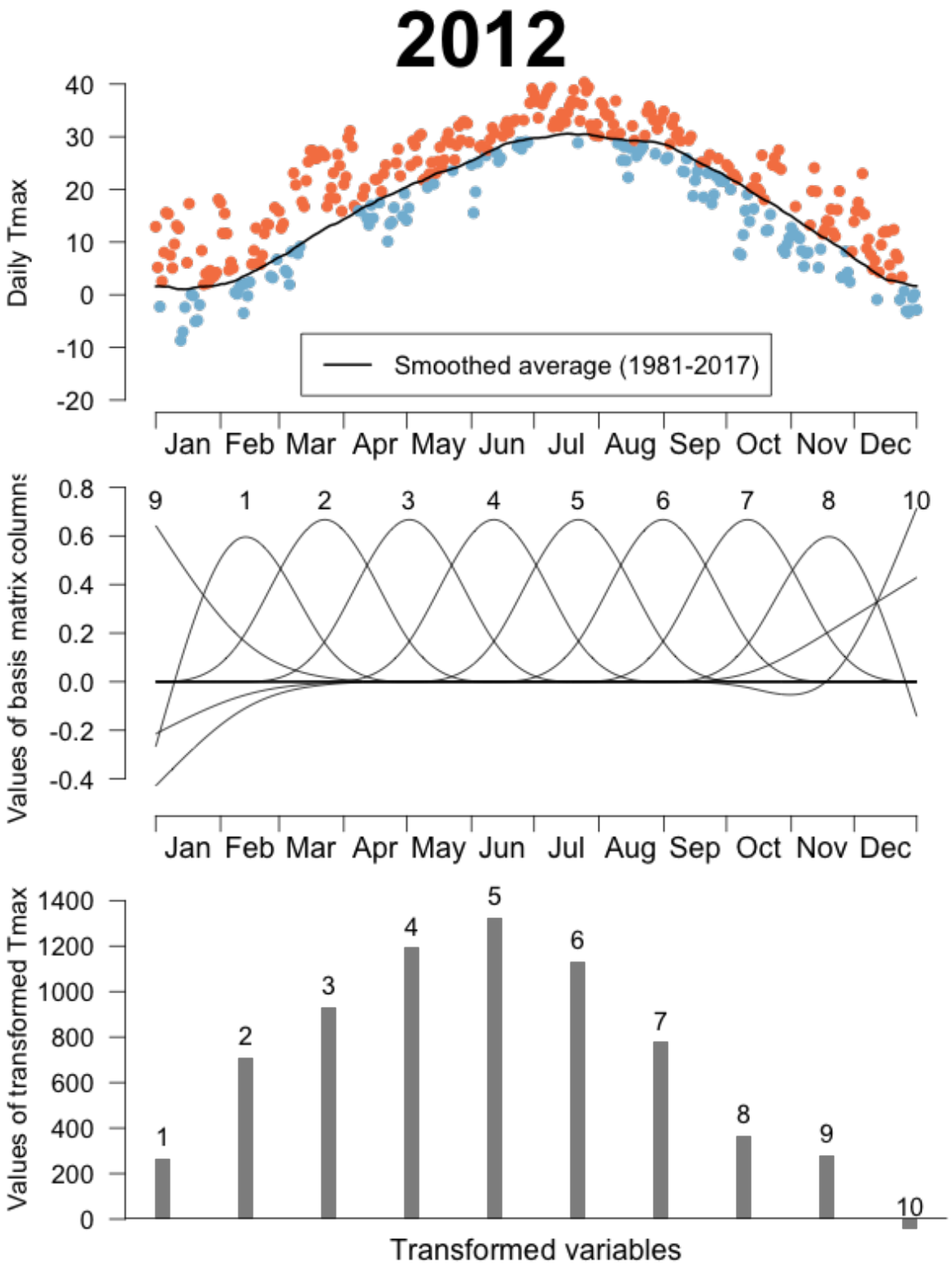

2014
Daily Tmax
40
30
20
10
0
-10
-20
Smoothed average (1981-2017)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of basis matrix columns
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
1 2 3 4 5 6
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of transformed Tmax
1400
1200
1000
800
600
400
200
0
1 2 3 4 5 6
Transformed variables

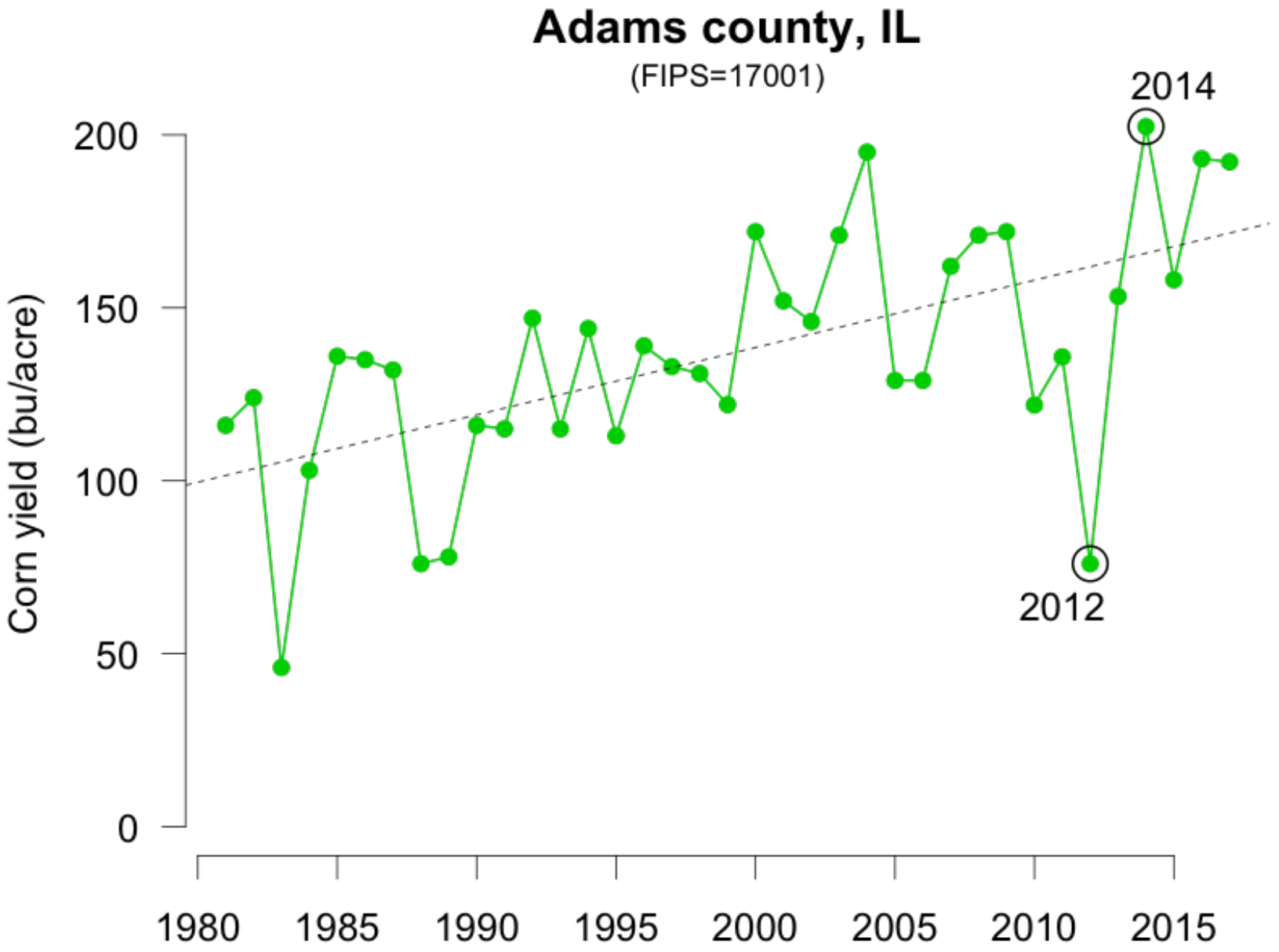

2014
Daily Tmax
40
30
20
10
0
-10
-20
Smoothed average (1981-2017)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of basis matrix columns
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
1 2 3 4 5 6 7
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of transformed Tmax
1400
1200
1000
800
600
400
200
0
1 2 3 4 5 6 7
Transformed variables


2014
Daily Tmax
40
30
20
10
0
-10
-20
Smoothed average (1981-2017)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of basis matrix columns
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
1 2 3 4 5 6 7 8
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of transformed Tmax
1400
1200
1000
800
600
400
200
0
1 2 3 4 5 6 7 8
Transformed variables


2014
Daily Tmax
40
30
20
10
0
-10
-20
Smoothed average (1981-2017)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of basis matrix columns
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
9 1 2 3 4 5 6 7 8
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of transformed Tmax
1400
1200
1000
800
600
400
200
0
1 2 3 4 5 6 7 8 9
Transformed variables


2014
Daily Tmax
40
30
20
10
0
-10
-20
Smoothed average (1981-2017)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of basis matrix columns
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
9 1 2 3 4 5 6 7 8 10
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of transformed Tmax
1400
1200
1000
800
600
400
200
0
1 2 3 4 5 6 7 8 9 10
Transformed variables


2012
Daily Tmax
40
30
20
10
0
-10
-20
Smoothed average (1981-2017)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of basis matrix columns
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
9 1 2 3 4 5 6 7 8 10
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Values of transformed Tmax
1400
1200
1000
800
600
400
200
0
1 2 3 4 5 6 7 8 9 10
Transformed variables

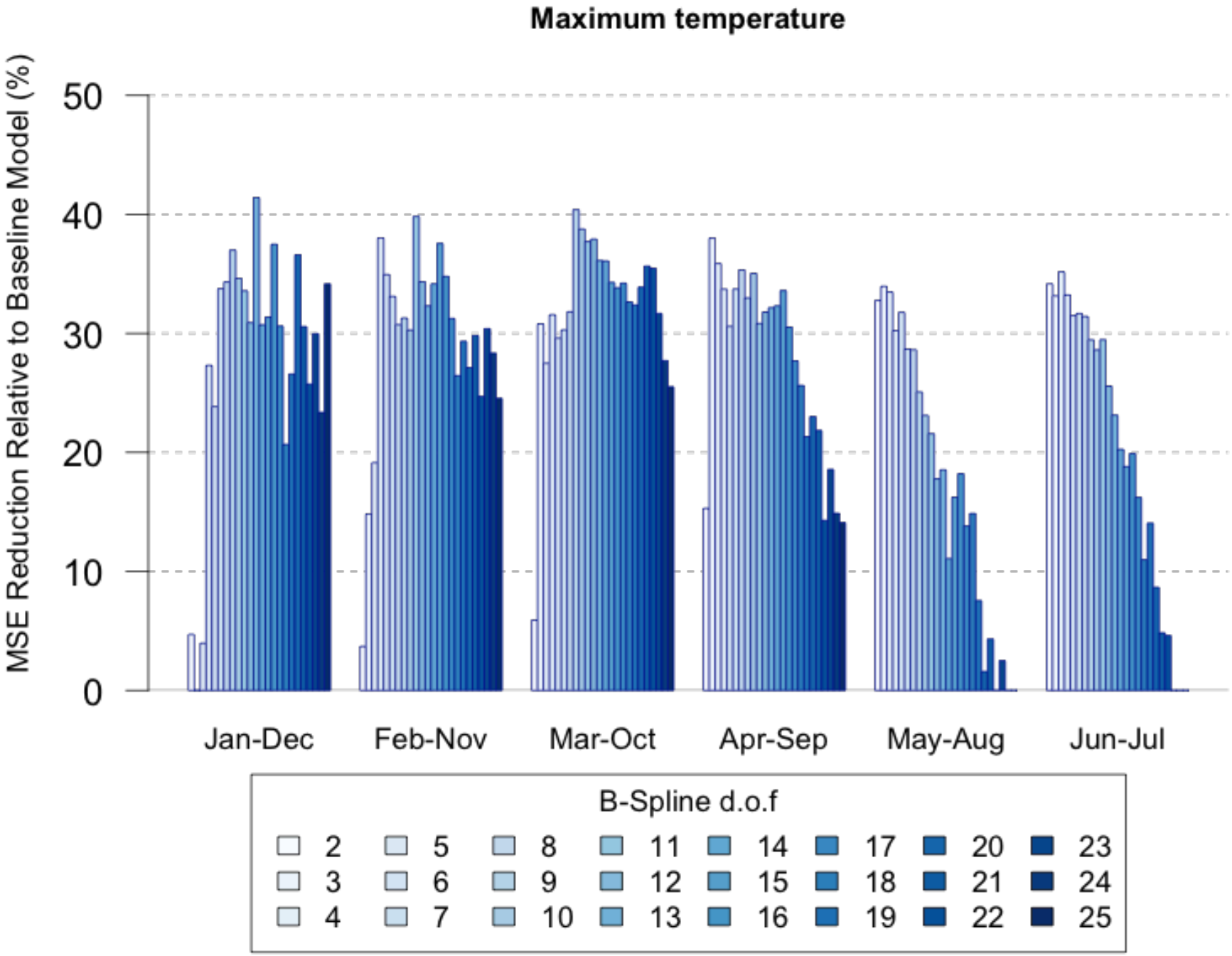

Adams county, IL
(FIPS=17001)
2014
2012
Corn yield (bu/acre)
200
150
100
50
0
1980
1985
1990
1995
2000
2005
2010
2015

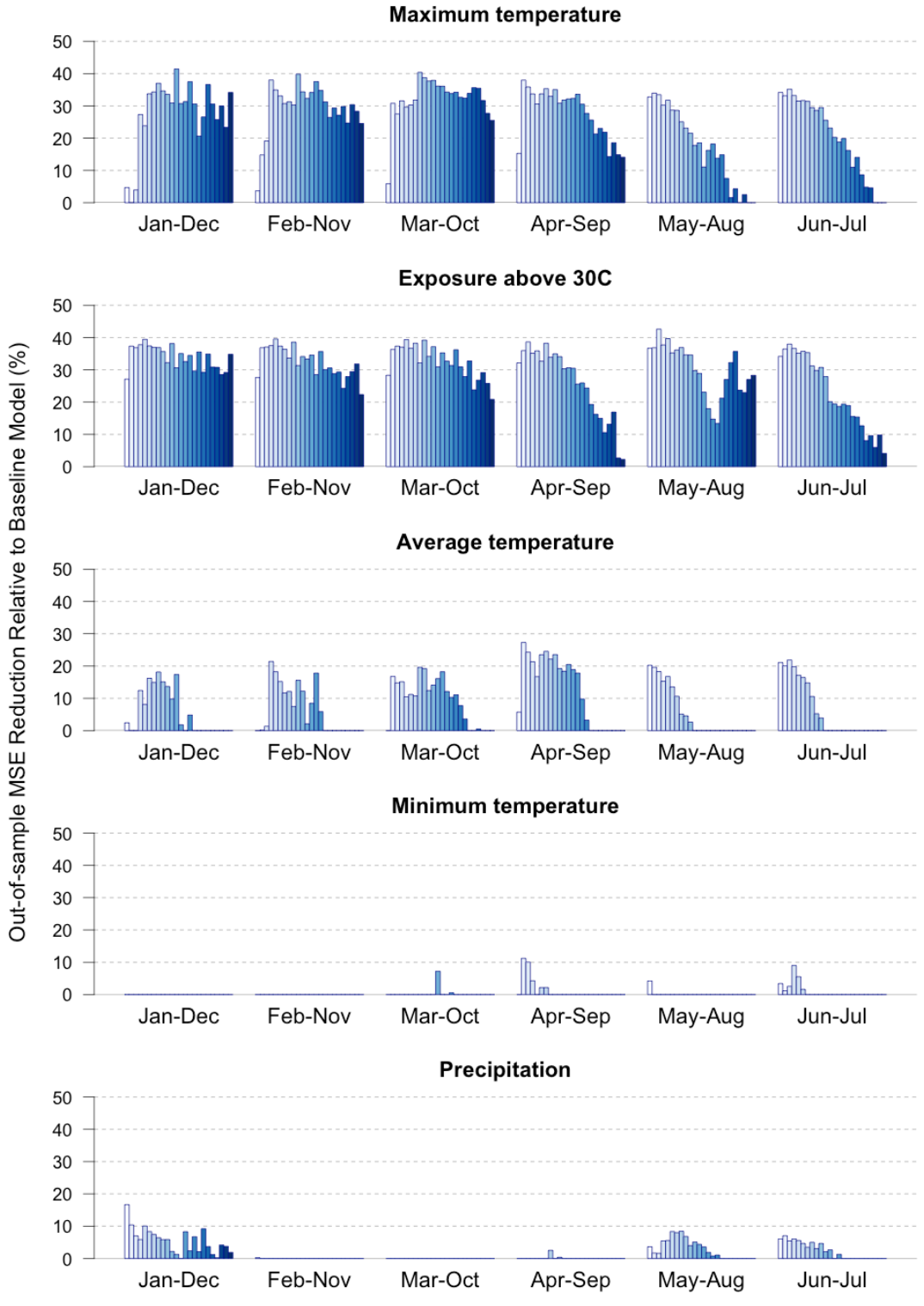
Cross-validation
(parameter search)

