IE 2014 NNN 1

Segmented Regression Models

Milo Schield
Augsburg College
Editor of www.StatLit.org
US Rep: International Statistical Literacy Project

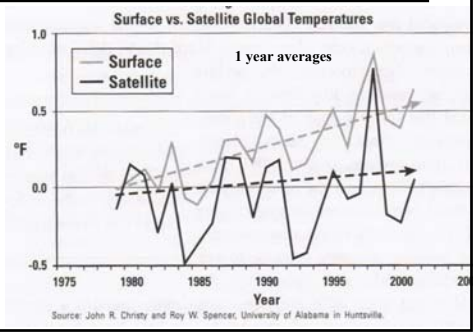
Fall 2014
National Numeracy Network Conference
www.StatLit.org/pdf/2014-Schield-NNN5-Slides.pdf

IE 2014 NNN 2

Are Global Temperatures Increasing

Which source?

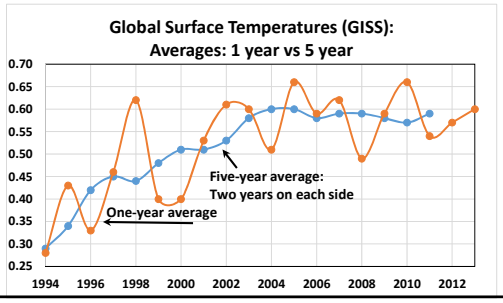
Surface or satellite based?



IE 2014 NNN 3

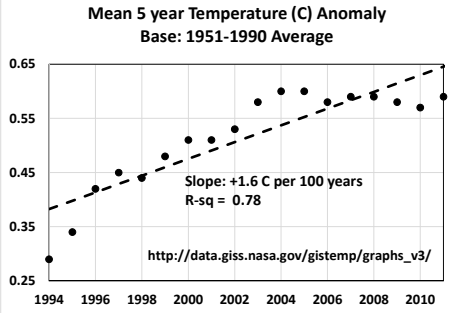
Are Global Surface Temperatures Still Increasing

Averaged over what time period? One-year or five?



IE 2014 NNN 4

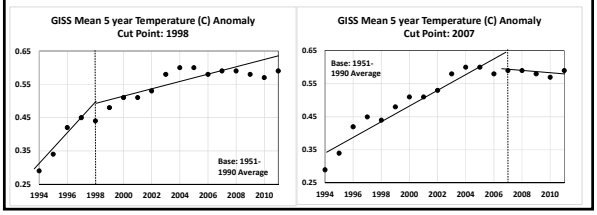
Global Surface Temperatures: Are they Still Increasing?



IE 2014 NNN 5

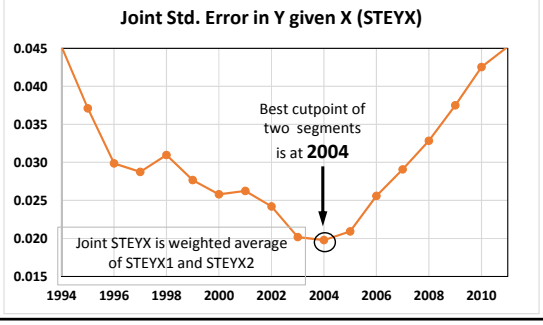
Using a Two-Segment Model

Least-squares regression works when data is nearly linear.
Rather than transform, consider a segmented linear model.
The goal is unchanged: minimum variation about model.



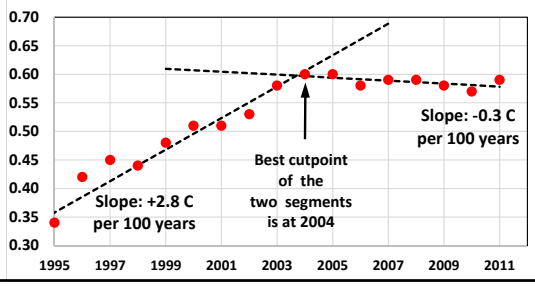
IE 2014 NNN 6

Minimize Total Error Relative to Predicted



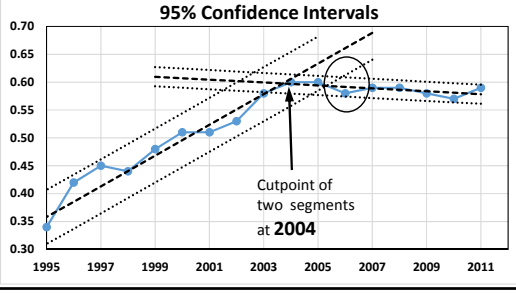
IE 2014 NNN 7

Best fit Two-Segment Model



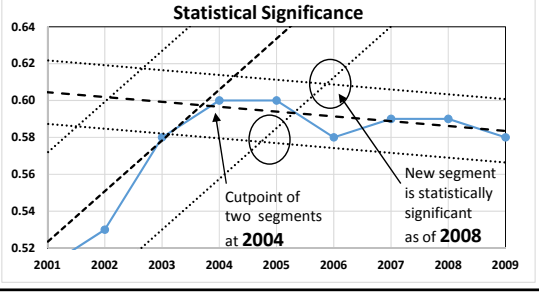
IE 2014 NNN 8

Two-Segment Model: 95% Confidence Intervals



IE 2014 NNN 9

Is the Segmentation Statistically Significant?



IE 2014 NNN 10

Conclusion

Five-year averages of global surface temperatures:

From 1994-2004, they trended up: 2.8° C per century.

