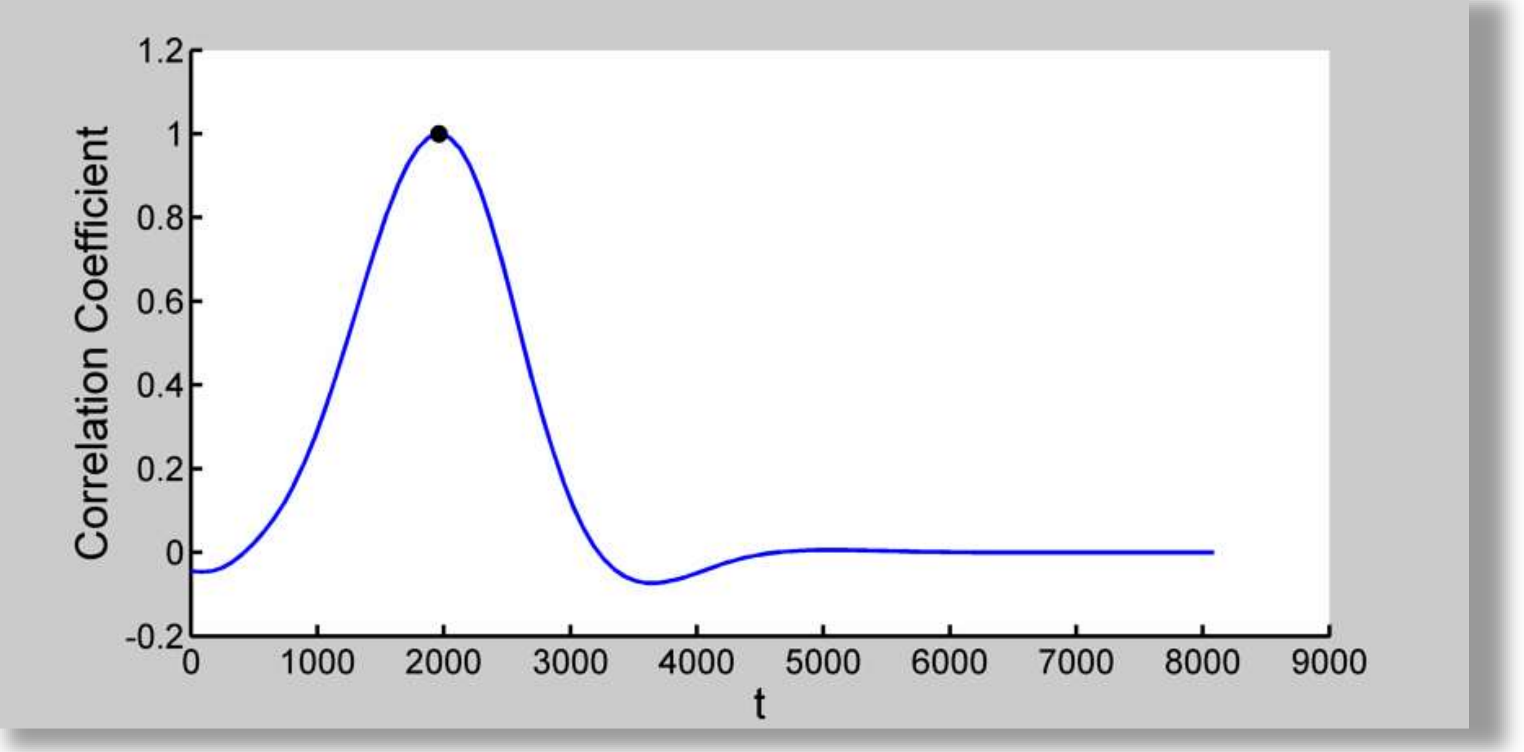
# Use and abuse of weighted means/linear regression - propagating random and systematic uncertainties