Cross-validation

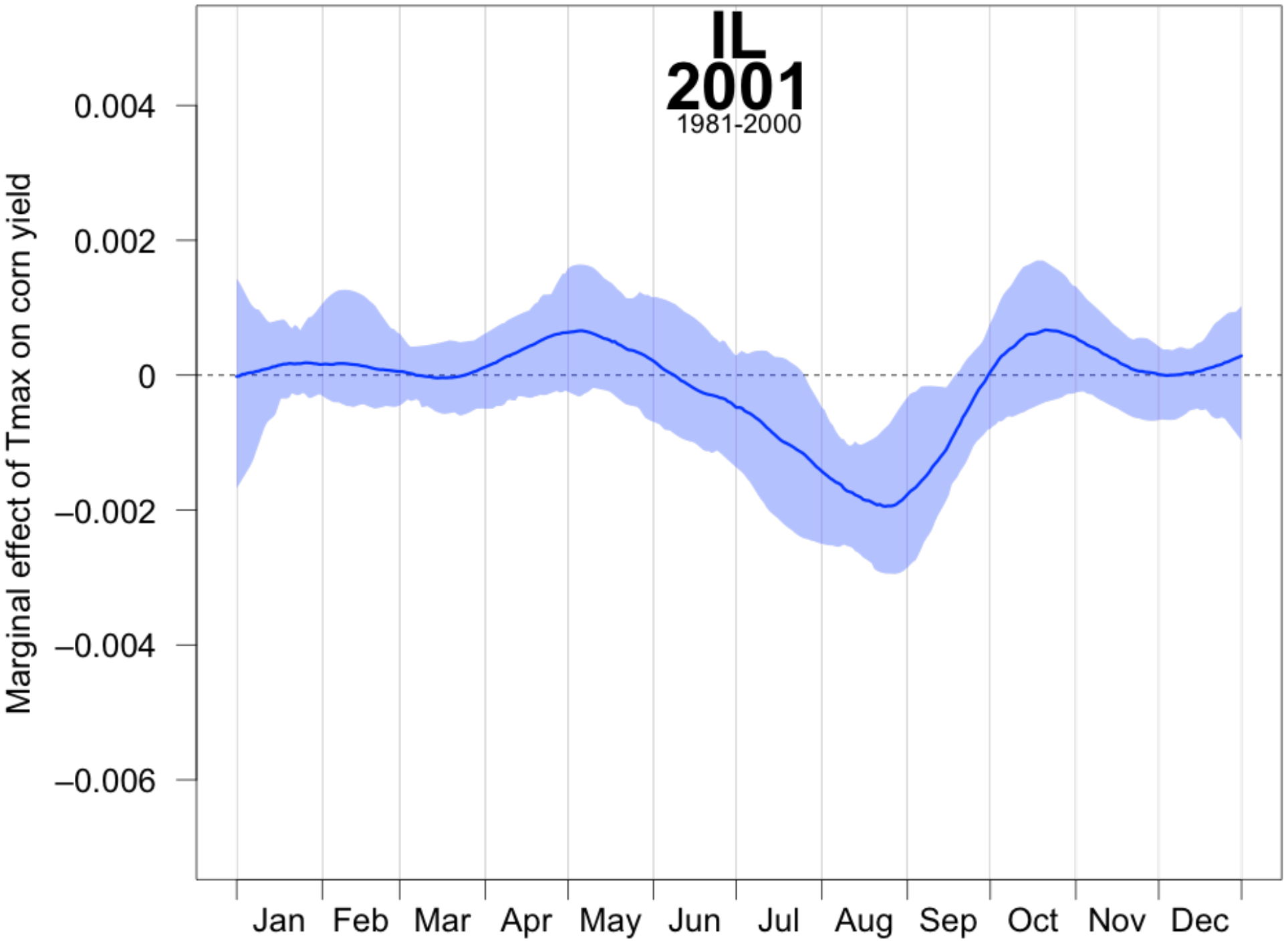
- Data: corn yields, 1981-2017, Illinois
- Cross-validation: 10 year folds
- Dimensions we explore:
  - Growing season period (e.g. Jan-Dec, Apr-Sep, etc.)
  - Spline with more/less flexibility
  - Different climate variables (Tmax, Precip, etc.)

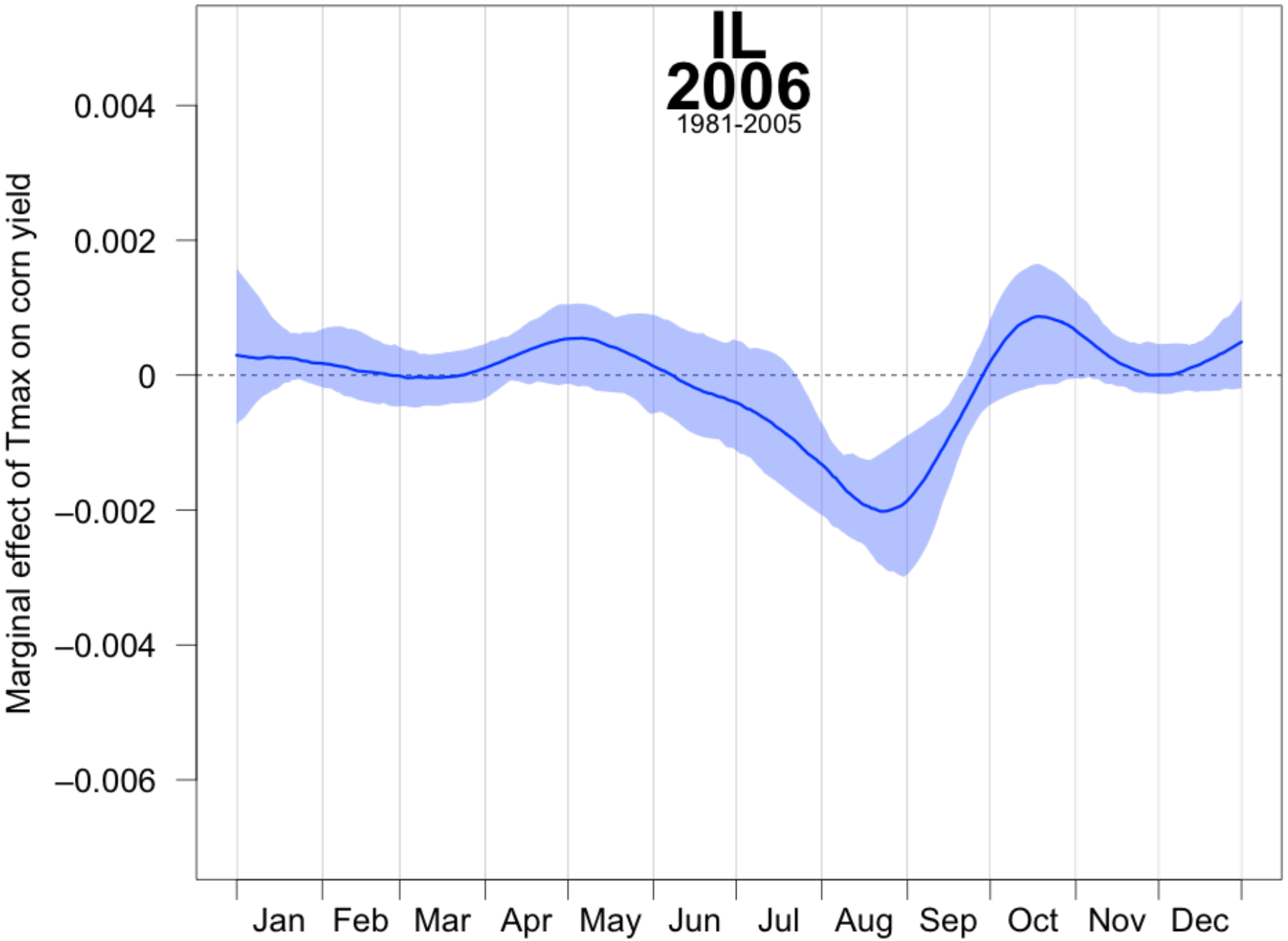
IL

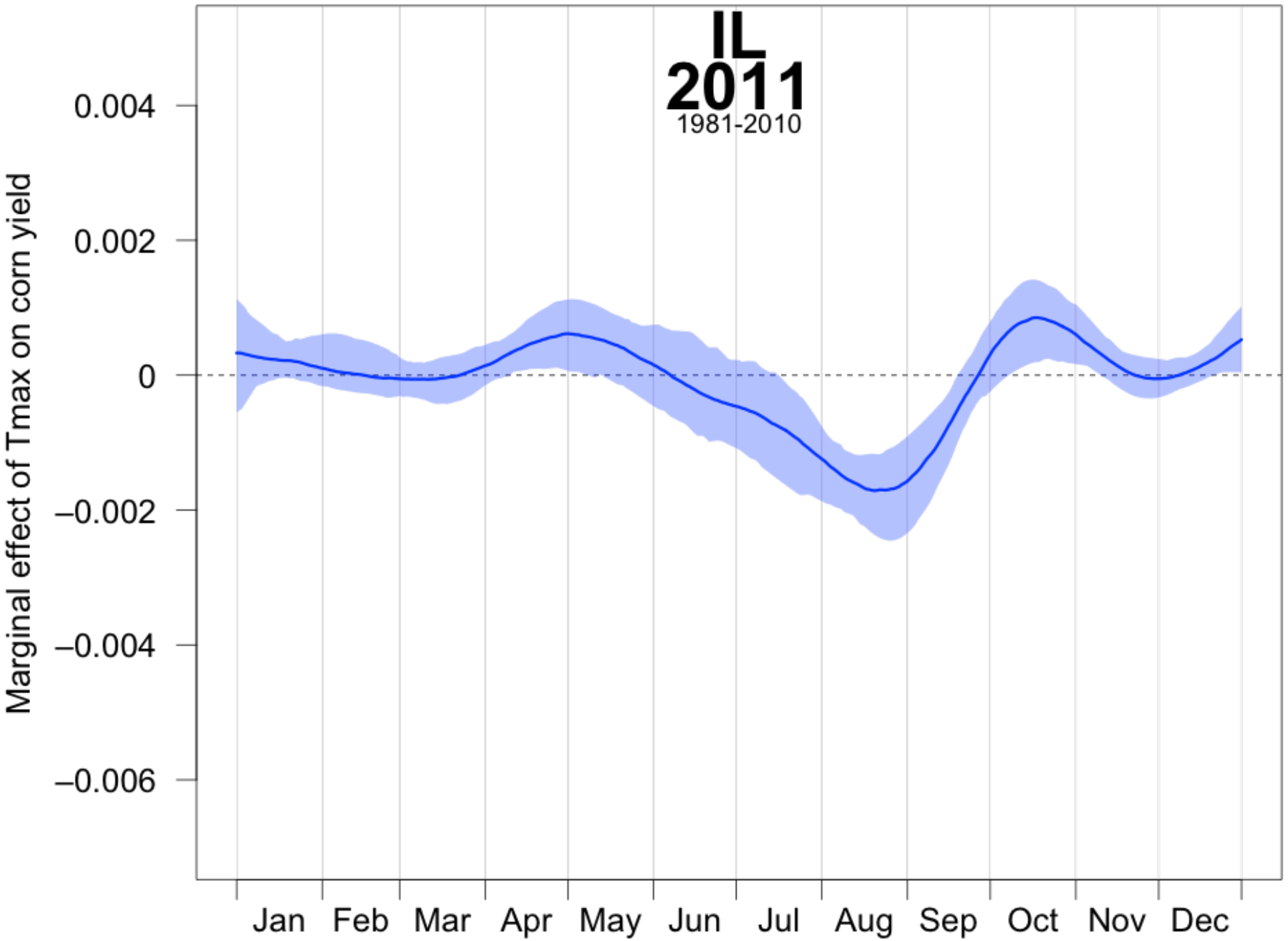

Maximum temperature
MSE Reduction Relative to Baseline Model (%)
50
40
30
20
10
0
Jan-Dec
Feb-Nov
Mar-Oct
Apr-Sep
May-Aug
Jun-Jul
B-Spline d.o.f
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

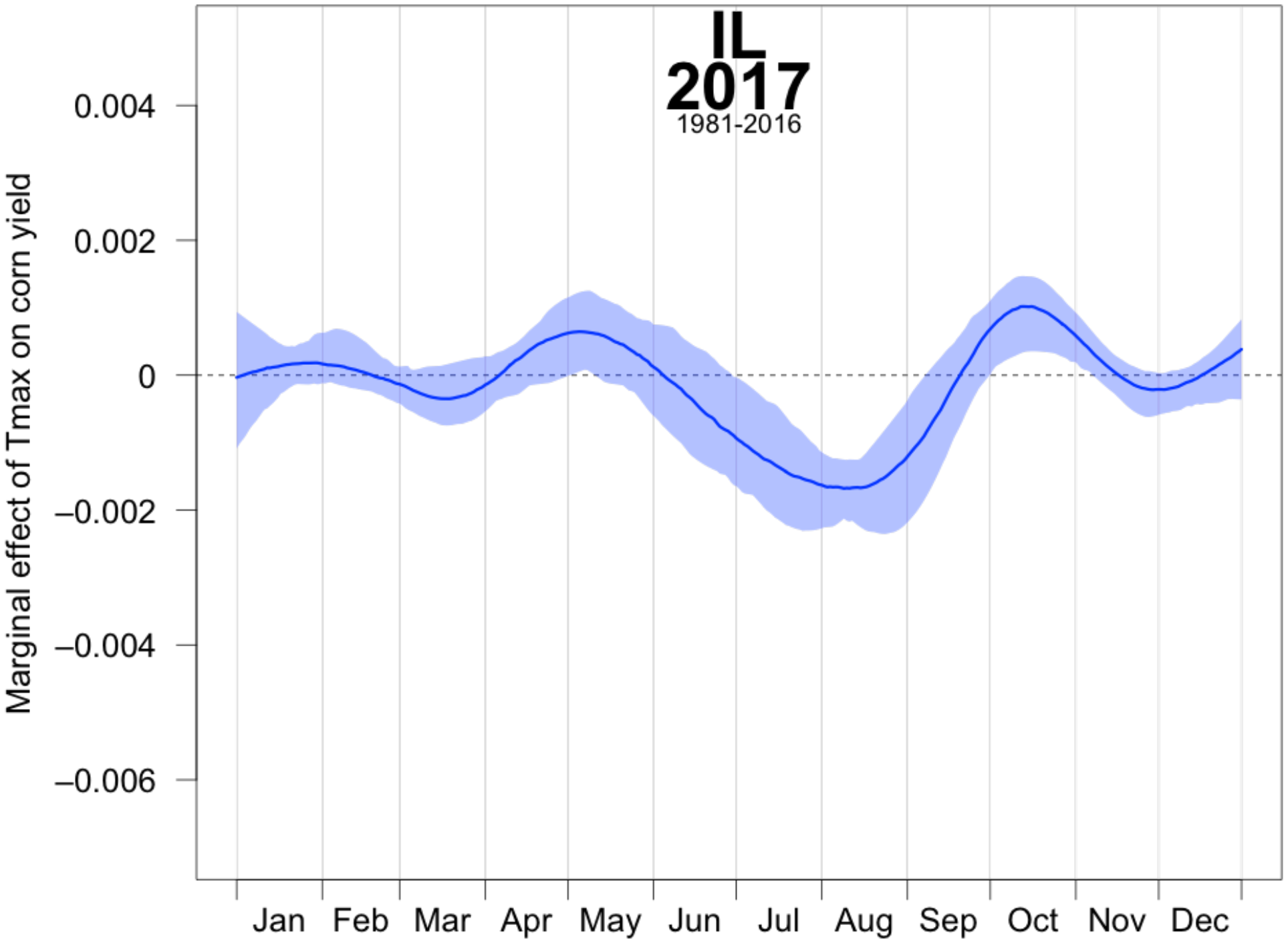

Maximum temperature
Exposure above 30C
Average temperature
Minimum temperature
Precipitation
Out-of-sample MSE Reduction Relative to Baseline Model (%)
0
10
20
30
40
50
Jan-Dec
Feb-Nov
Mar-Oct
Apr-Sep
May-Aug
Jun-Jul