Since 2004, they trended down: -0.3° C per century

After 2008 a statistician could say: "In 2004 - 2013, the trend in five-year averaged global surface temperatures changed from positive (2.8 C per 100 years) to negative (-0.3 C per 100 years) and this change in trend was statistically-significant."

IE 2014 NNN 11

Create Line1

1. Current row = 1995.
2. Fit 5 year data from 1994 to current row.
3. Calculate slope b1 using Excel SLOPE.
3. Calculate Std. Error of Y given X using Excel STEYX.
4. Increase current row; Repeat 2, 3 & 4.

Out-of-control???

	DATA		LINE1	LINE1
Year	Ave5yr		b1	STEYX1
1994	0.29			
1995	0.34		0.050	
1996	0.42		0.065	0.012
1997	0.45		0.056	0.014
1998	0.44		0.041	0.030
1999	0.48		0.037	0.028
2000	0.51		0.034	0.026
2001	0.51		0.030	0.028
2002	0.53		0.028	0.028
2003	0.58		0.028	0.026
2004	0.60		0.028	0.025
2005	0.60		0.026	0.025
2006	0.58		0.024	0.030
2007	0.59		0.022	0.033
2008	0.59		0.020	0.036
2009	0.58		0.019	0.040
2010	0.57		0.017	0.044
2011	0.59		0.015	0.045

IE 2014 NNN 12

Create Line 2 Series; Calculate Joint STEYX

Out of control?

	DATA		LINE1	LINE1		LINE2	LINE2		Joint
Year	Ave5yr		b1	STEYX1		b2	STEYX2		STEYX
1994	0.29					0.015	0.045		0.0452
1995	0.34		0.050			0.013	0.038		0.0371
1996	0.42		0.065	0.012		0.011	0.031		0.0299
1997	0.45		0.056	0.014		0.011	0.031		0.0288
1998	0.44		0.041	0.030		0.010	0.031		0.0310
1999	0.48		0.037	0.028		0.008	0.028		0.0277
2000	0.51		0.034	0.026		0.006	0.026		0.0258
2001	0.51		0.030	0.028		0.005	0.025		0.0262
2002	0.53		0.028	0.028		0.002	0.020		0.0242
2003	0.58		0.028	0.026		-0.001	0.010		0.0202
2004	0.60		0.028	0.025		-0.003	0.009		**0.0198**
2005	0.60		0.026	0.025		-0.002	0.009		0.0209
2006	0.58		0.024	0.030		-0.001	0.009		0.0256
2007	0.59		0.022	0.033		-0.002	0.010		0.0291
2008	0.59		0.020	0.036		-0.001	0.012		0.0328
2009	0.58		0.019	0.040		0.005	0.012		0.0375
2010	0.57		0.017	0.044		0.020			0.0425
2011	0.59		0.015	0.045					0.0452

References

Wikipedia: Change Detection
Wikipedia: Time-series segmentation
Wikipedia: Time Series [Segmentation]
Wikipedia: Regression Analysis

Segmented Regression Models

Milo Schield
Augsburg College
Editor of www.StatLit.org
US Rep: International Statistical Literacy Project

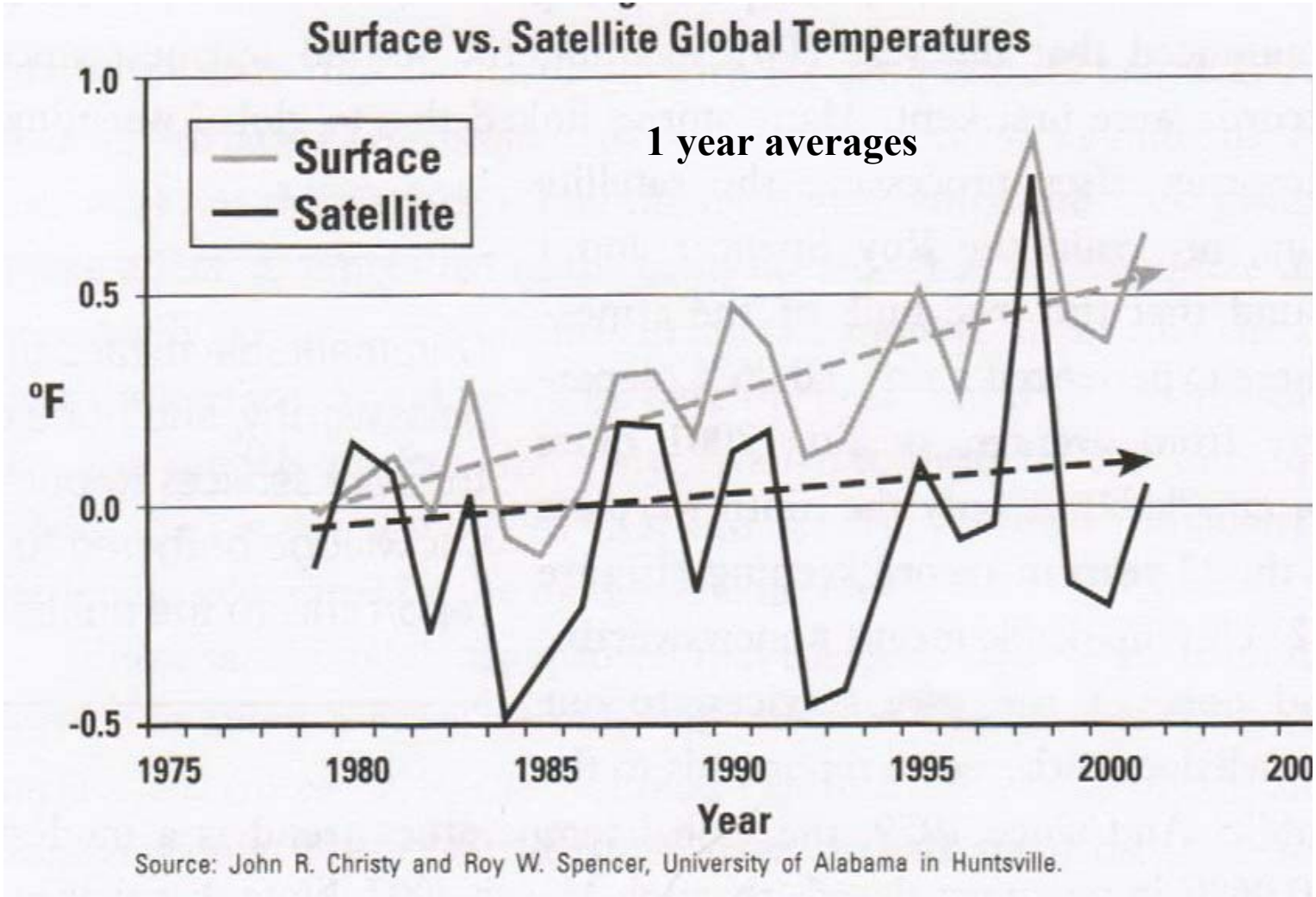
Fall 2014
National Numeracy Network Conference