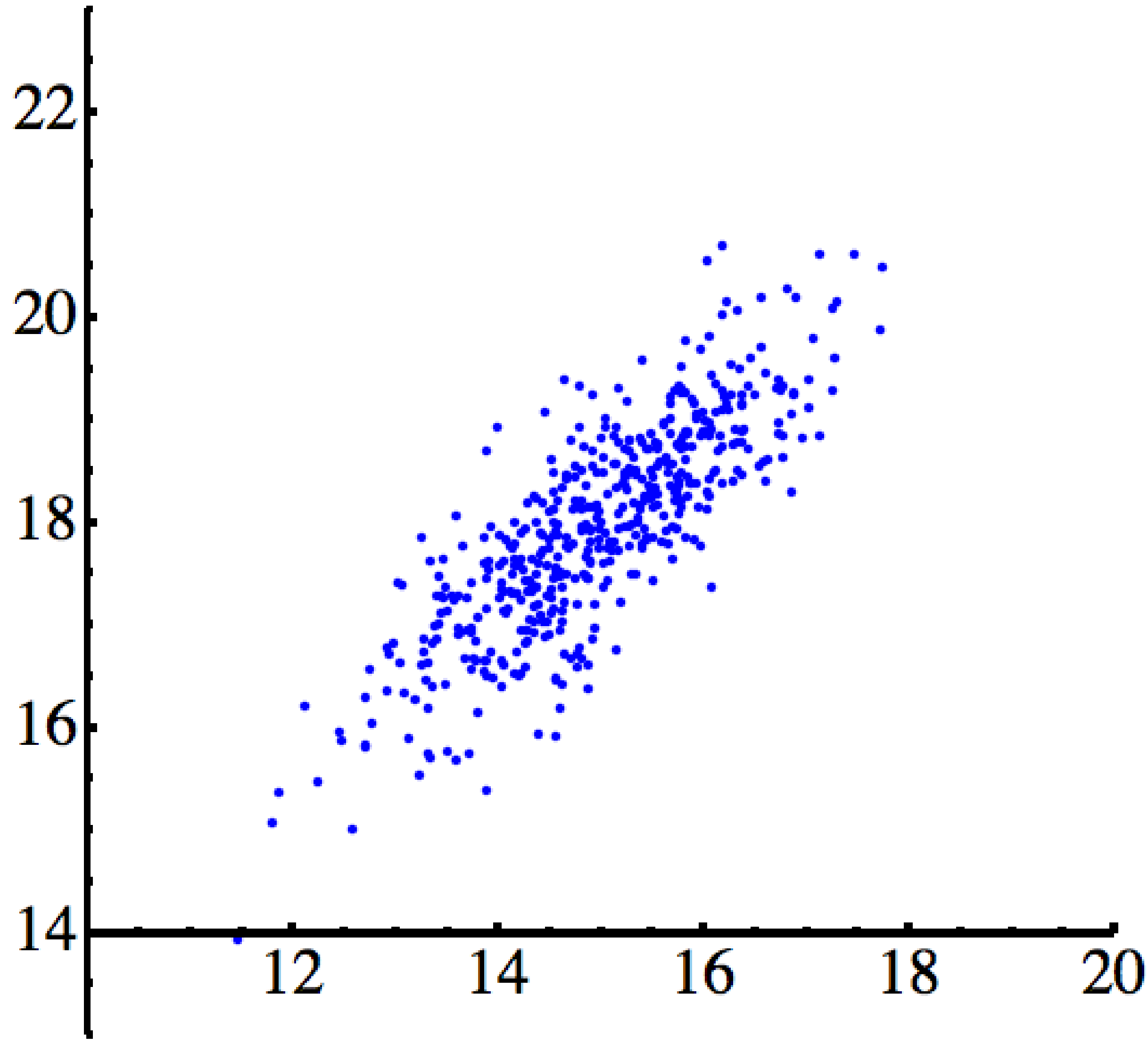


# Propagating systematic uncertainties

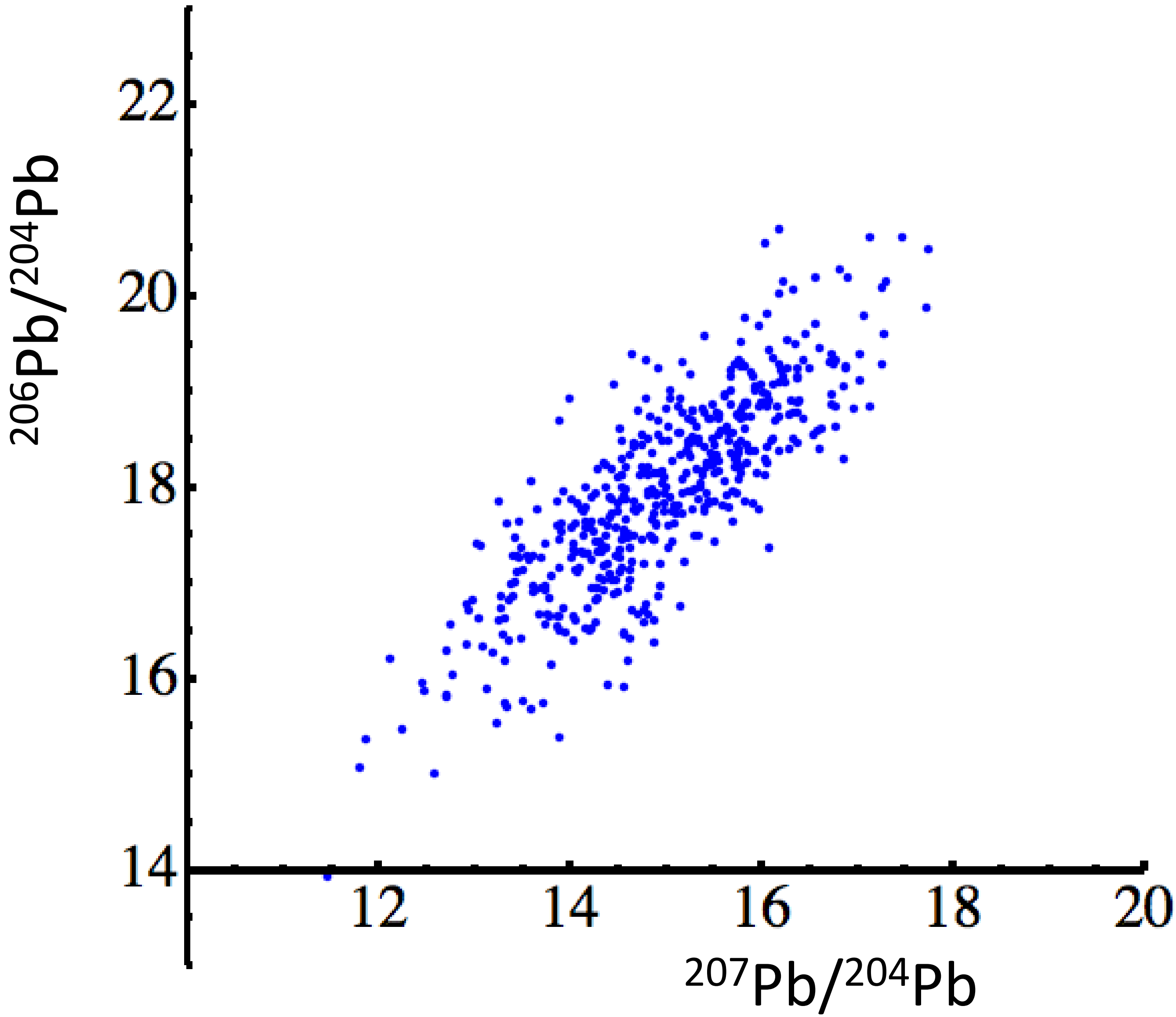
- if *z* is a function of *x* and *y*

$$\sigma_z^2 = \sigma_x^2 \left(\frac{dz}{dx}\right)^2 + 2\sigma_{xy}^2 \left(\frac{dz}{dx}\right)\left(\frac{dz}{dy}\right) + \sigma_y^2 \left(\frac{dz}{dy}\right)^2$$