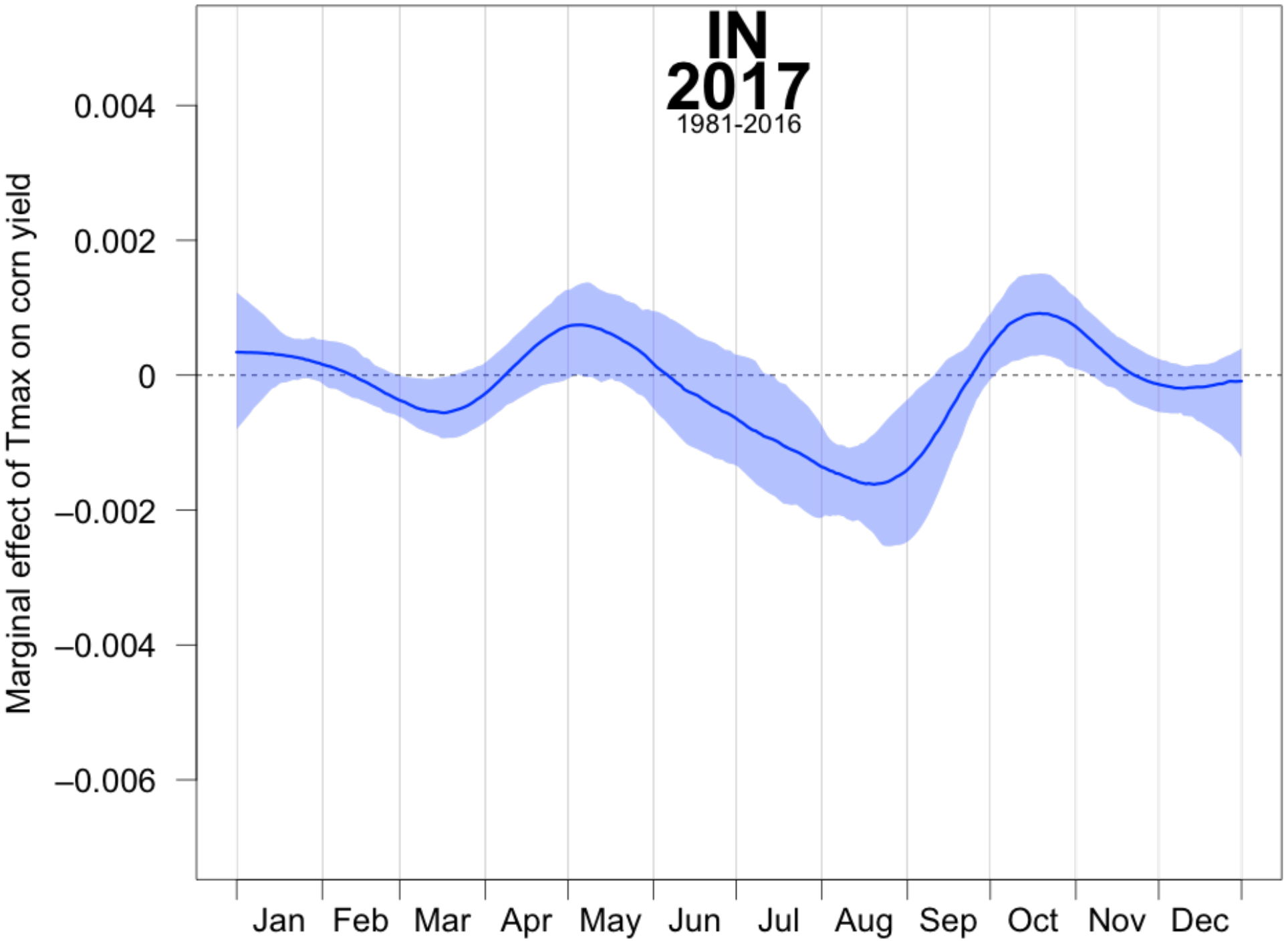
Marginal effects

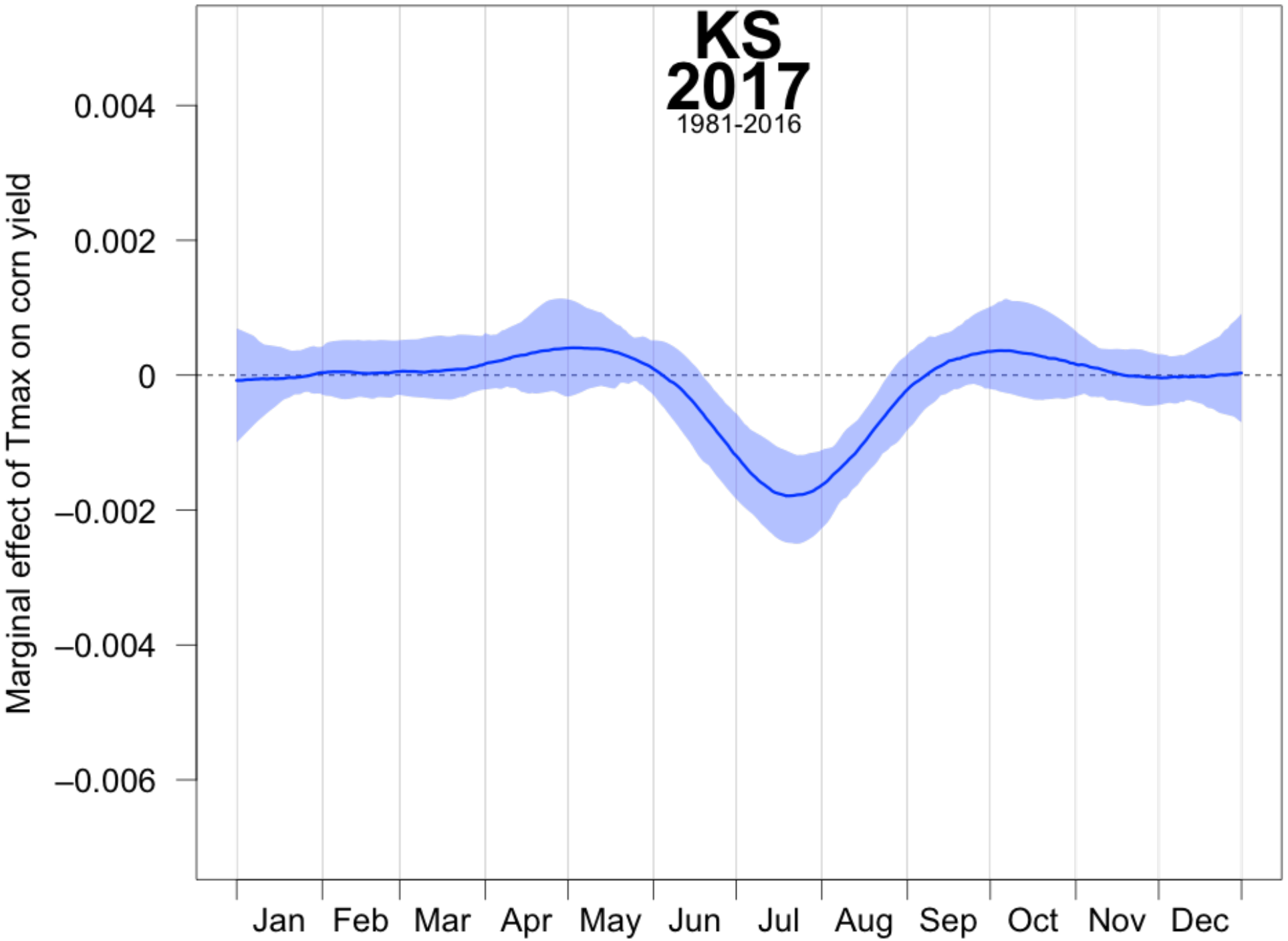

IL
2001
1981-2000
0.004
0.002
0
−0.002
−0.004
−0.006
Marginal effect of Tmax on corn yield
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

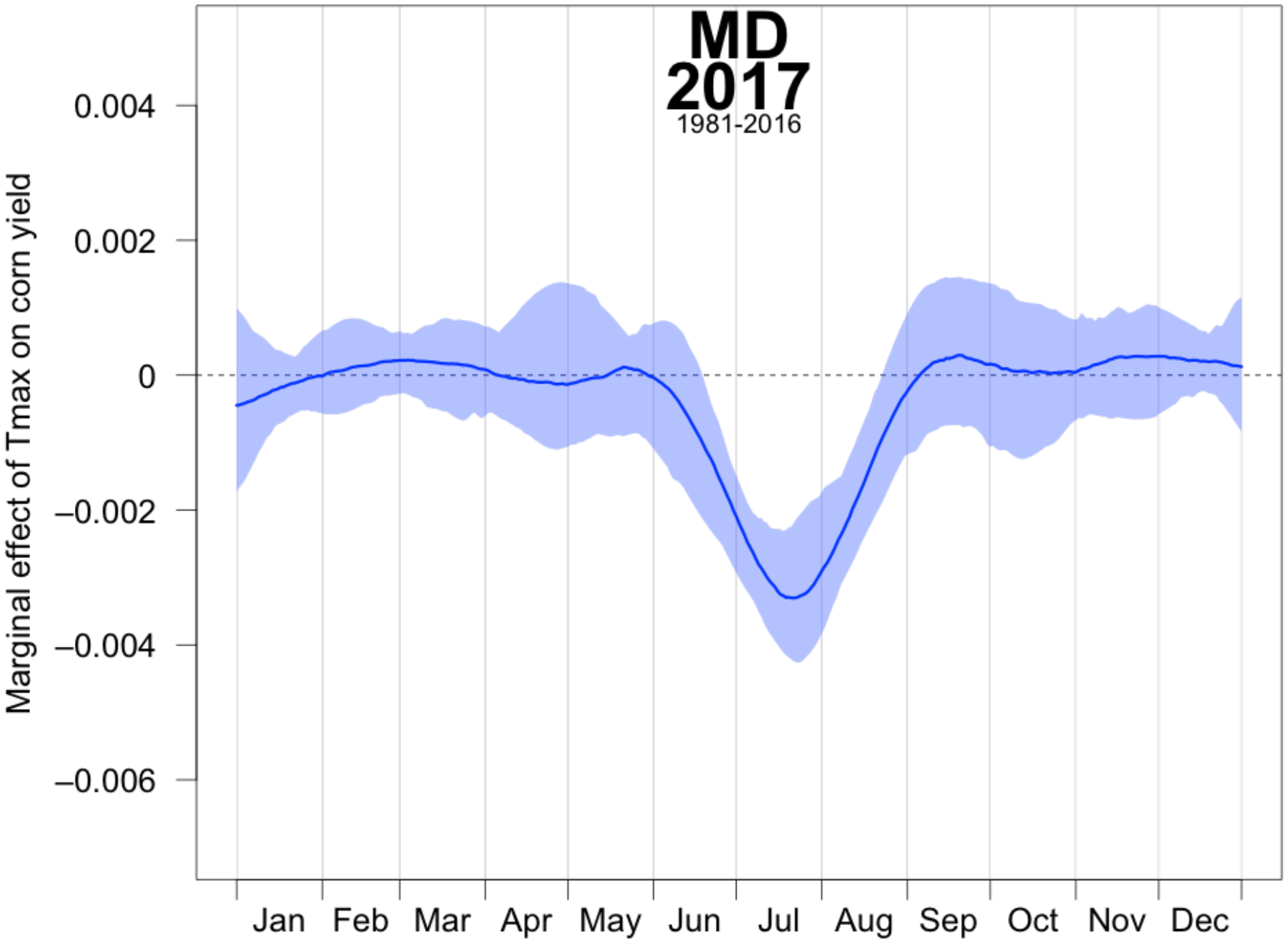

IL
2006
1981-2005
Marginal effect of Tmax on corn yield
0.004
0.002
0
–0.002
–0.004
–0.006
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


IL
2011
1981-2010
0.004
0.002
0
−0.002
−0.004
−0.006
Marginal effect of Tmax on corn yield
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

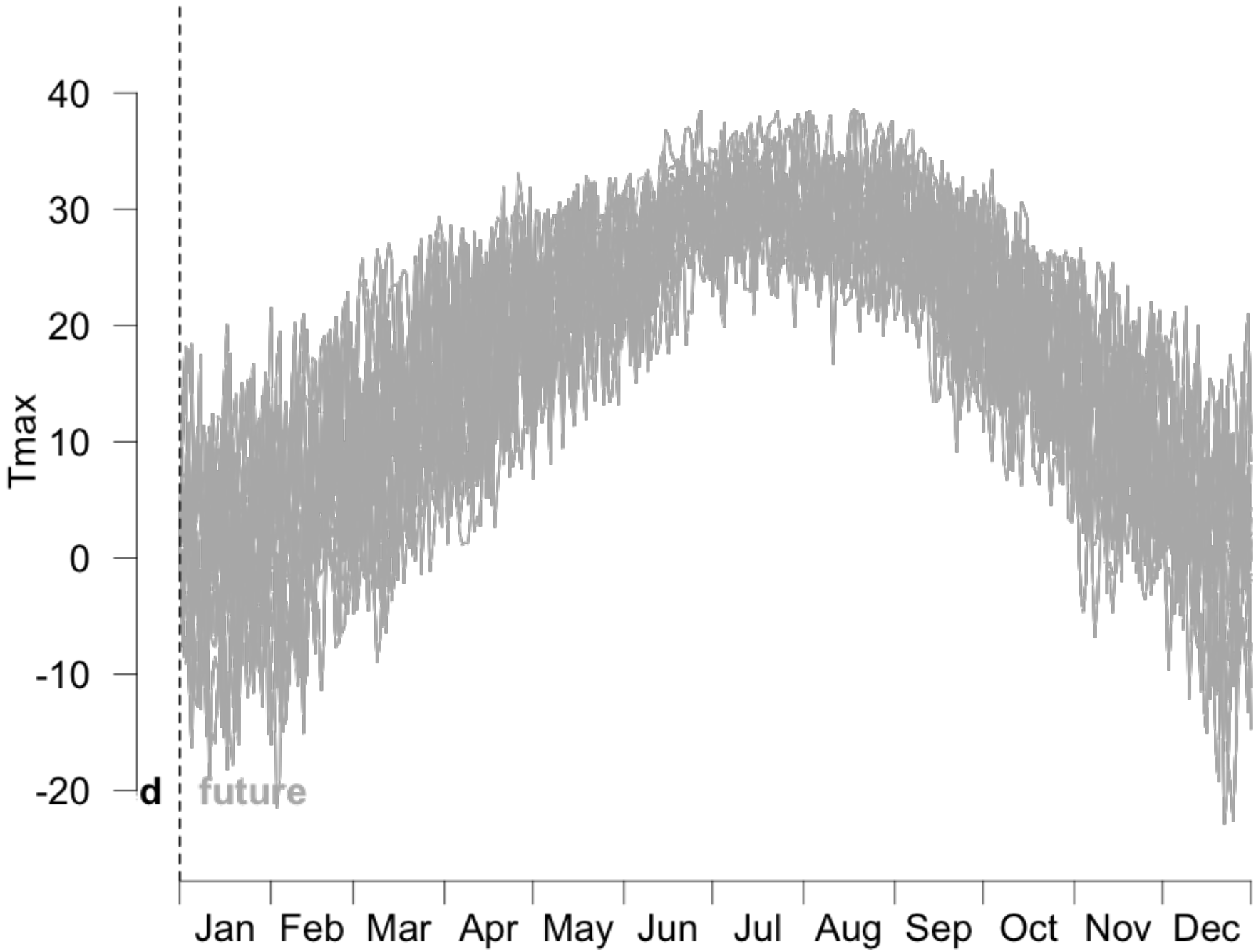

IL
2017
1981-2016
Marginal effect of Tmax on corn yield
0.004
0.002
0
–0.002
–0.004
–0.006
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