www.StatLit.org/pdf/2014-Schield-NNN5-Slides.pdf

Are Global Temperatures Increasing

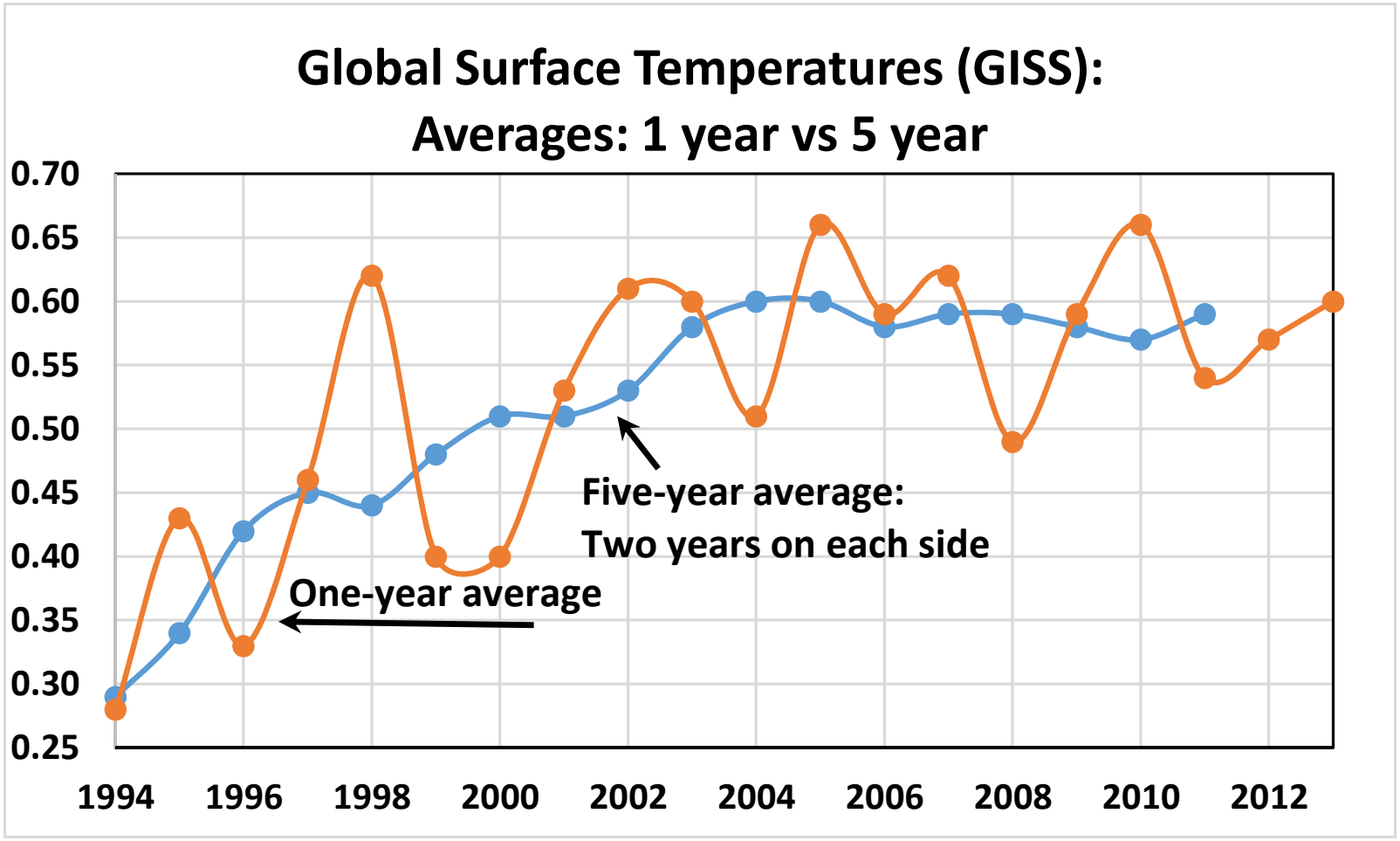
Which source?