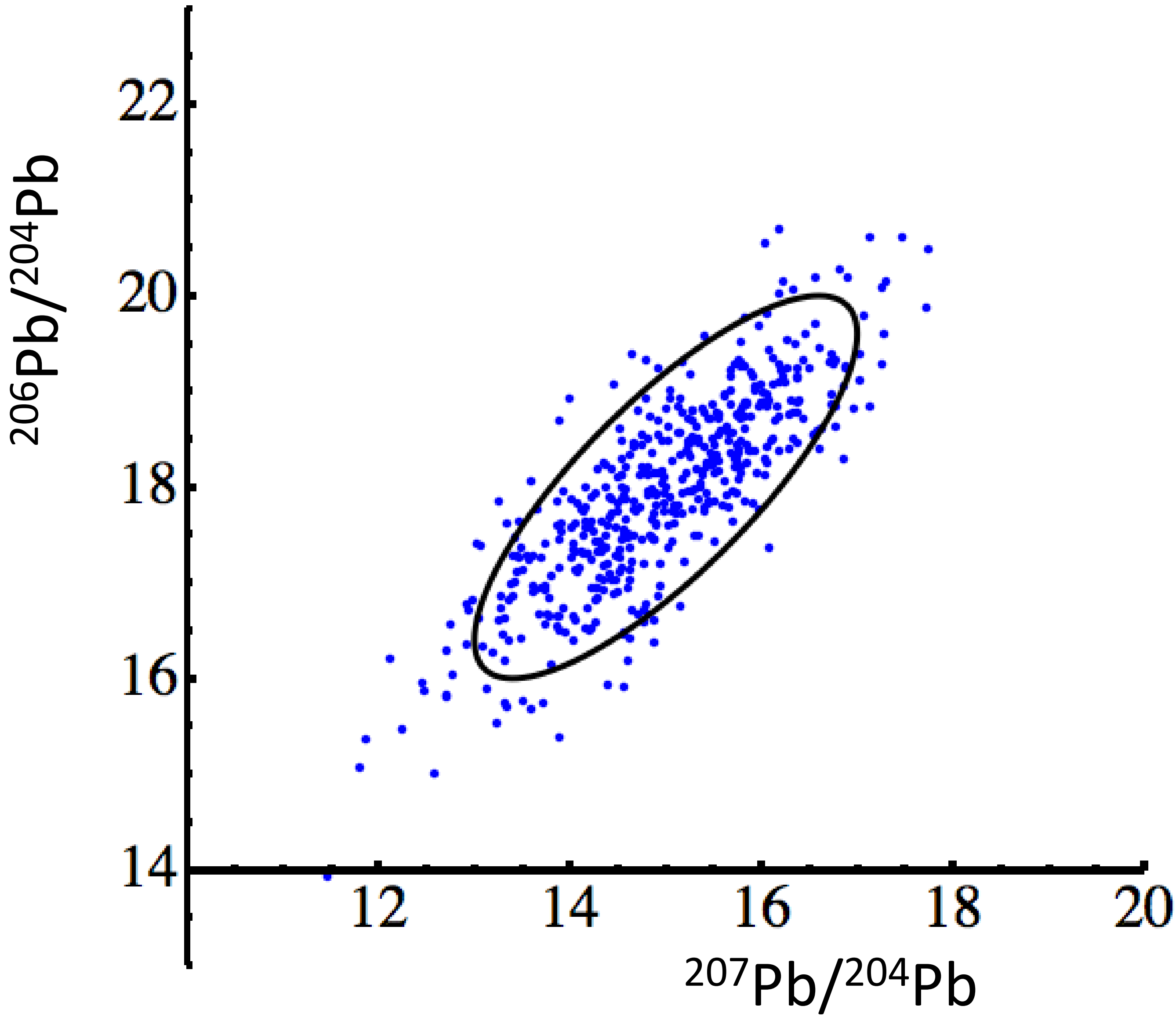
# Covariance/Correlation



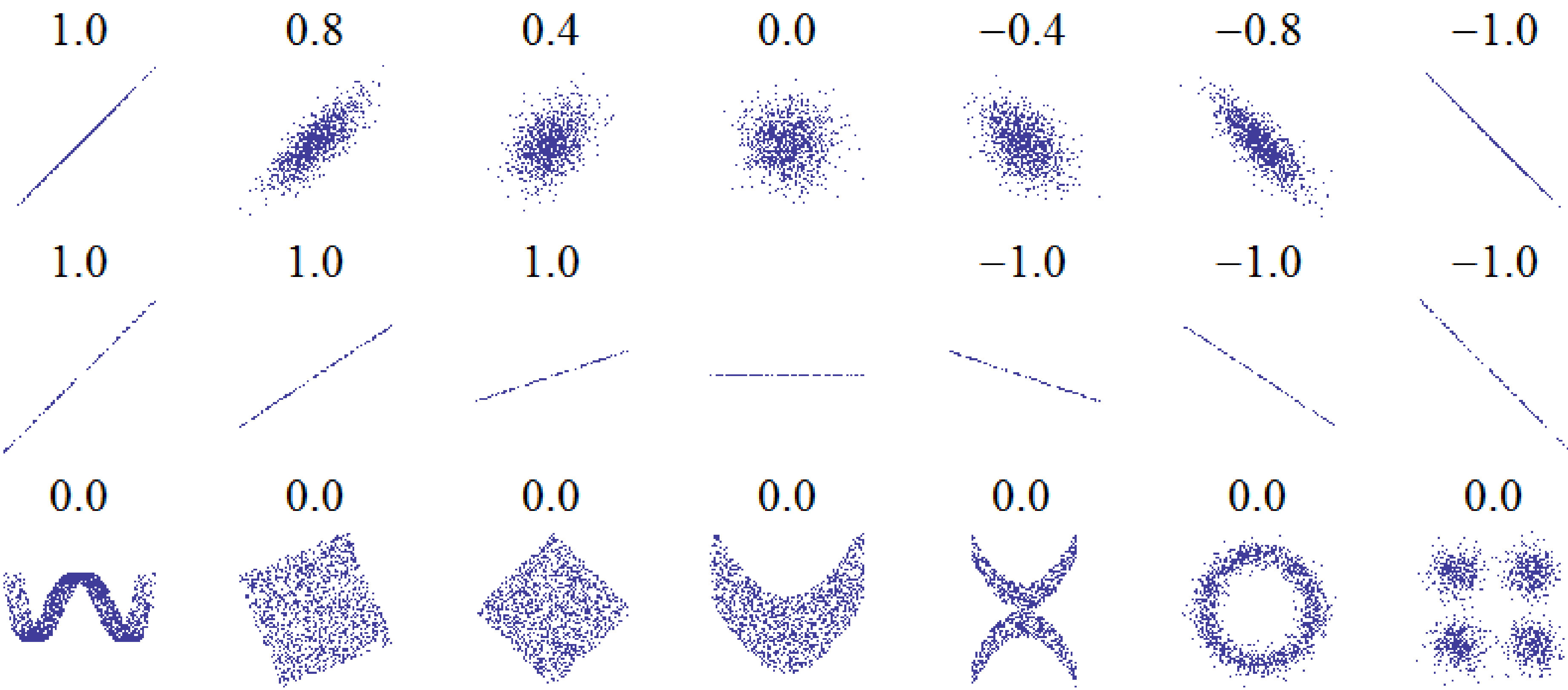
# Covariance/Correlation



# Covariance/Correlation



# Other Examples




from www.wikipedia.org

# Propagating systematic uncertainties

- if *z* is a function of *x* and *y*

$$\sigma_z^2 = \sigma_x^2 \left(\frac{dz}{dx}\right)^2 + 2\sigma_{xy}^2 \left(\frac{dz}{dx}\right)\left(\frac{dz}{dy}\right) + \sigma_y^2 \left(\frac{dz}{dy}\right)^2$$