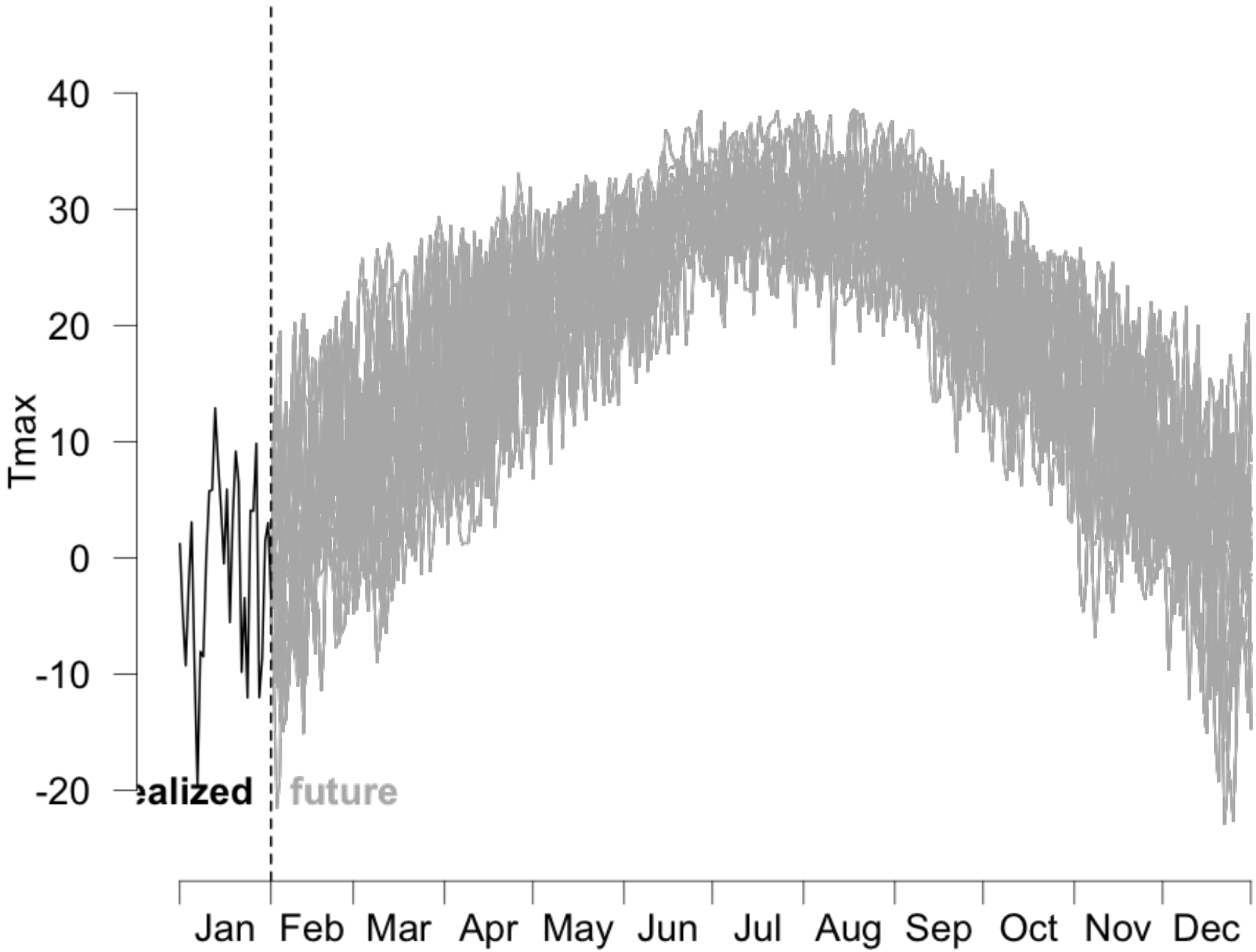

IN
2017
1981-2016
Marginal effect of Tmax on corn yield
0.004
0.002
0
−0.002
−0.004
−0.006
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

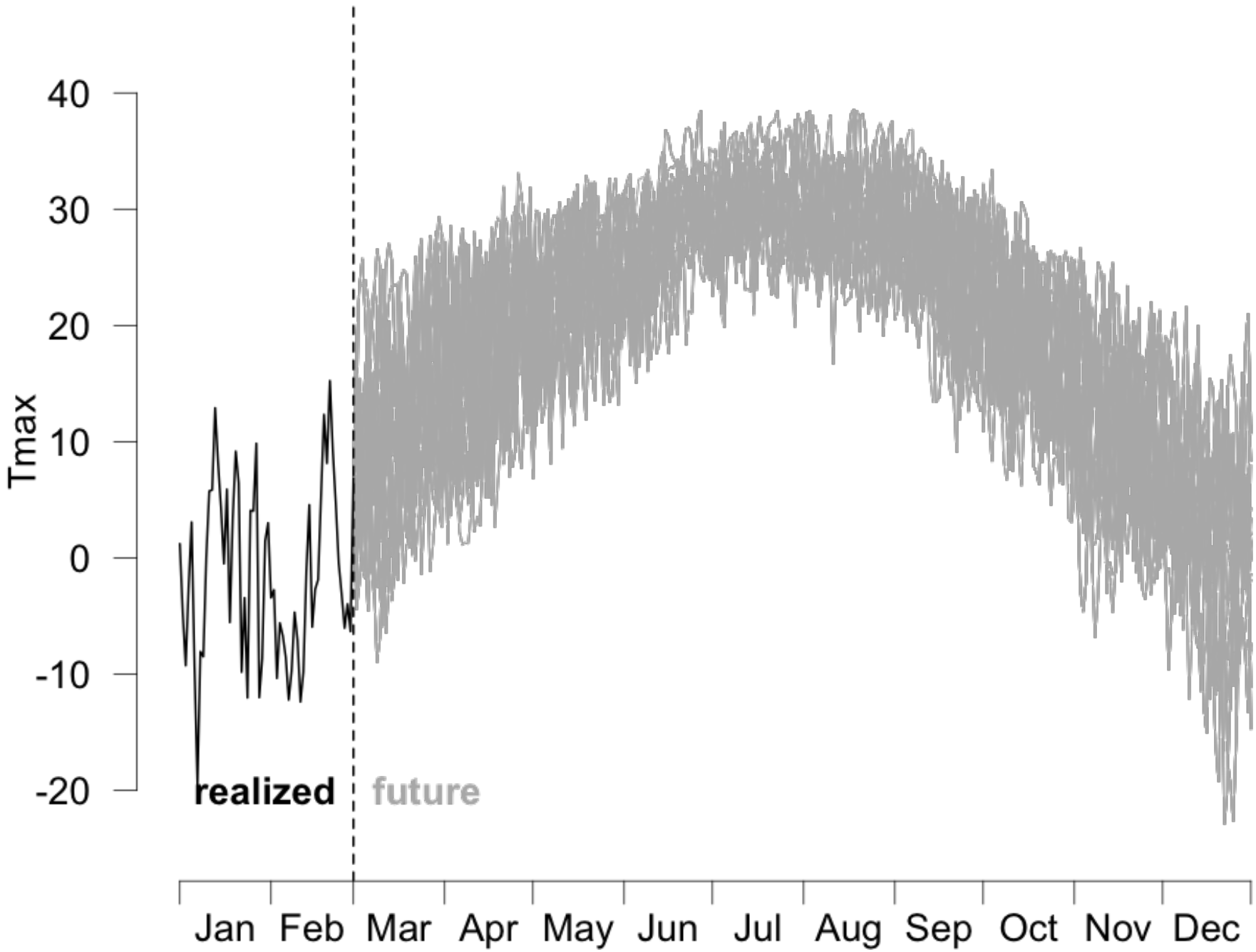

KS
2017
1981-2016
Marginal effect of Tmax on corn yield
0.004
0.002
0
–0.002
–0.004
–0.006
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

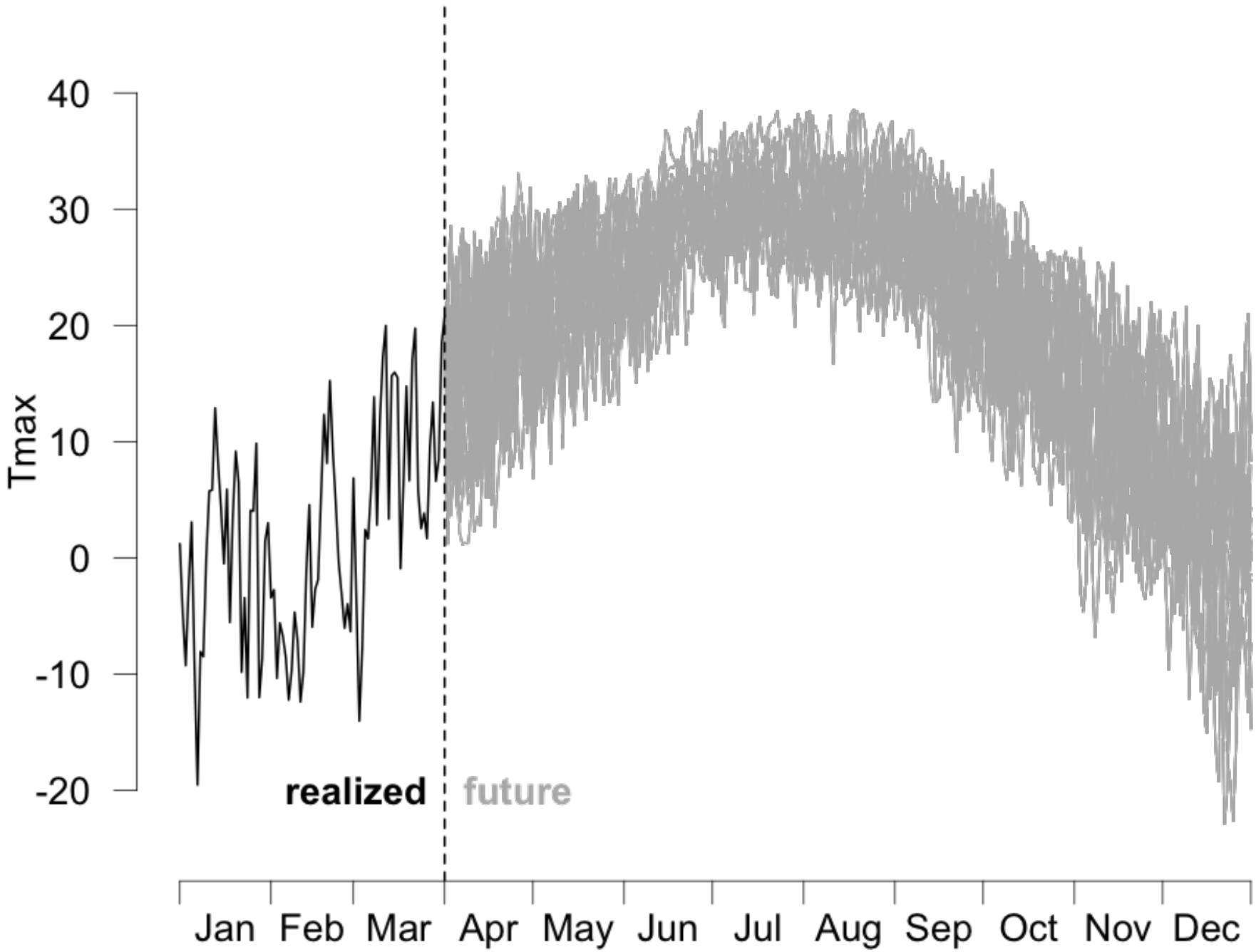

MD
2017
1981-2016
Marginal effect of Tmax on corn yield
0.004
0.002
0
−0.002
−0.004
−0.006
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