Surface or satellite based?

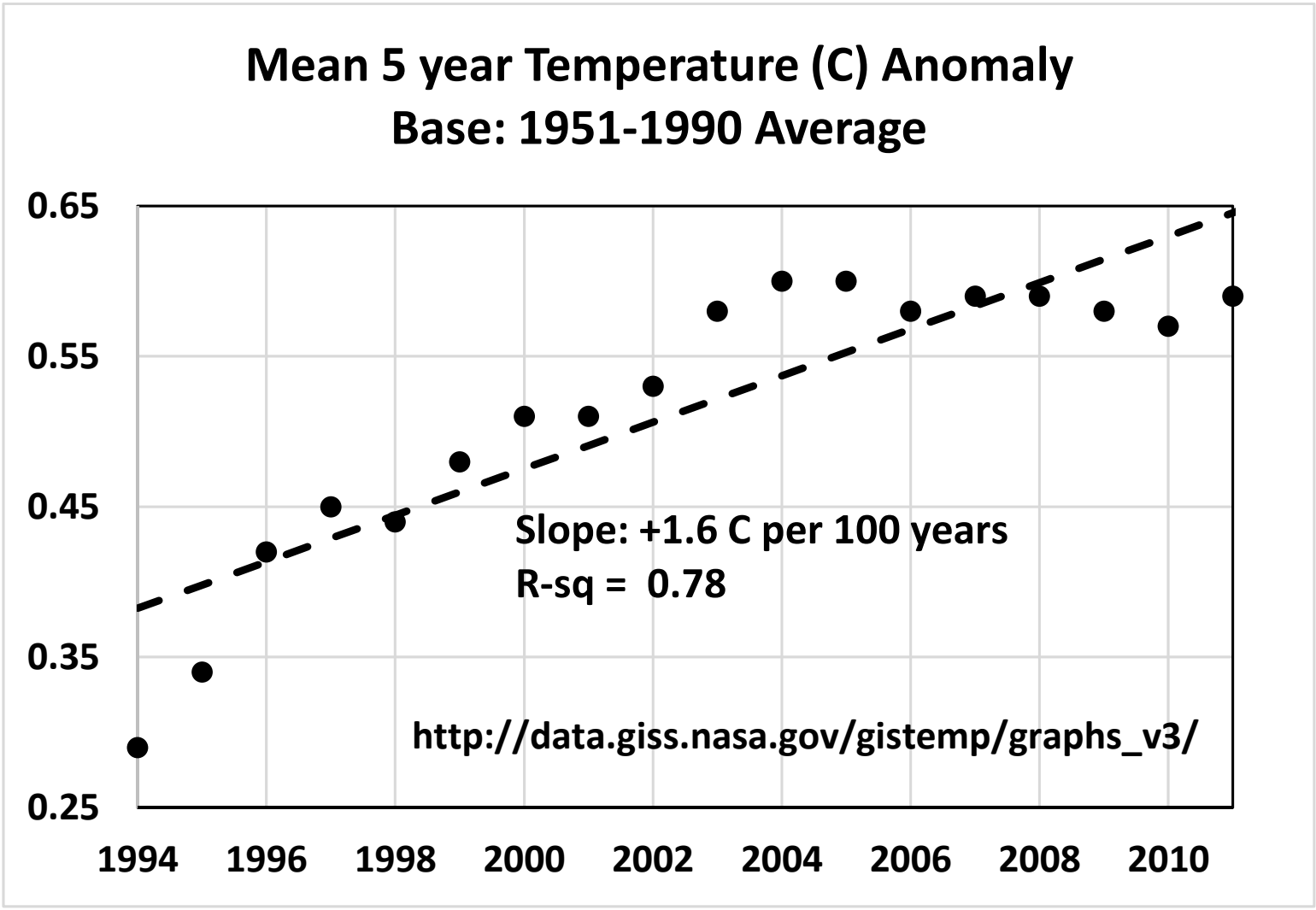


Are Global Surface Temperatures Still Increasing

Averaged over what time period? One-year or five?



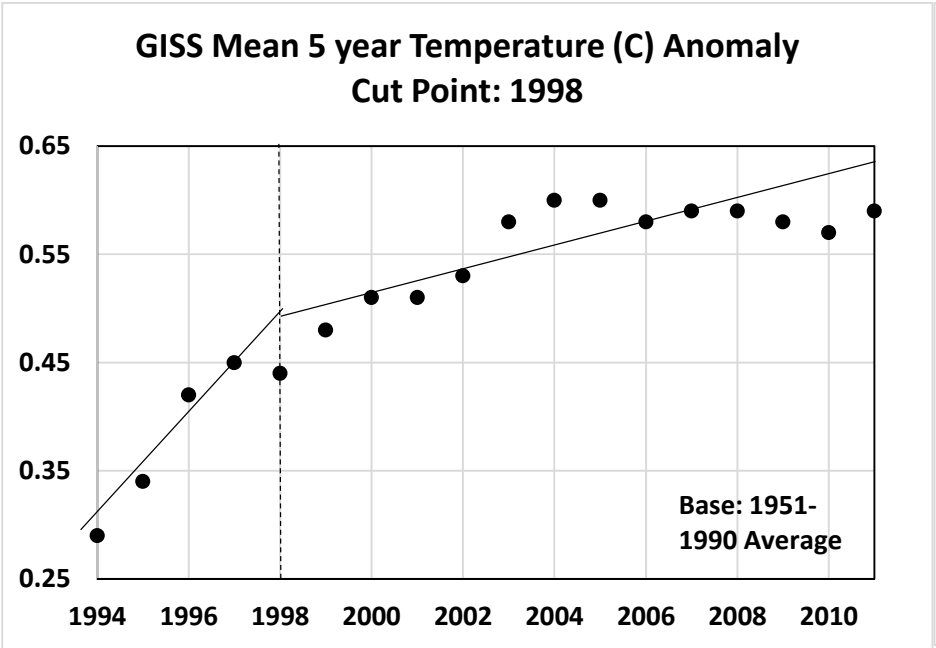
Global Surface Temperatures: Are they Still Increasing?