McLean et al., 2011 (G^3)

# Propagating systematic uncertainties

- if *z* is a function of *x* and *y*

$$\sigma_z^2 = \sigma_x^2 \left(\frac{dz}{dx}\right)^2 + 2\sigma_{xy}^2 \left(\frac{dz}{dx}\right)\left(\frac{dz}{dy}\right) + \sigma_y^2 \left(\frac{dz}{dy}\right)^2$$

- *or*

$$\sigma_z^2 = \begin{bmatrix} \frac{dz}{dx} & \frac{dz}{dy} \end{bmatrix} \begin{bmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{bmatrix} \begin{bmatrix} \frac{dz}{dx} \\ \frac{dz}{dy} \end{bmatrix}$$

McLean et al., 2011 (G^3)

# Covariance vs. Correlation

- Covariance, like variance, expresses an average sum of multiplied distances

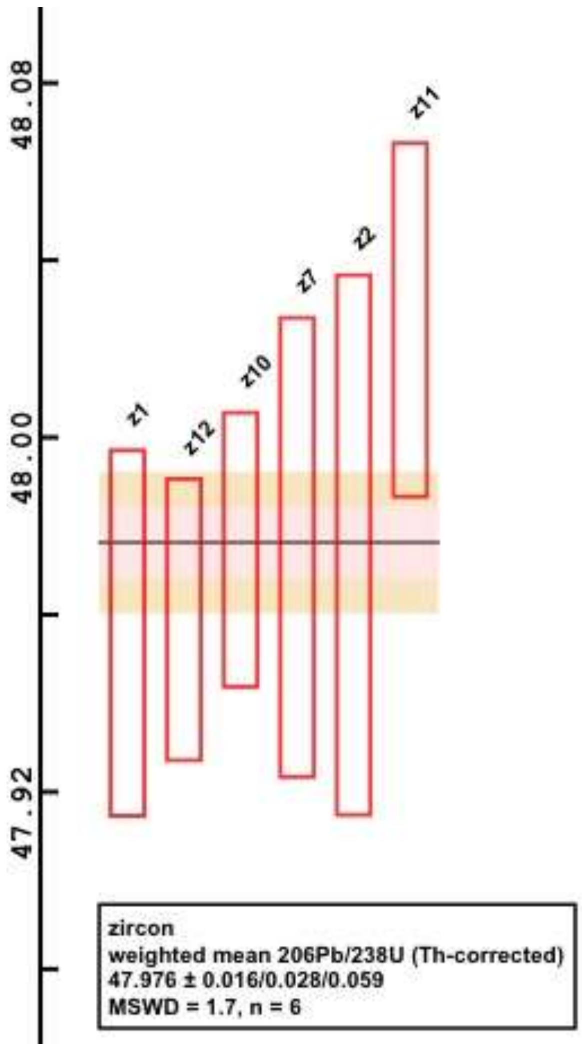
$$\sigma_{xy}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$\sigma_{xy}^2 = \sigma_c^2 \left(\frac{\partial x}{\partial c}\right)\left(\frac{\partial y}{\partial c}\right)$$

$$\rho_{xy} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}$$

[-1, 1]

# Calculating a weighted mean




- Gives more weight to more precise analyses, un-weights less precise analyses

$$\bar{t} = \sum_{i=1}^{n} \alpha_i t_i = \sum_{i=1}^{n} \left(\frac{t_i}{\sigma_i^2}\right) \Big/ \sum_{i=1}^{n} \left(\frac{1}{\sigma_i^2}\right)$$

- But there is no room here for systematic uncertainties, from sample-standard bracketing, standard ICs, decay constants…

# Systematic uncertainties are covariance

$$\bar{t} = \alpha_1 t_1 + \alpha_2 t_2 \qquad \alpha_1 + \alpha_2 = 1$$

$$\sigma_{\bar{t}}^2 = \begin{bmatrix} \frac{d\bar{t}}{dt_1} & \frac{d\bar{t}}{dt_2} \end{bmatrix} \begin{bmatrix} \sigma_{t_1}^2 & \sigma_{t_1 t_2}^2 \\ \sigma_{t_1 t_2}^2 & \sigma_{t_2}^2 \end{bmatrix} \begin{bmatrix} \frac{d\bar{t}}{dt_1} \\ \frac{d\bar{t}}{dt_2} \end{bmatrix}$$

- What are the weights that minimize the uncertainty in the weighted mean?