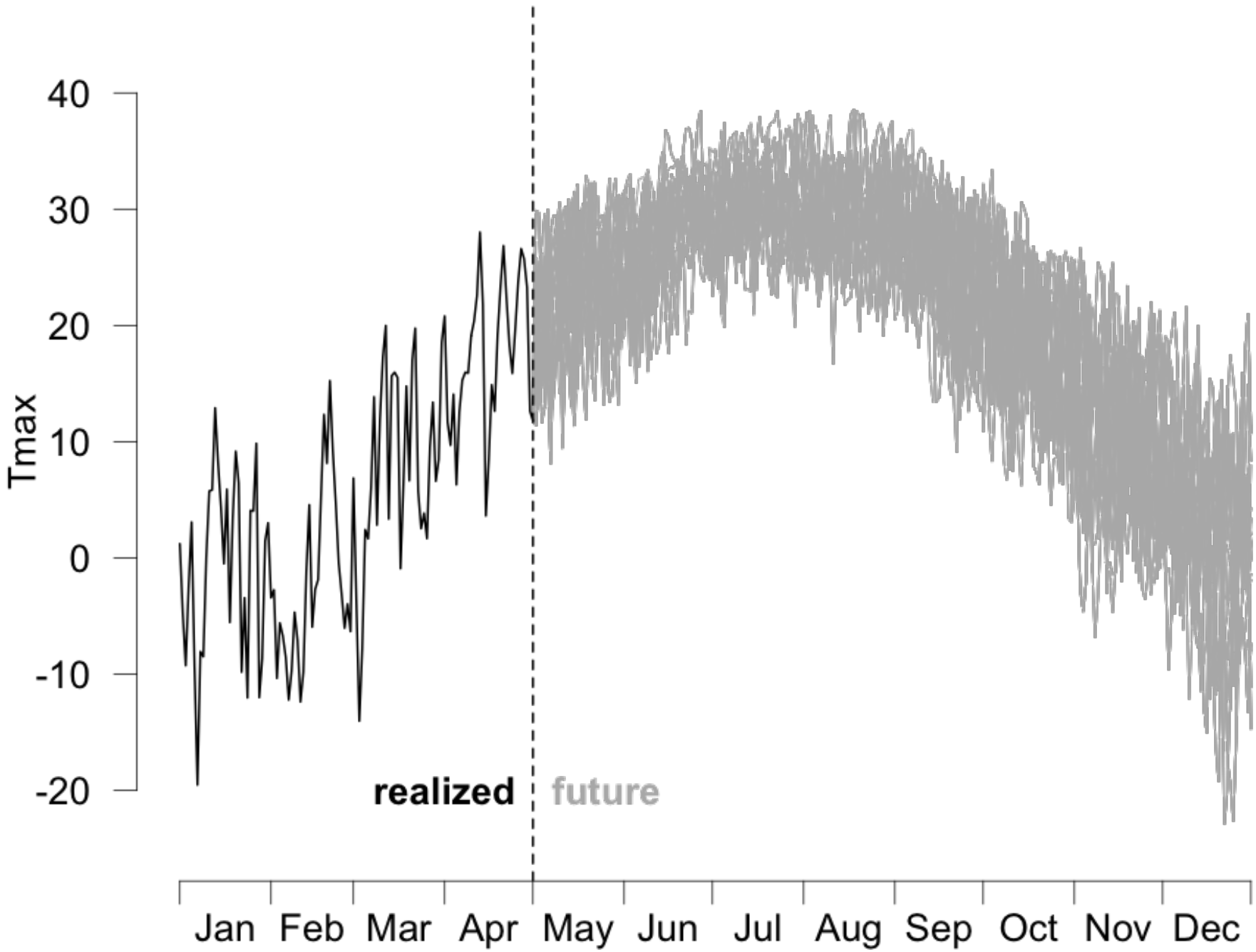
Imputing weather for forecasting

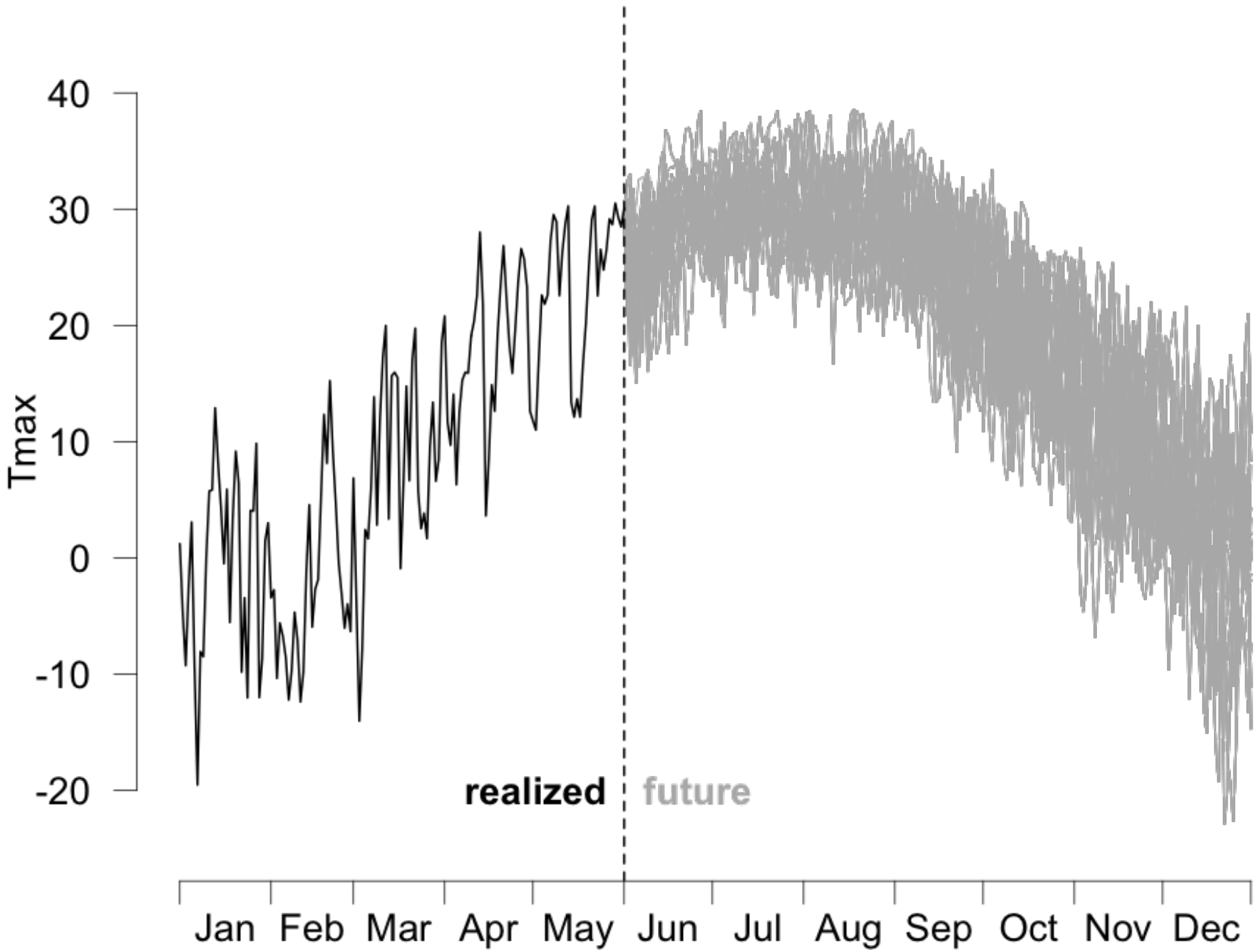

40
30
20
10
0
-10
-20
Tmax
d
future
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

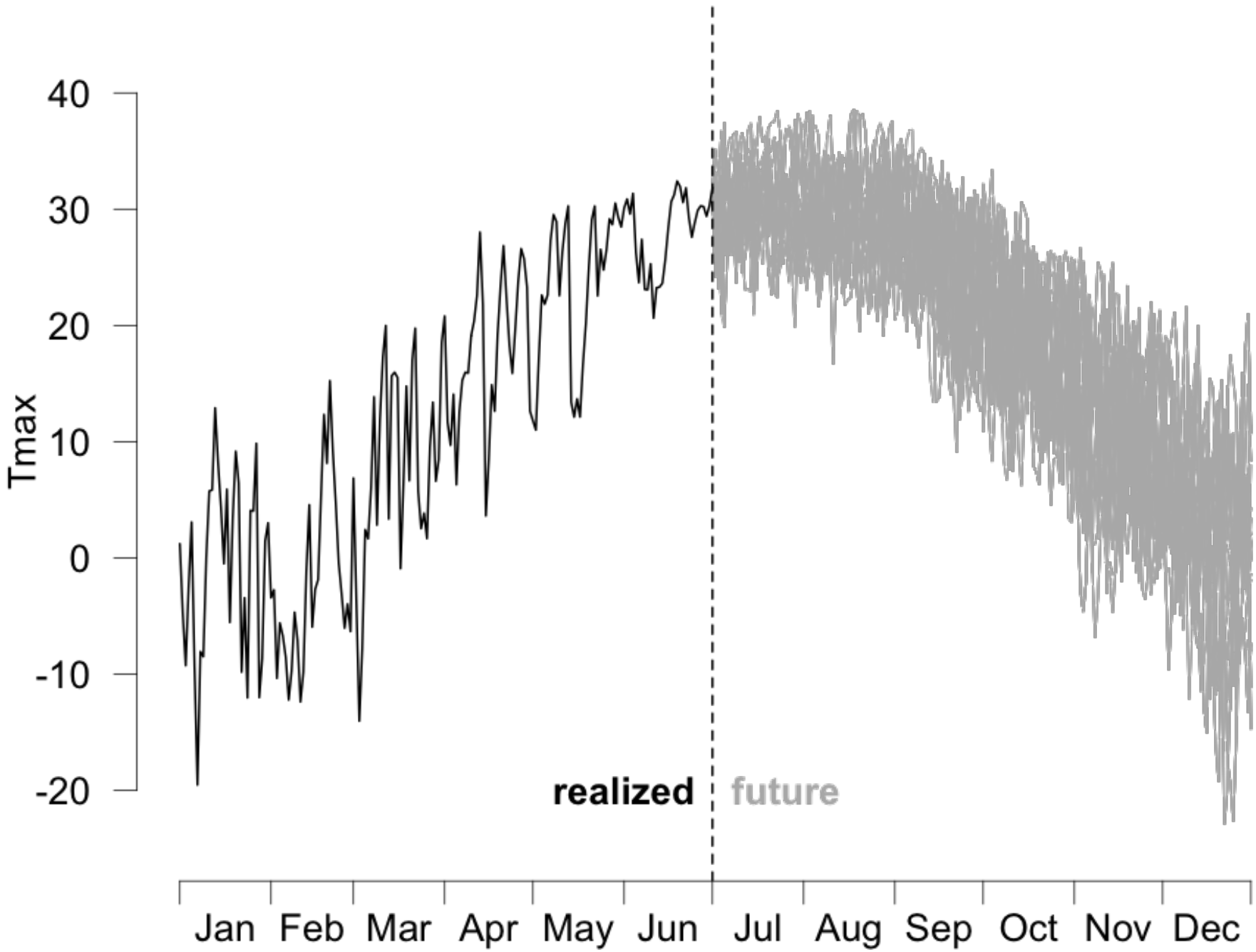

40
30
20
10
0
-10
-20
Tmax
ealized
future
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

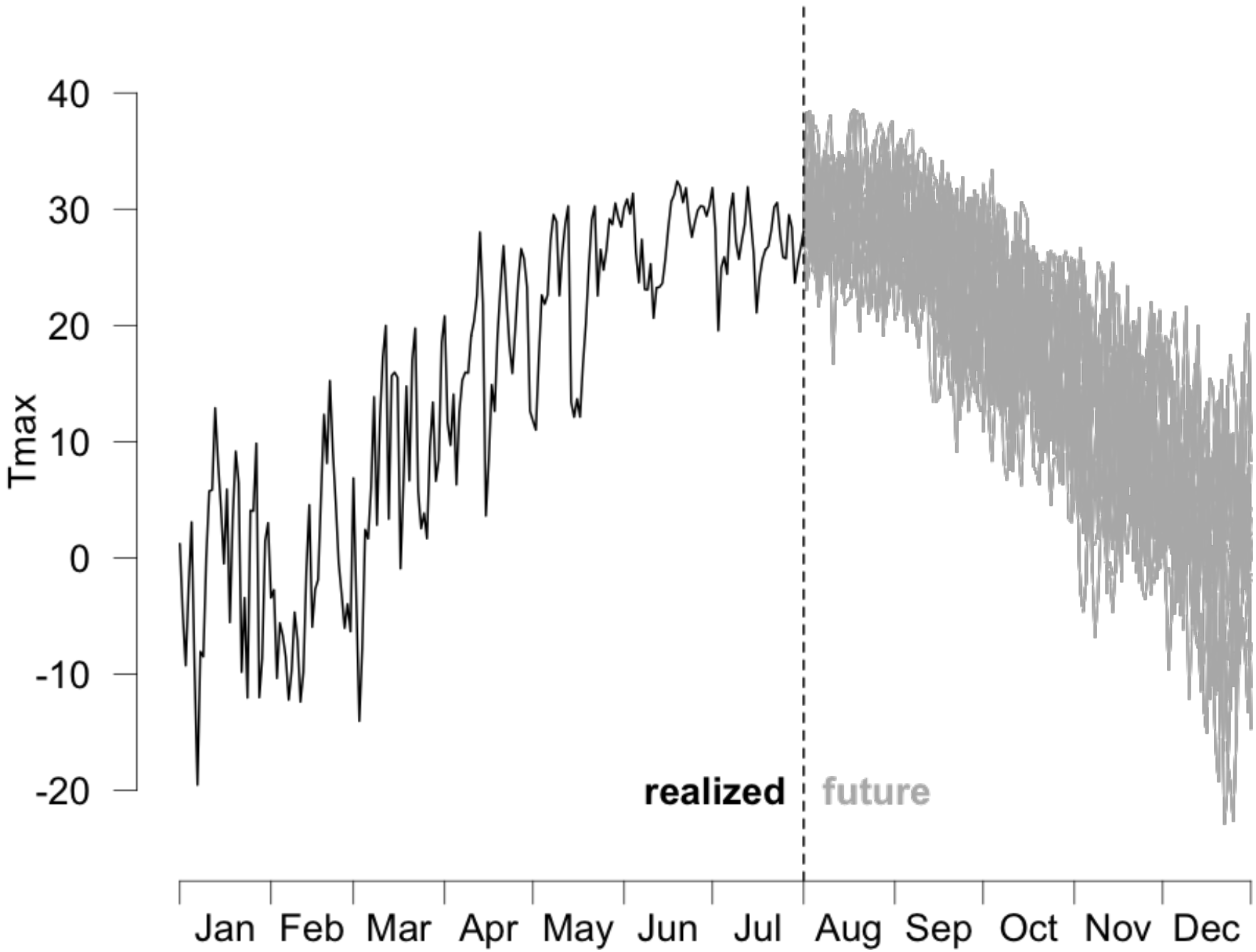

40
30
20
10
0
-10
-20
Tmax
realized
future
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

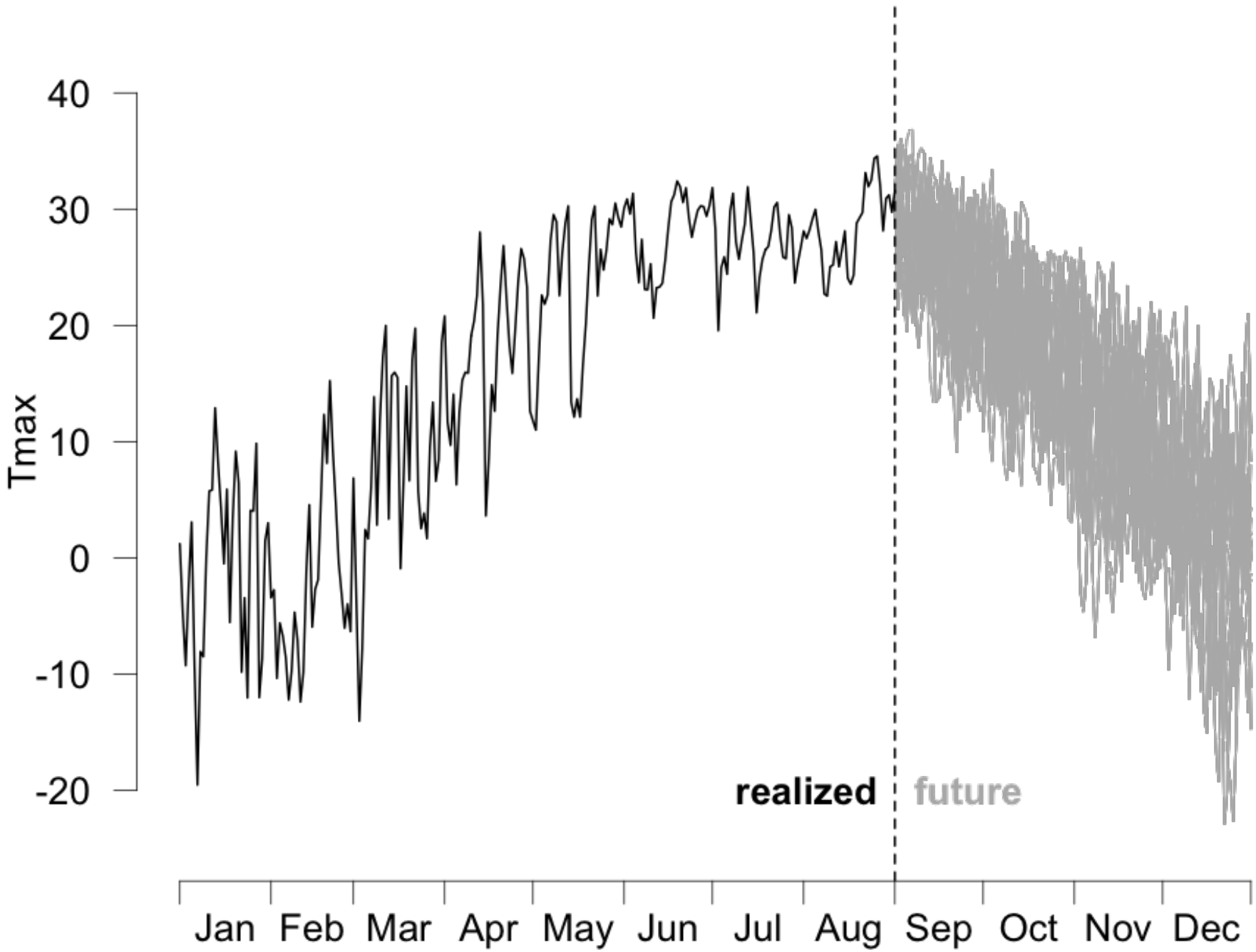

40
30
20
10
0
-10
-20
Tmax
realized
future
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