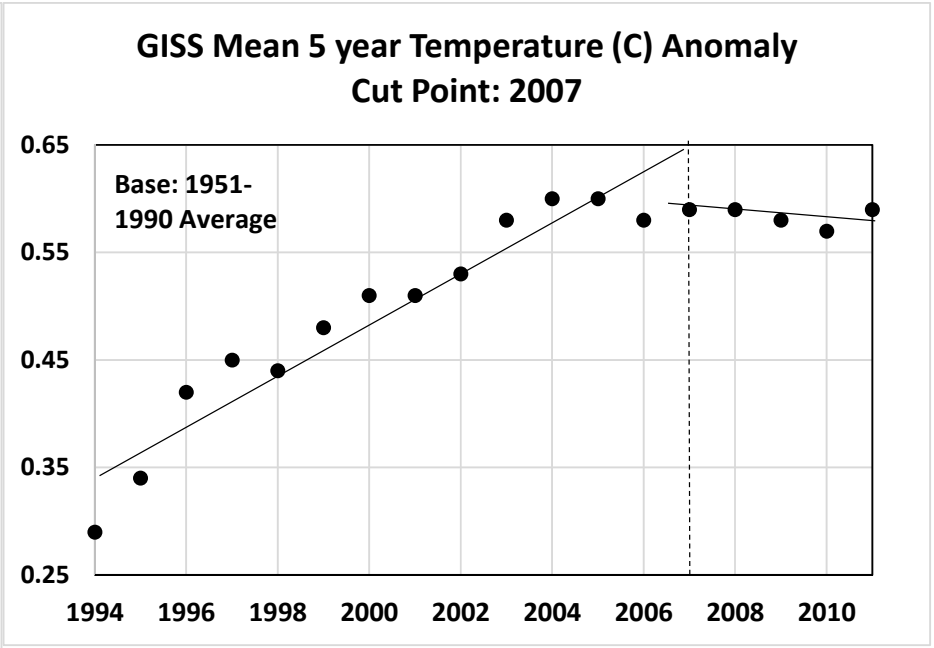


Using a Two-Segment Model

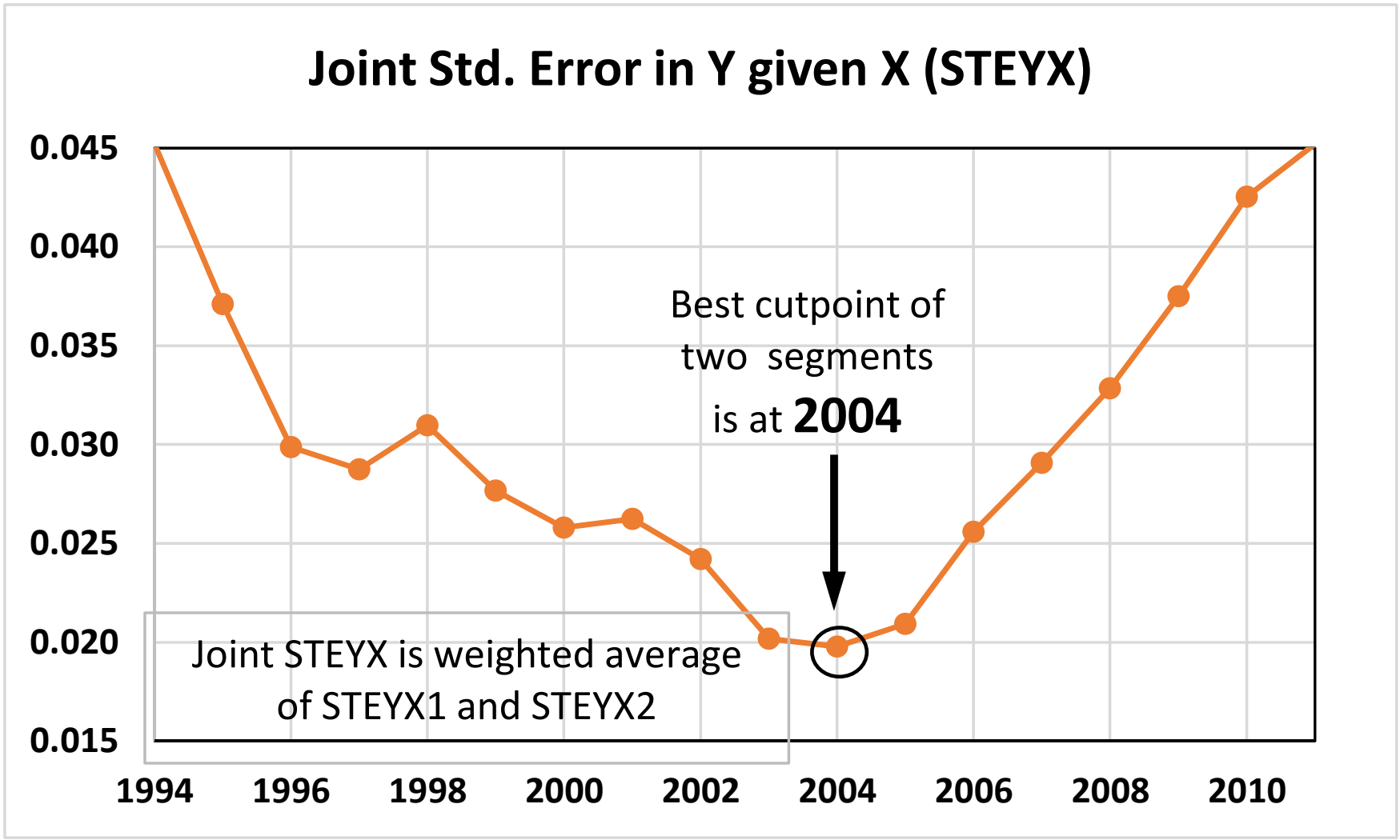
Least-squares regression works when data is nearly linear. Rather than transform, consider a segmented linear model.