McLean et al., 2011 (G^3)

# Systematic uncertainties are covariance

$$\bar{t} = \alpha_1\, t_1 + \alpha_2\, t_2 \qquad\qquad \alpha_1 + \alpha_2 = 1$$

$$\sigma_{\bar{t}}^2 = \begin{bmatrix} \dfrac{d\bar{t}}{dt_1} & \dfrac{d\bar{t}}{dt_2} \end{bmatrix} \begin{bmatrix} \sigma_{t_1}^2 & \sigma_{t_1 t_2}^2 \\ \sigma_{t_1 t_2}^2 & \sigma_{t_2}^2 \end{bmatrix} \begin{bmatrix} \dfrac{d\bar{t}}{dt_1} \\ \dfrac{d\bar{t}}{dt_2} \end{bmatrix}$$

$$\alpha = \mathbf{\Sigma}^{-1}\mathbf{1} \Big/ \left(\mathbf{1}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{1}\right)$$

McLean et al., 2011 (G^3)

# Systematic uncertainties are covariance

$$\bar{t} = \alpha_1\, t_1 + \alpha_2\, t_2 \qquad\qquad \alpha_1 + \alpha_2 = 1$$

$$\sigma^2_{\bar{t}} = \begin{bmatrix} \dfrac{d\bar{t}}{dt_1} & \dfrac{d\bar{t}}{dt_2} \end{bmatrix} \begin{bmatrix} \sigma^2_{t_1} & \sigma^2_{t_1 t_2} \\ \sigma^2_{t_1 t_2} & \sigma^2_{t_2} \end{bmatrix} \begin{bmatrix} \dfrac{d\bar{t}}{dt_1} \\ \\ \dfrac{d\bar{t}}{dt_2} \end{bmatrix}$$

$$\bar{t} = \mathbf{1}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{t} \Big/ \left( \mathbf{1}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{1} \right)$$

McLean et al., 2011 (G^3)

# In general:

- Random uncertainties (analytical; random variability from analysis to analysis) can be decreased with more analyses. These appear only on the diagonal of the covariance matrix
- Systematic uncertainties are usually constant biases and/or biases that vary predictably. These contribute to both the on- and off-diagonal terms of the covariance matrix.