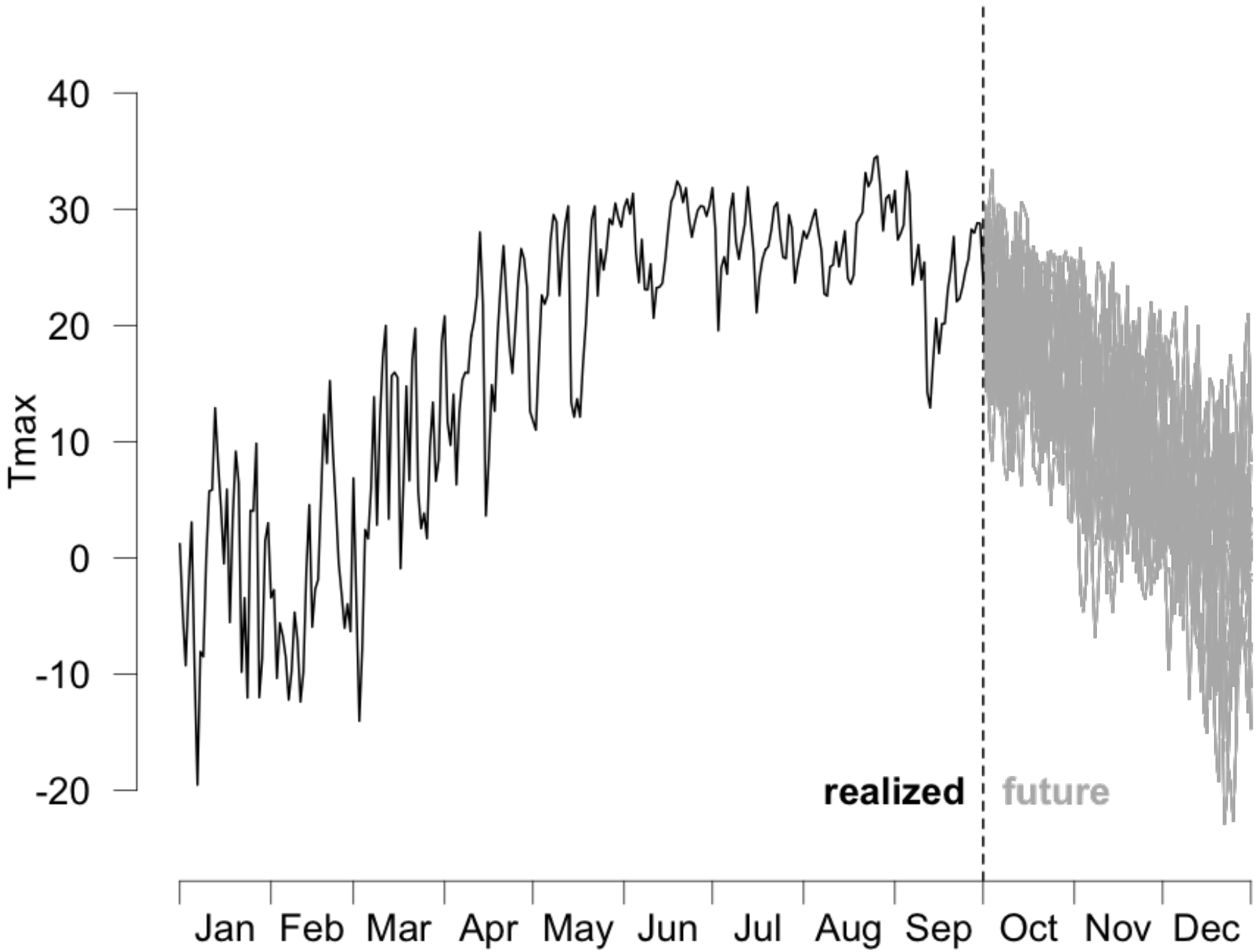

40
30
20
10
0
-10
-20
Tmax
realized
future
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

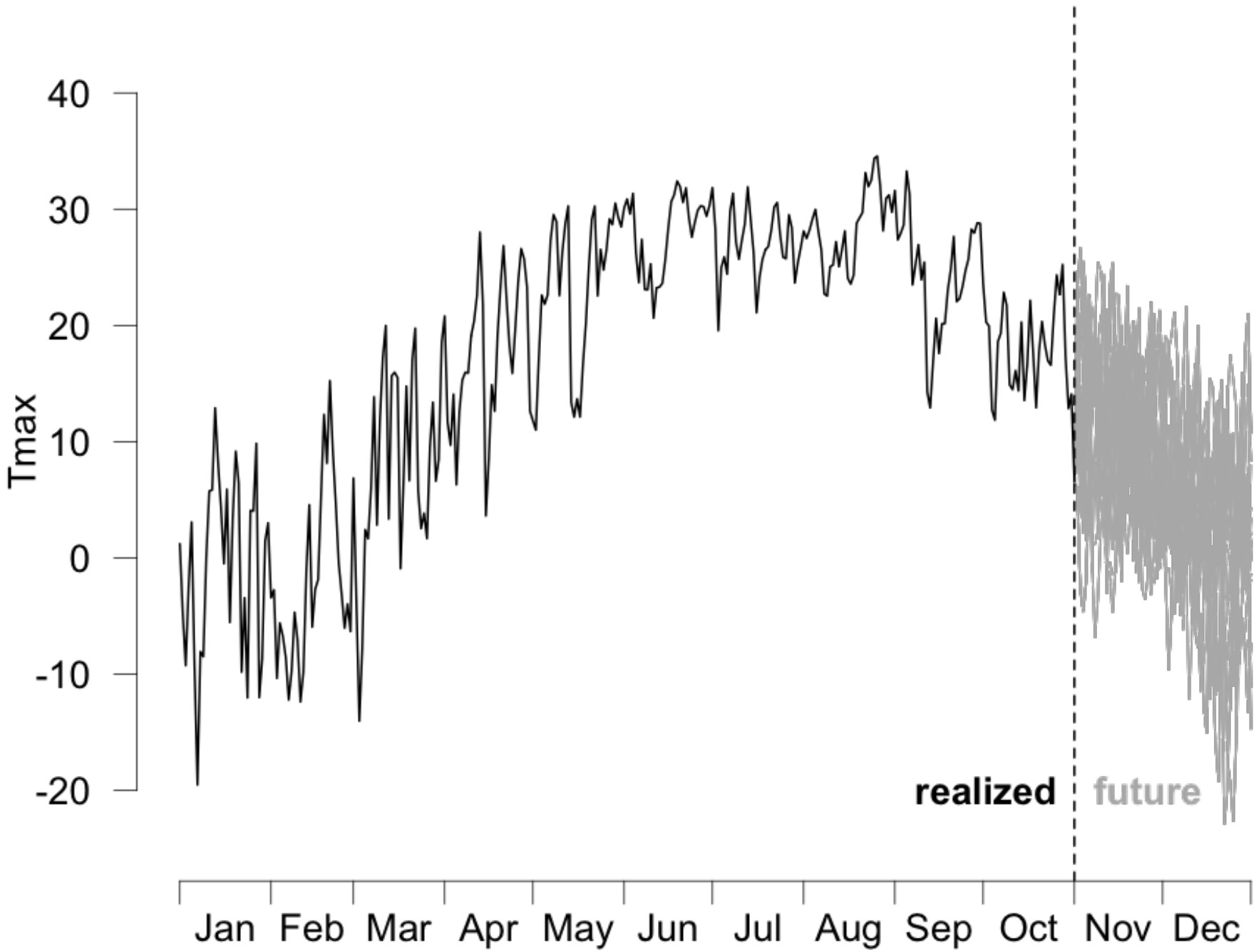

40
30
20
10
0
-10
-20
Tmax
realized
future
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

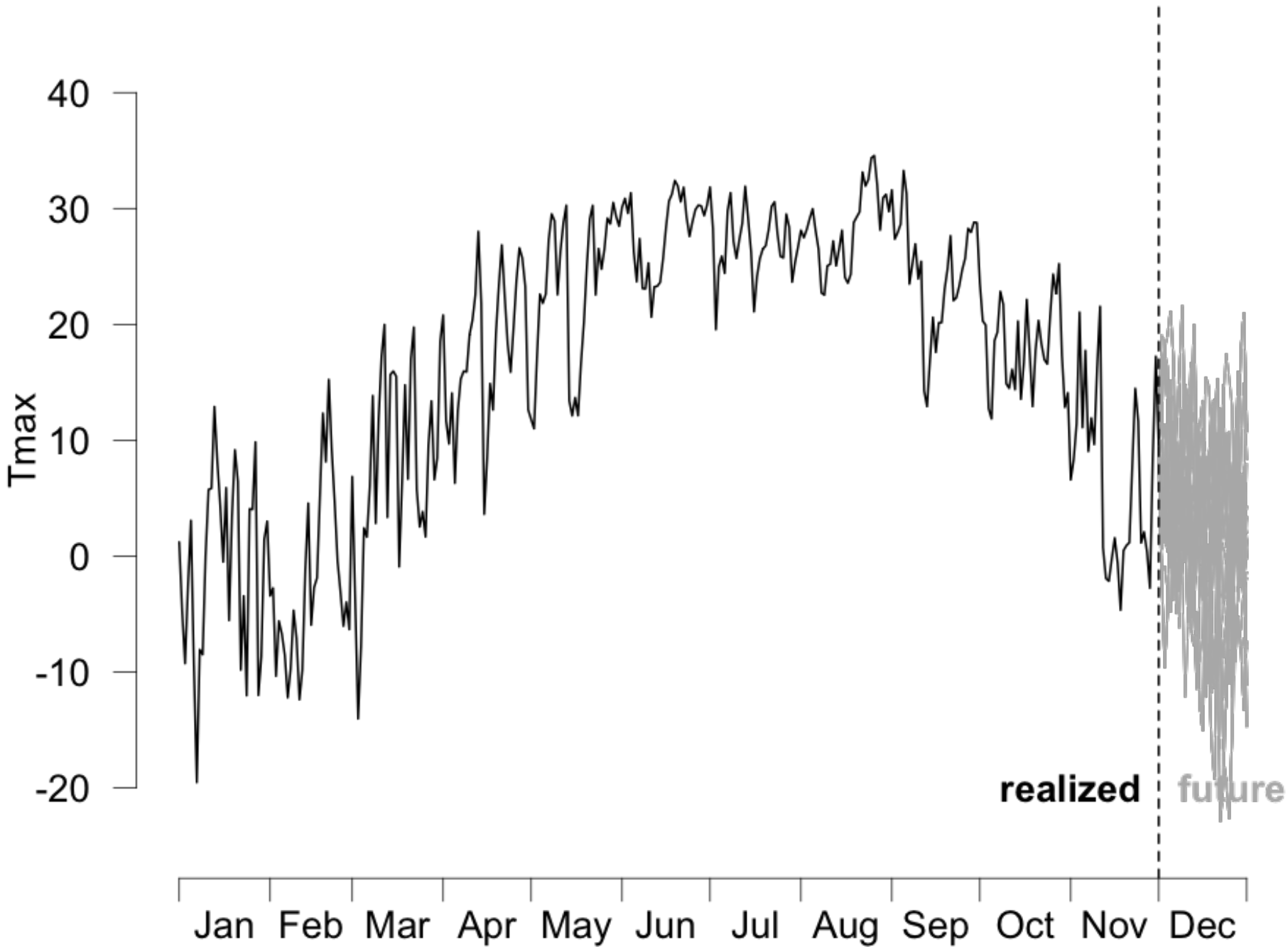

40
30
20
10
0
-10
-20
Tmax
realized
future
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

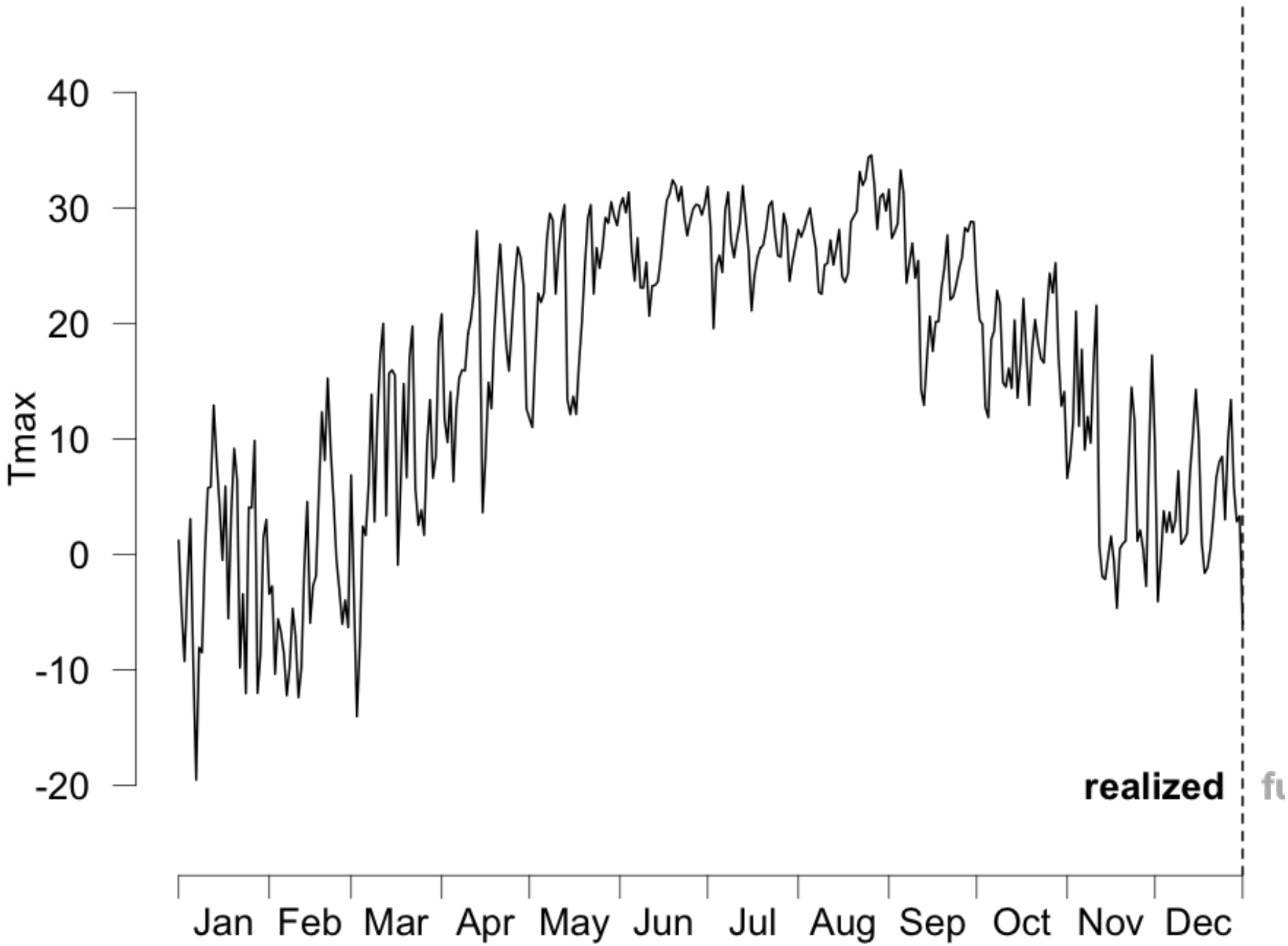

40
30
20
10
0
-10
-20
Tmax
realized
future
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


40
30
20
10
0
-10
-20
Tmax
realized
future
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