The goal is unchanged: minimum variation about model.

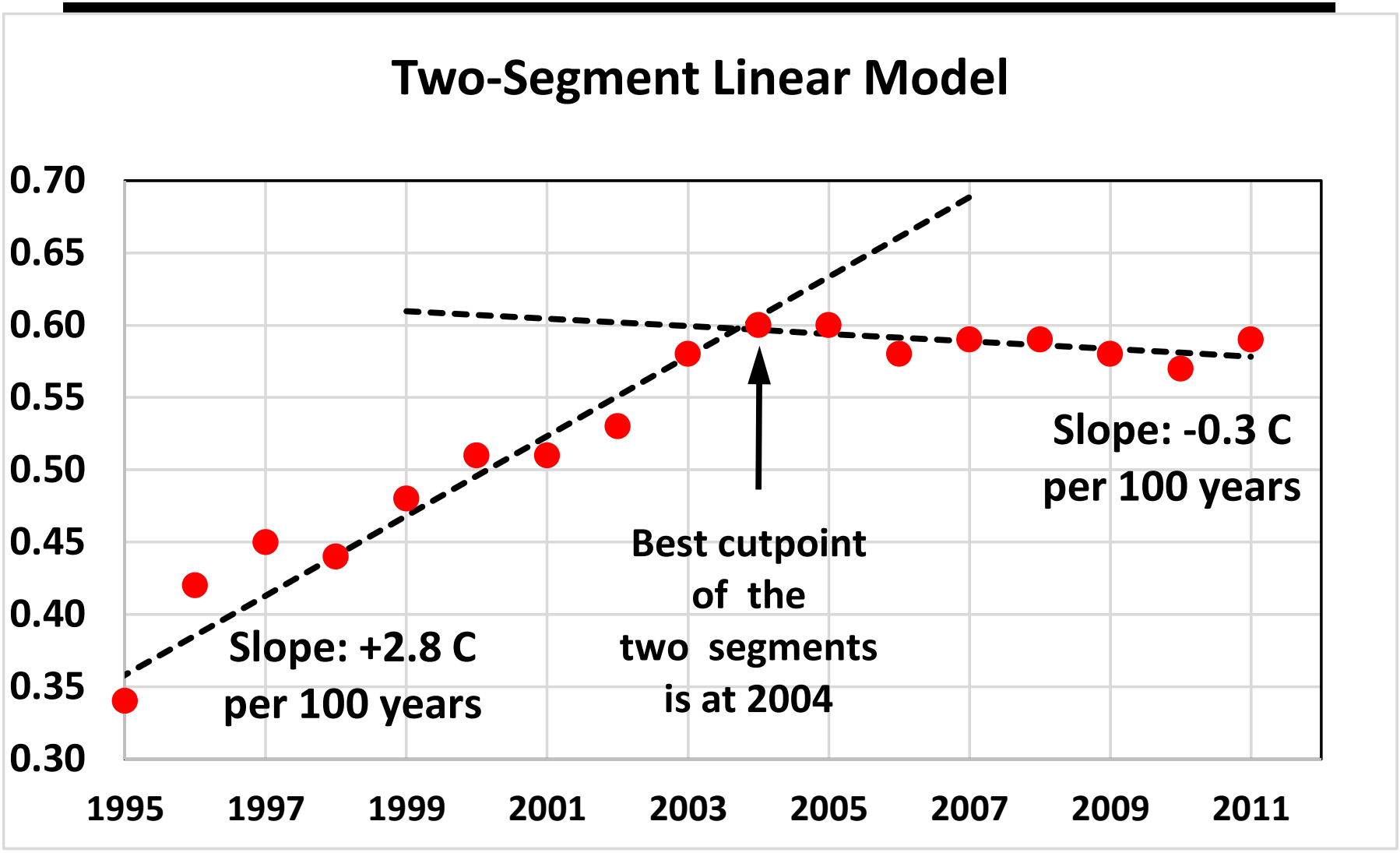