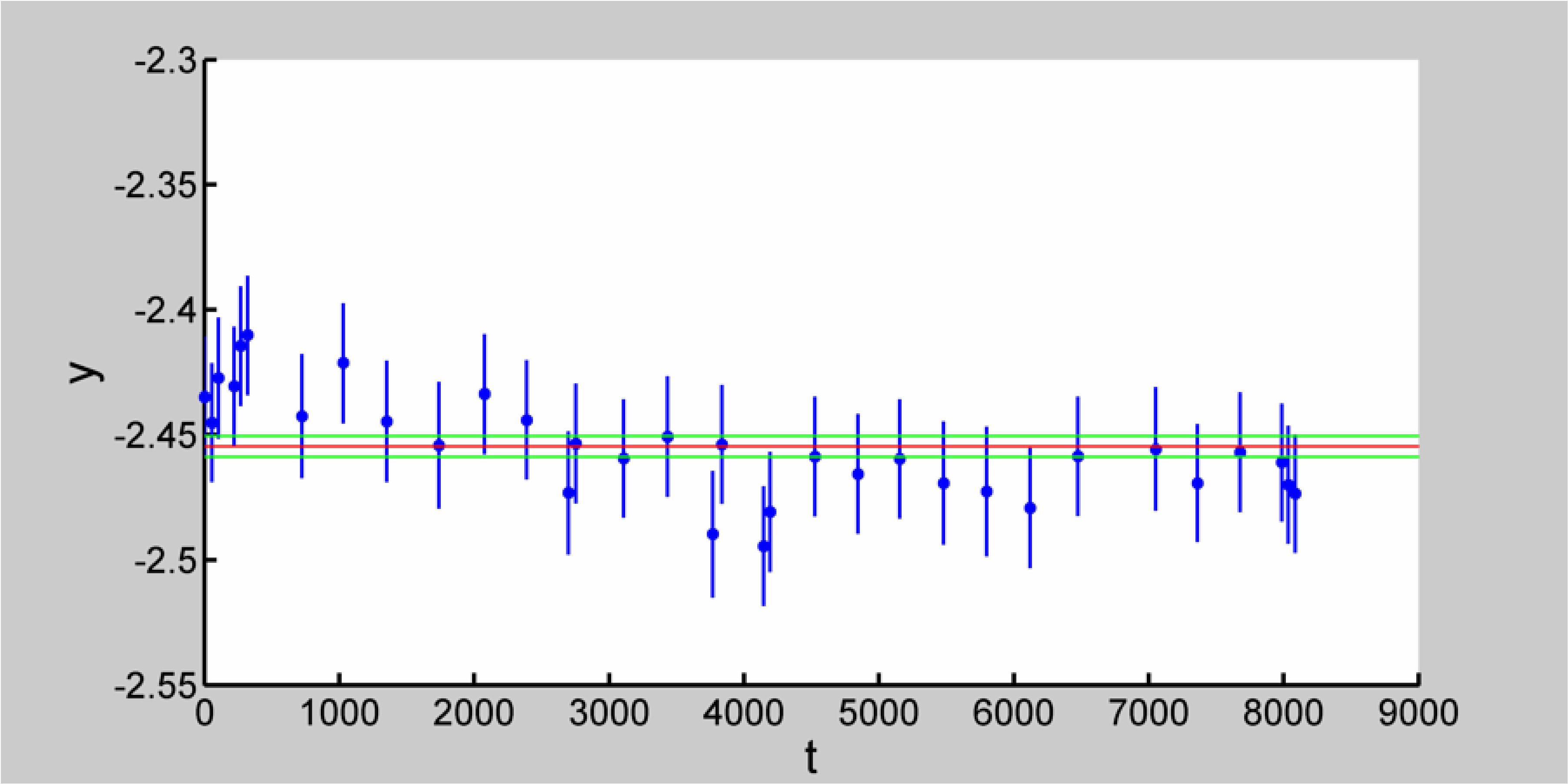
# Sample-Standard Bracketing

- Central problem:
  - When we measure the standard, we don't get the true value.
  - Some normalization factor is needed to make up the difference.
  - Often times in IRMS, this takes the form an equation like $X_{true} = A\ X_{meas}$, where *A* is the fudge factor that will transform a measured value to a true value, and X is an isotope ratio.
  - *A* might or might not be time-dependent.

# Sample-Standard Bracketing

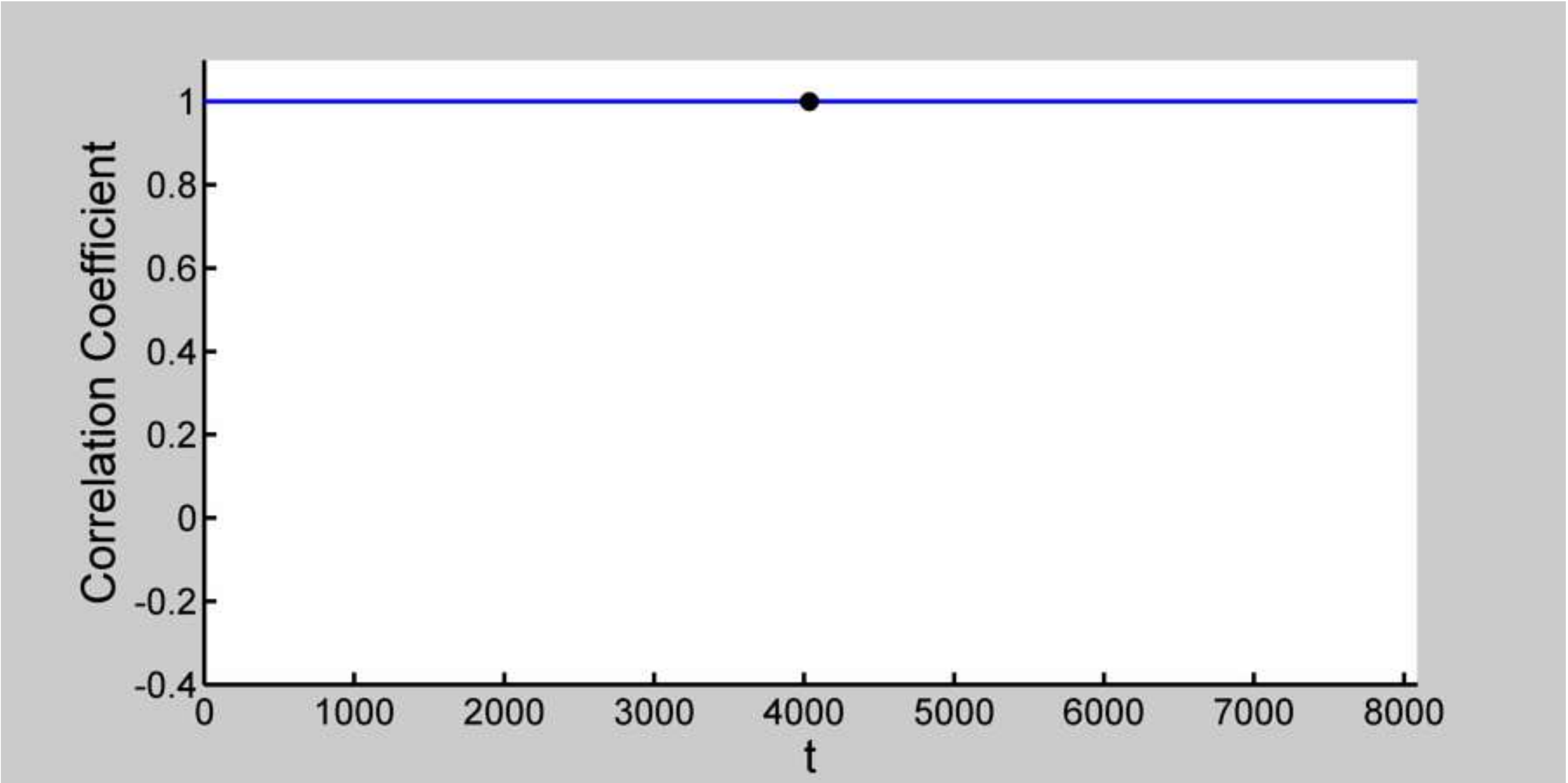
- **Solution:** periodically measure standards, calculate *A*, test for time-dependence. Fit some kind of function to the values of *A* through time, interpolate to find the predicted values for the unknowns, propagate uncertainties
- This is all straightforward… what's the catch?
  - The uncertainties in the interpolated unknowns are not independent—they are affected by the standards. The structure of this dependency is important, and can be determined analytically.