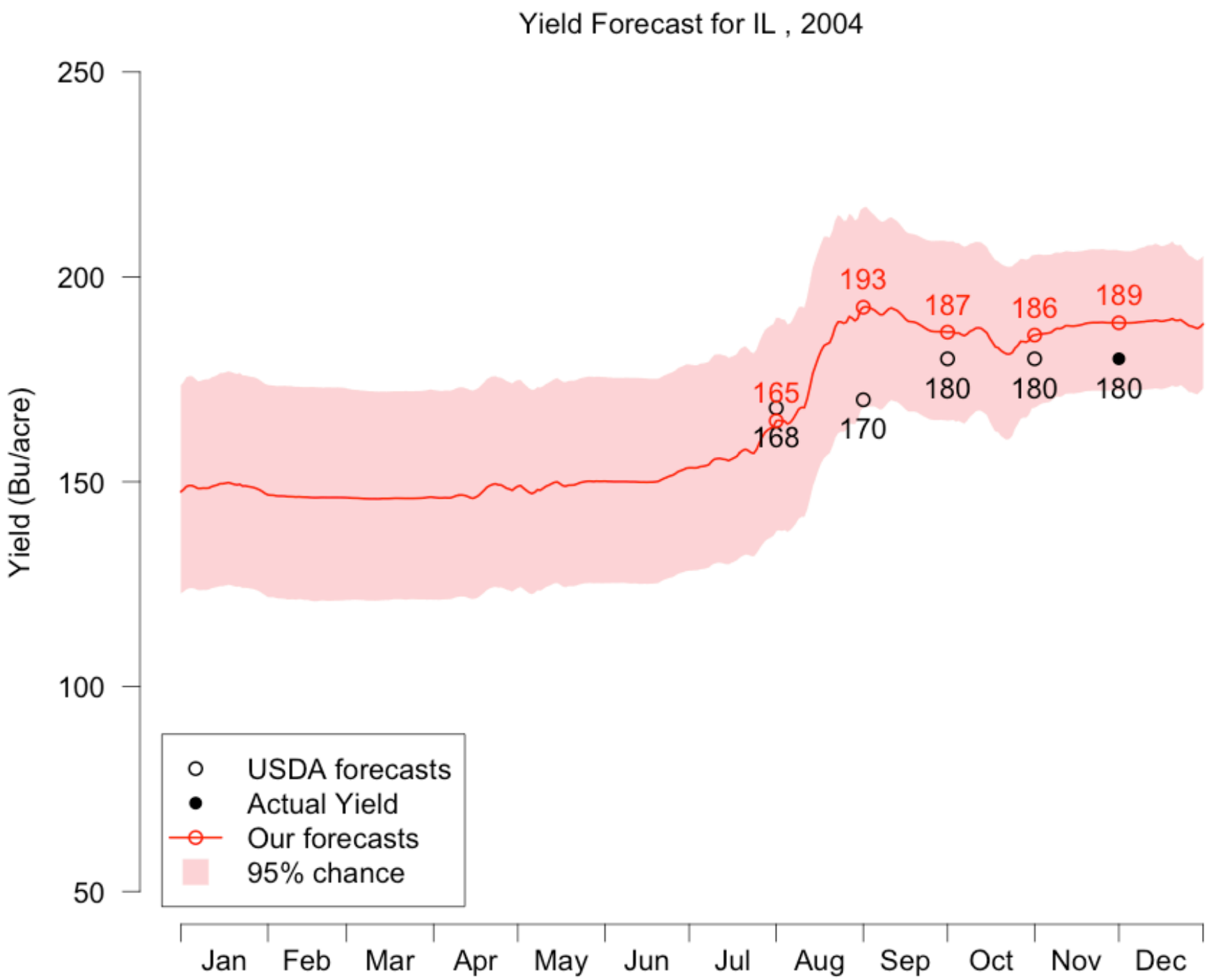

40
30
20
10
0
-10
-20
Tmax
realized
future
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

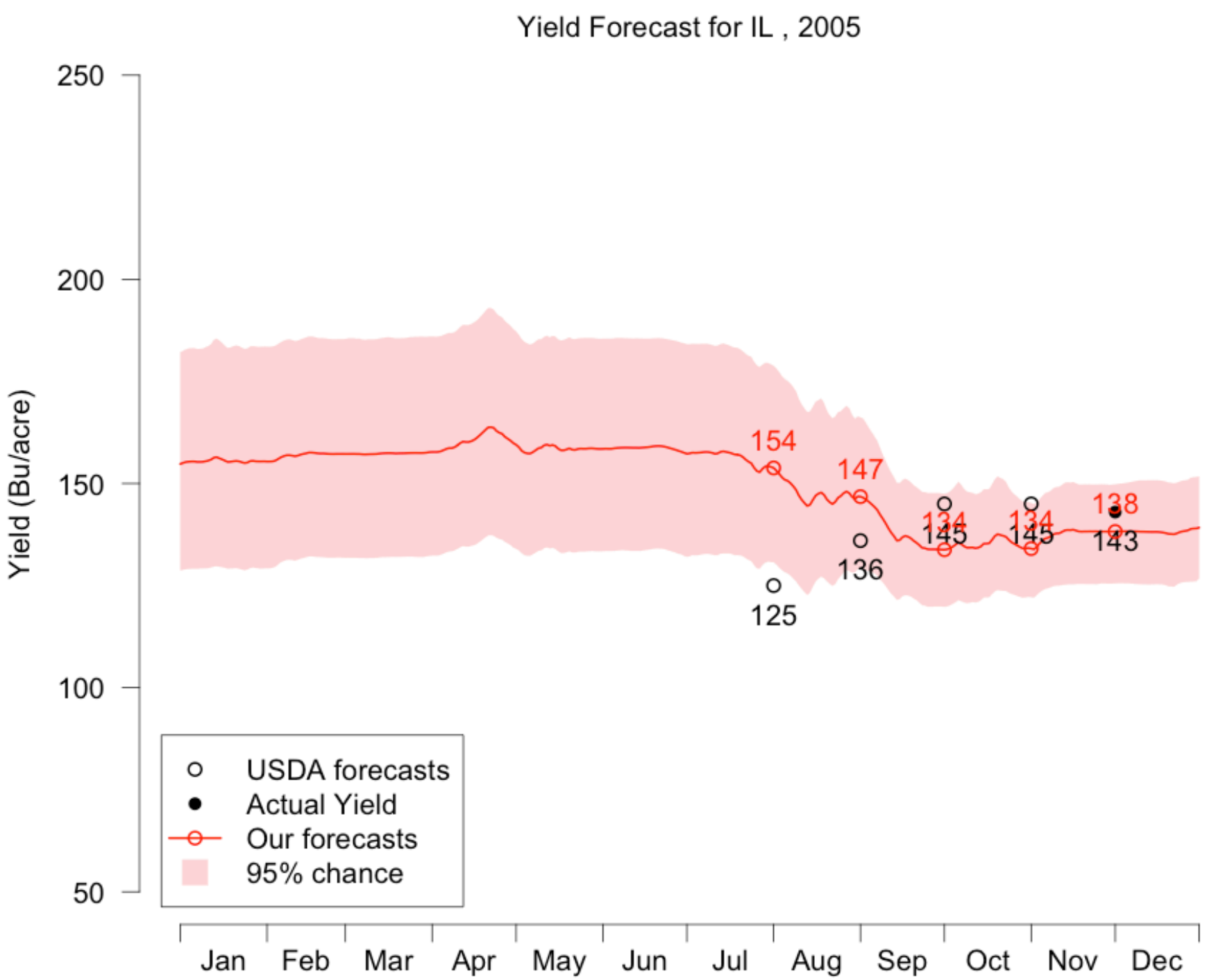

40
30
20
10
0
-10
-20
Tmax
realized
future
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

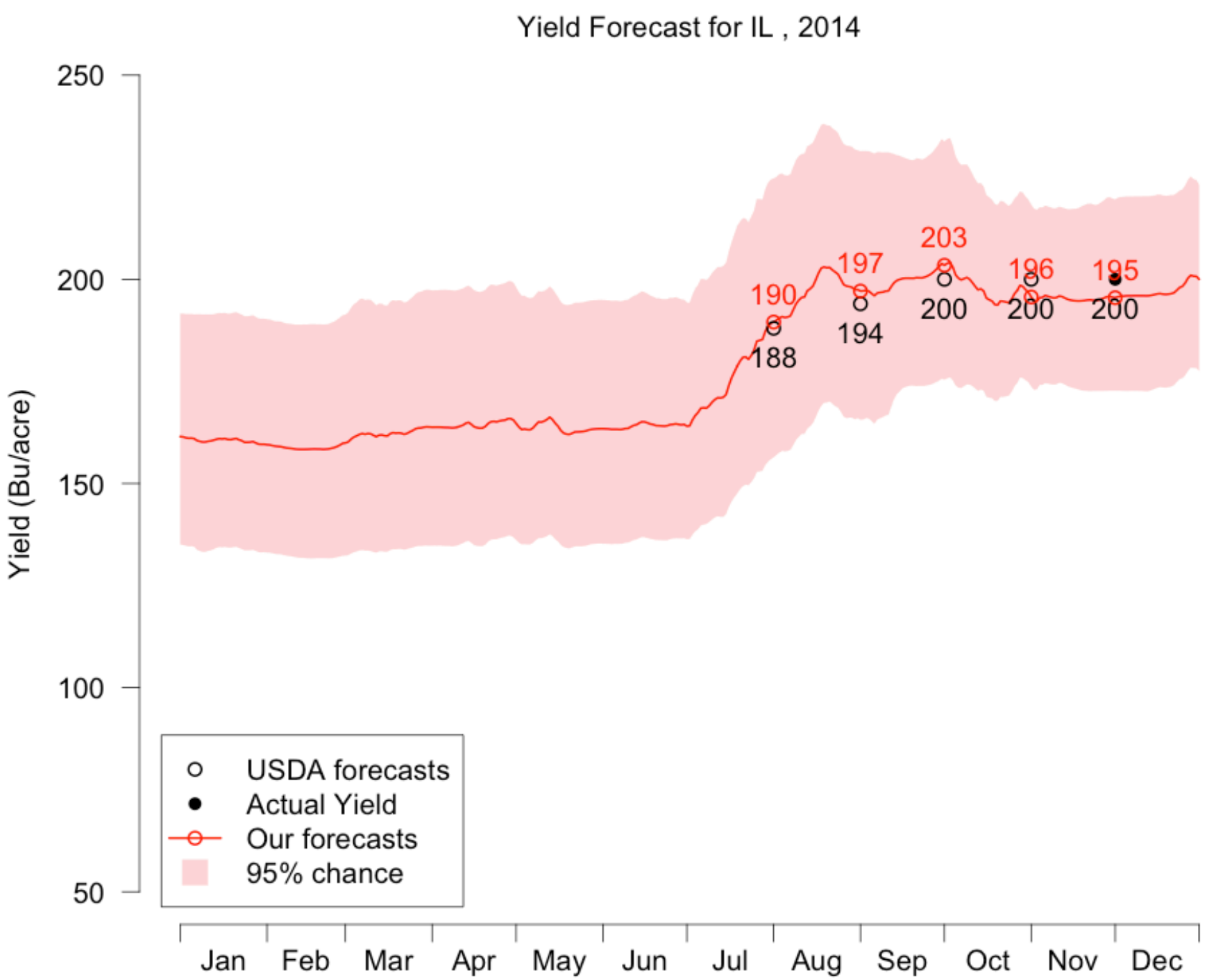

40
30
20
10
0
-10
-20
Tmax
realized
future
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

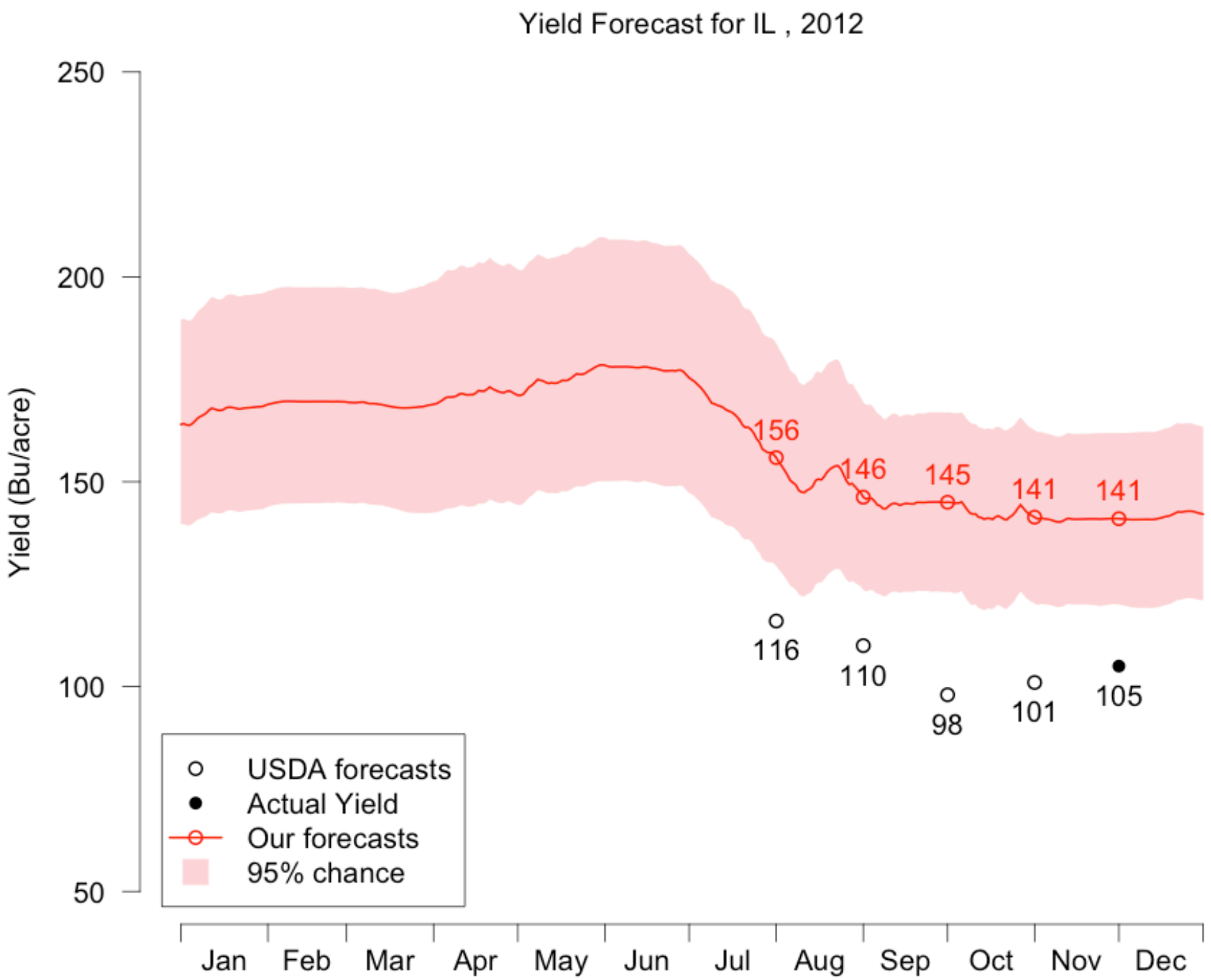

40
30
20
10
0
-10
-20
Tmax
realized
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

Forecasting (state-level yields)


Yield Forecast for IL , 2004
Yield (Bu/acre)
250
200
150
100
50
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
193
187
186
189
165
168
170
180
180
180
USDA forecasts
Actual Yield
Our forecasts
95% chance


Yield Forecast for IL , 2005
Yield (Bu/acre)
250
200
150
100
50
154
147
125
136
134
145
134
145
138
143
USDA forecasts
Actual Yield
Our forecasts
95% chance
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