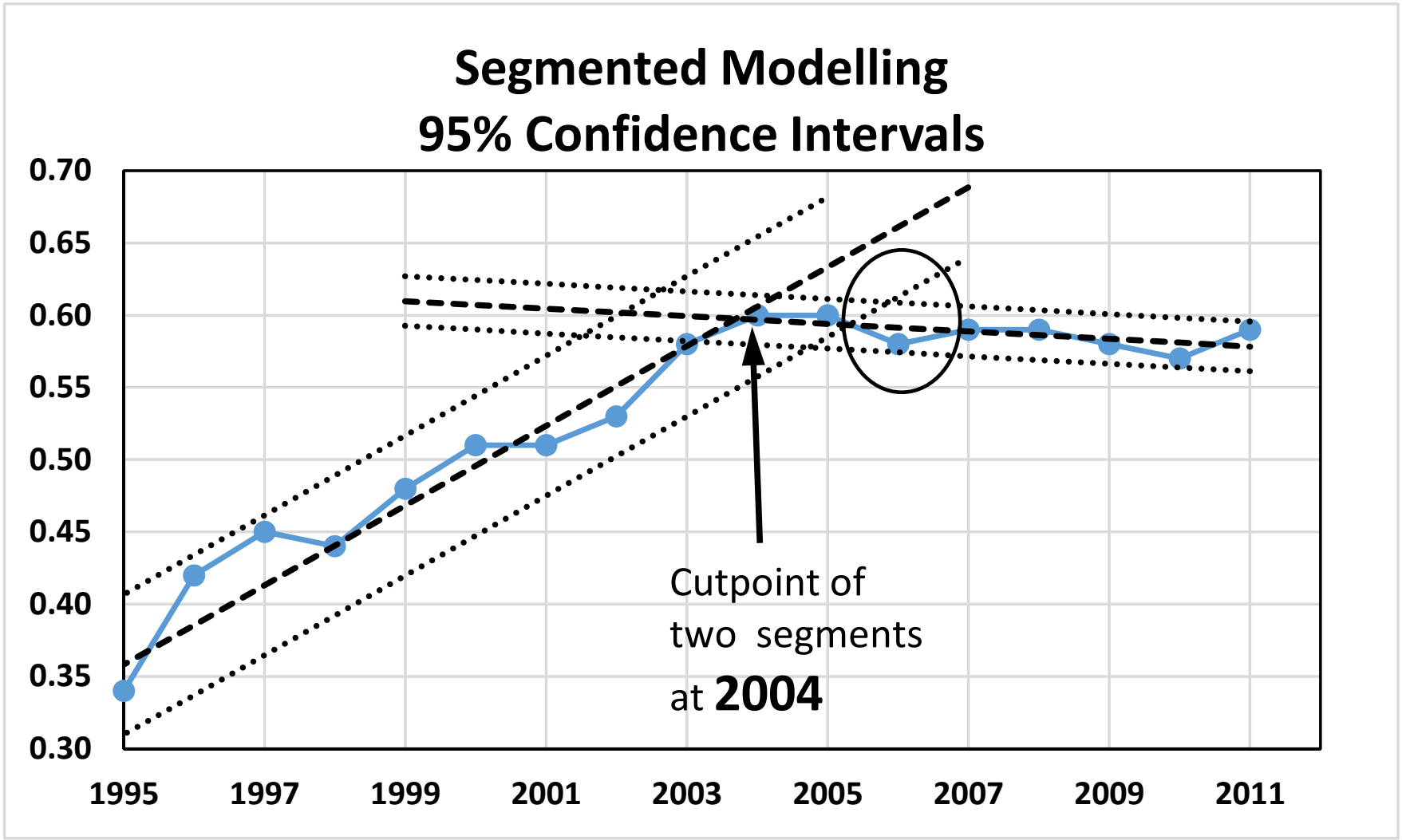
Minimize Total Error Relative to Predicted



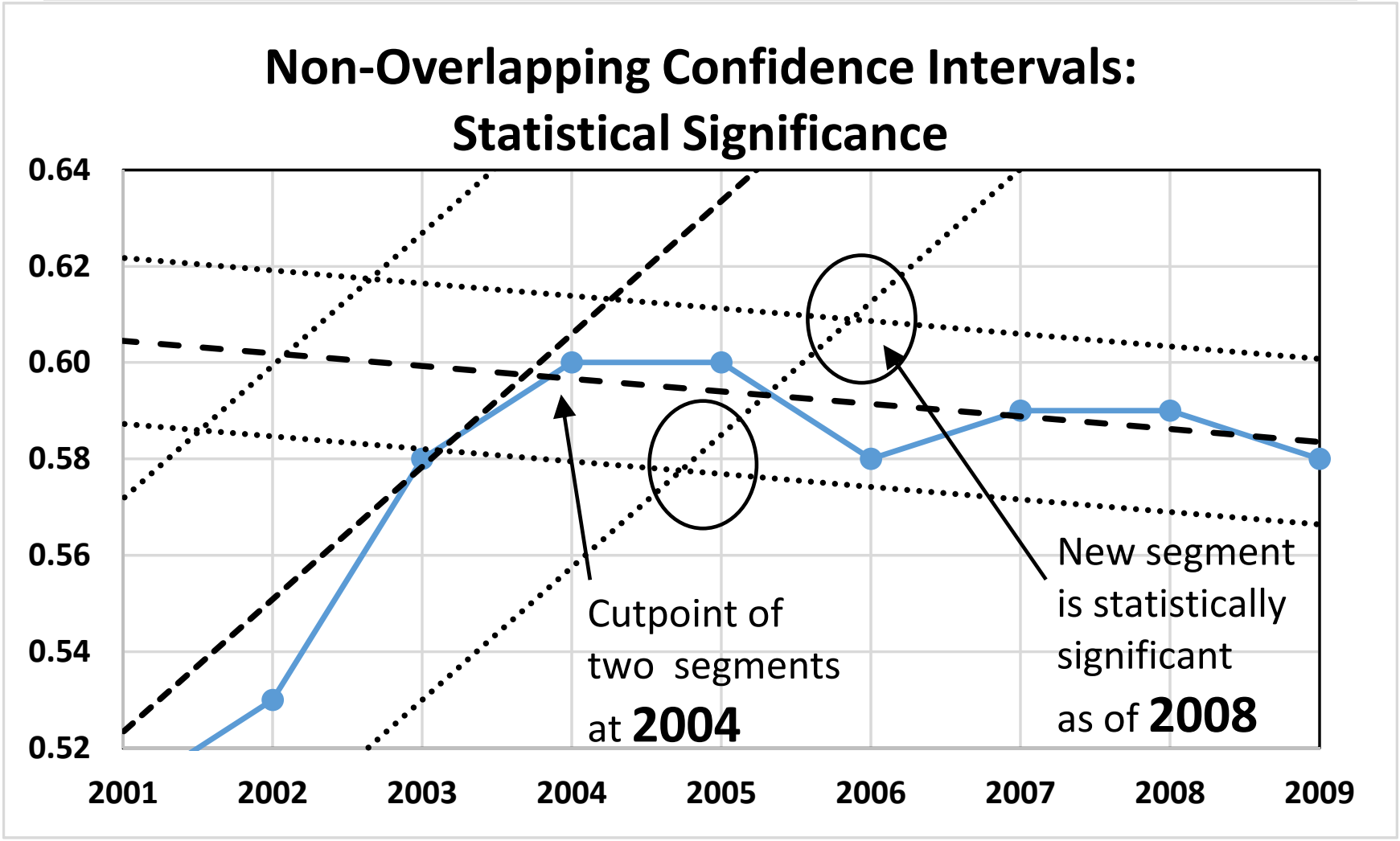
Best fit
Two-Segment Model



Two-Segment Model: 95% Confidence Intervals



Is the Segmentation Statistically Significant?



Conclusion

Five-year averages of global surface temperatures:

From 1994-2004, they trended up: 2.8° C per century.

Since 2004, they trended down: -0.3° C per century

After 2008 a statistician could say: “In 2004 - 2013, the trend in five-year averaged global surface temperatures changed from positive (2.8 C per 100 years) to negative (-0.3 C per 100 years) and this change in trend was statistically-significant.”

Create Line1

1. Current row = 1995.
2. Fit 5 year data from 1994 to current row.
3. Calculate slope b1 using Excel SLOPE.
3. Calculate Std. Error of Y given X using Excel STEYX.
4. Increase current row; Repeat 2, 3 & 4.

Out-of-control???

	DATA		LINE1	LINE1
Year	Ave5yr		b1	STEYX1
1994	0.29			
1995	0.34		0.050	
1996	0.42		0.065	0.012
1997	0.45		0.056	0.014
1998	0.44		0.041	0.030
1999	0.48		0.037	0.028
2000	0.51		0.034	0.026
2001	0.51		0.030	0.028
2002	0.53		0.028	0.028
2003	0.58		0.028	0.026
2004	0.60		0.028	0.025
2005	0.60		0.026	0.025
2006	0.58		0.024	0.030
2007	0.59		0.022	0.033
2008	0.59		0.020	0.036
2009	0.58		0.019	0.040
2010	0.57		0.017	0.044
2011	0.59		0.015	0.045

Create Line 2 Series; Calculate Joint STEYX

Out of control?

Year	DATA Ave5yr		LINE1 b1	LINE1 STEYX1		LINE2 b2	LINE2 STEYX2		Joint STEYX
1994	0.29					0.015	0.045		0.0452
1995	0.34		0.050			0.013	0.038		0.0371
1996	0.42		0.065	0.012		0.011	0.031		0.0299
1997	0.45		0.056	0.014		0.011	0.031		0.0288
1998	0.44		0.041	0.030		0.010	0.031		0.0310
1999	0.48		0.037	0.028		0.008	0.028		0.0277
2000	0.51		0.034	0.026		0.006	0.026		0.0258
2001	0.51		0.030	0.028		0.005	0.025		0.0262
2002	0.53		0.028	0.028		0.002	0.020		0.0242
2003	0.58		0.028	0.026		-0.001	0.010		0.0202
2004	0.60		0.028	0.025		-0.003	0.009		**0.0198**
2005	0.60		0.027	0.025		-0.002	0.009		0.0209
2006	0.58		0.024	0.030		-0.001	0.009		0.0256
2007	0.59		0.022	0.033		-0.002	0.010		0.0291
2008	0.59		0.020	0.036		-0.001	0.012		0.0328
2009	0.58		0.019	0.040		0.005	0.012		0.0375
2010	0.57		0.017	0.044		0.020			0.0425
2011	0.59		0.015	0.045					0.0452

References

Wikipedia: Change Detection

Wikipedia: Time-series segmentation

Wikipedia: Time Series [Segmentation]

Wikipedia: Regression Analysis