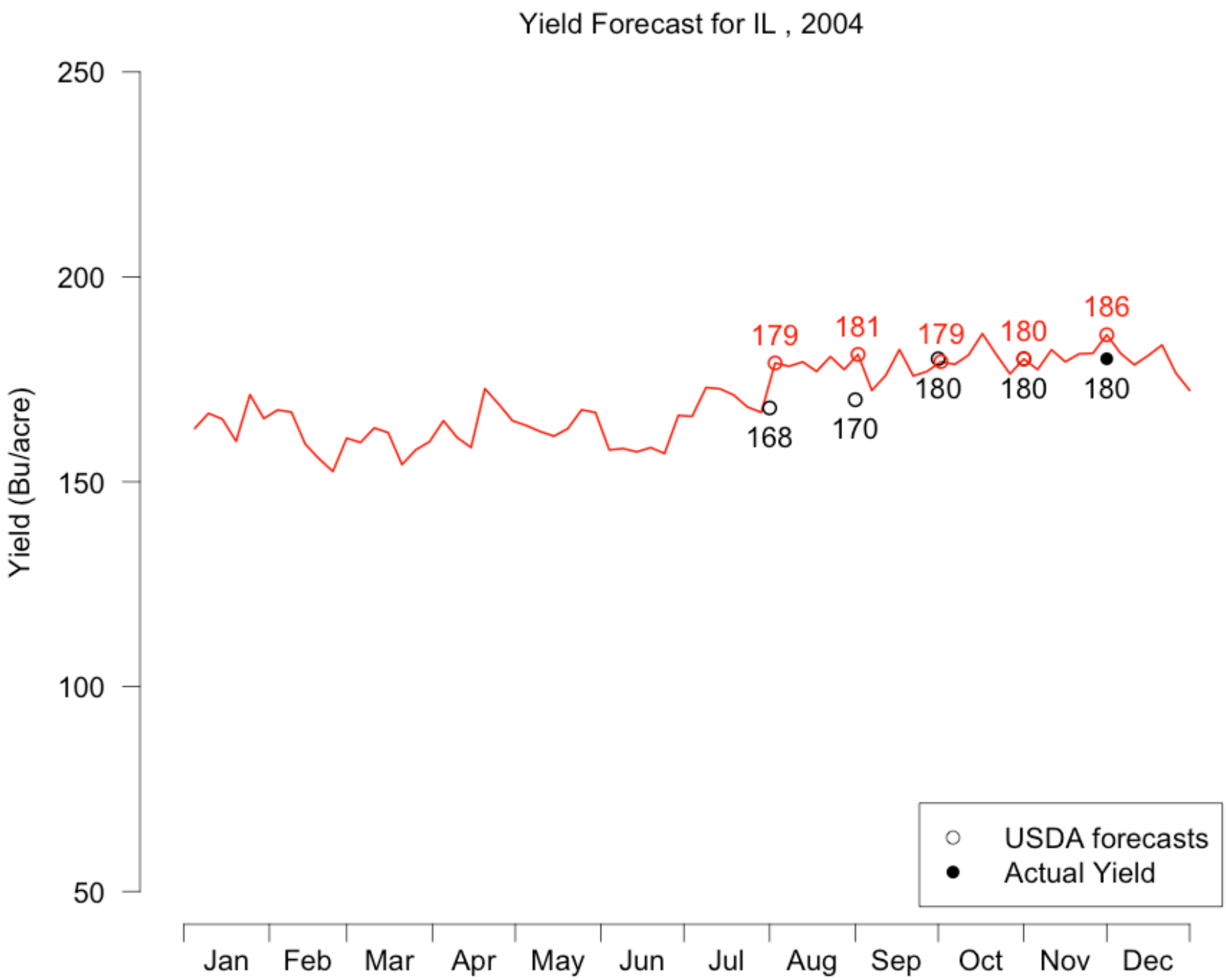

Yield Forecast for IL , 2014
Yield (Bu/acre)
250
200
150
100
50
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
190
197
203
196
195
188
194
200
200
200
USDA forecasts
Actual Yield
Our forecasts
95% chance

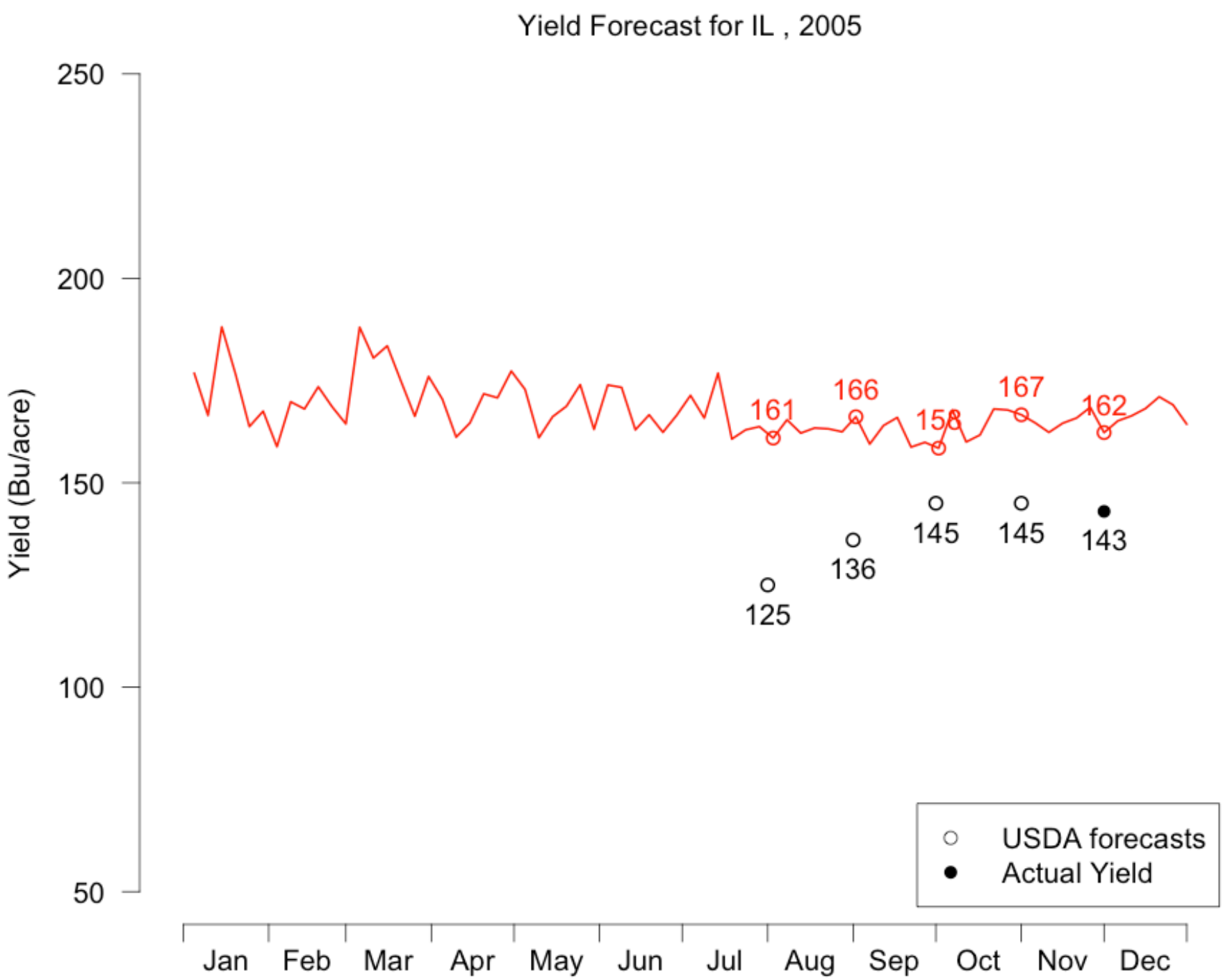

Yield Forecast for IL , 2012
Yield (Bu/acre)
250
200
150
100
50
156
146
145
141
141
116
110
98
101
105
USDA forecasts
Actual Yield
Our forecasts
95% chance
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

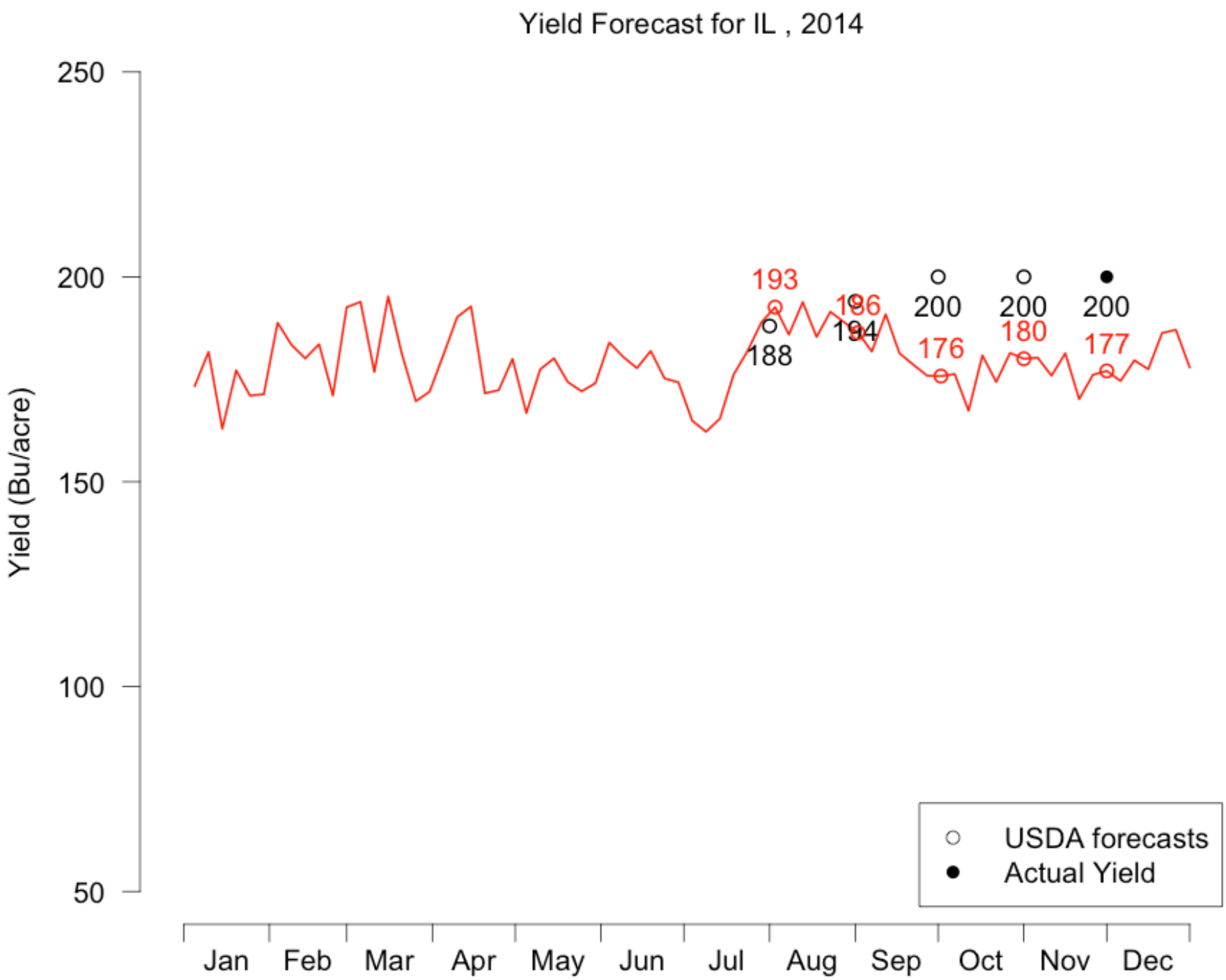
Serial model approach

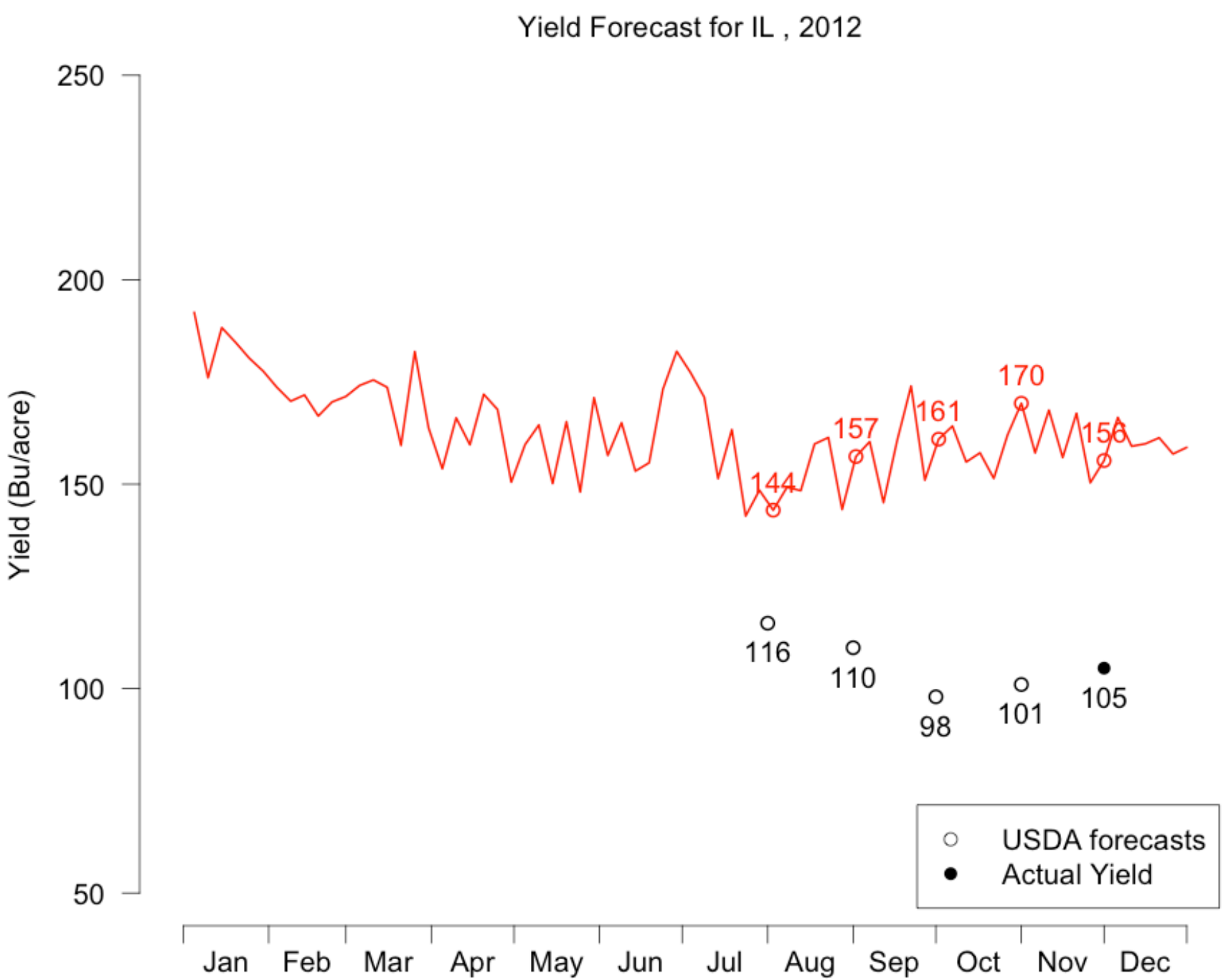
Alternative model

- Implement a different model for each period (e.g. day, pentad, week, month)
- Model used is a boosted regression tree (XGBoost)


Yield Forecast for IL , 2004
Yield (Bu/acre)
250
200
150
100
50
179
181
179
180
186
168
170
180
180
180
USDA forecasts
Actual Yield
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


Yield Forecast for IL , 2005
Yield (Bu/acre)
250
200
150
100
50
161
166
158
167
162
125
136
145
145
143
USDA forecasts
Actual Yield
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


Yield Forecast for IL , 2014
Yield (Bu/acre)
250
200
150
100
50
193
188
186
194
200
176
200
180
200
177
USDA forecasts
Actual Yield
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


Yield Forecast for IL , 2012
Yield (Bu/acre)
250
200
150
100
50
144
157
161
170
156
116
110
98
101
105
USDA forecasts
Actual Yield
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

Next steps

- Improve the one-model approach (weather uncertainty, tensor splines)
- Improve the serial model approach
- Incorporate forecasted weather in both approaches

Thank you

Ariel Ortiz-Bobea
Assistant Professor

Twitter @ArielOrtizBobea
Email ao332@cornell.edu
Web ortiz-bobea.dyson.cornell.edu