# For a mean:

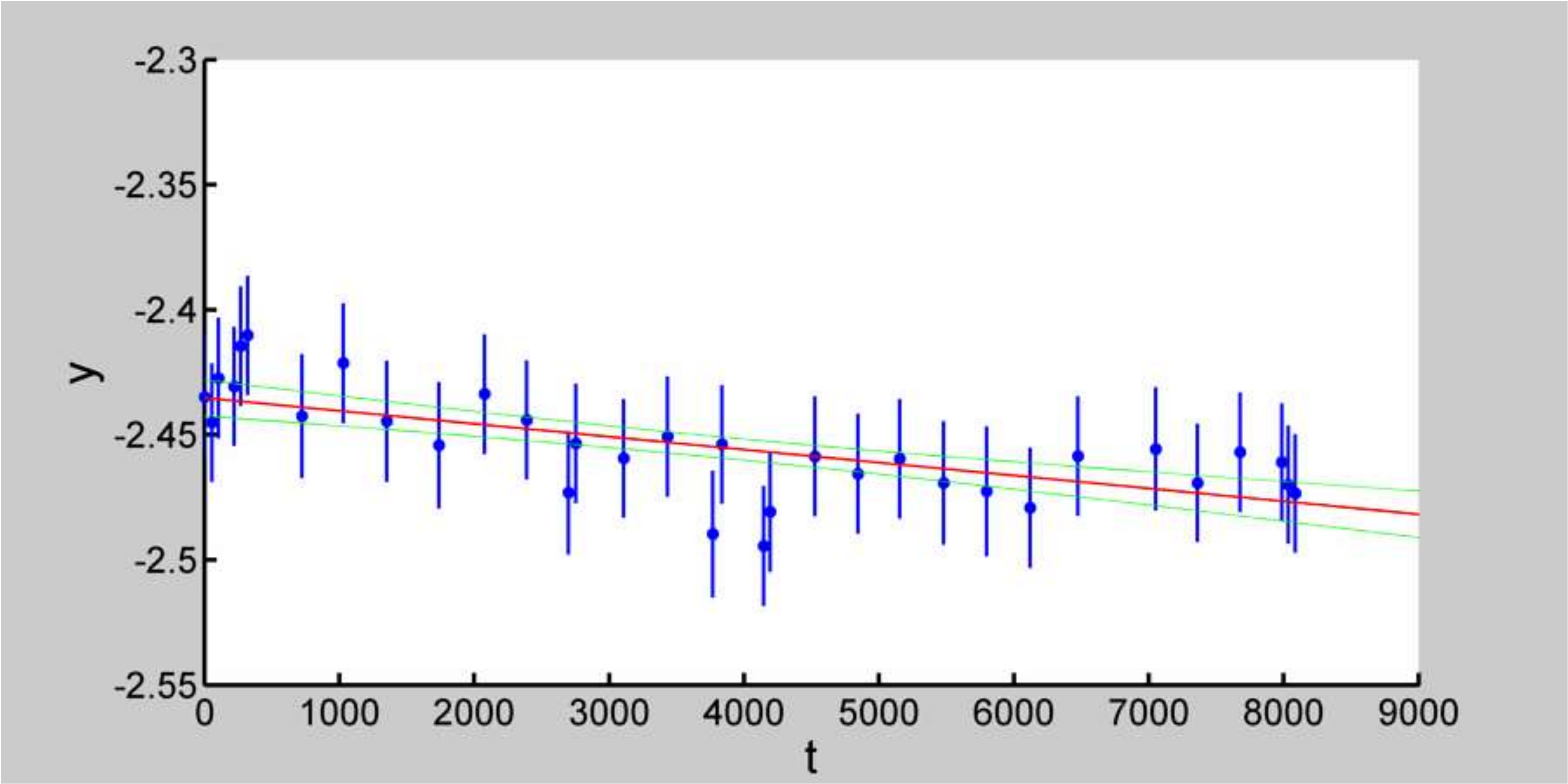


-2.3
-2.35
-2.4
-2.45
-2.5
-2.55
y
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
t

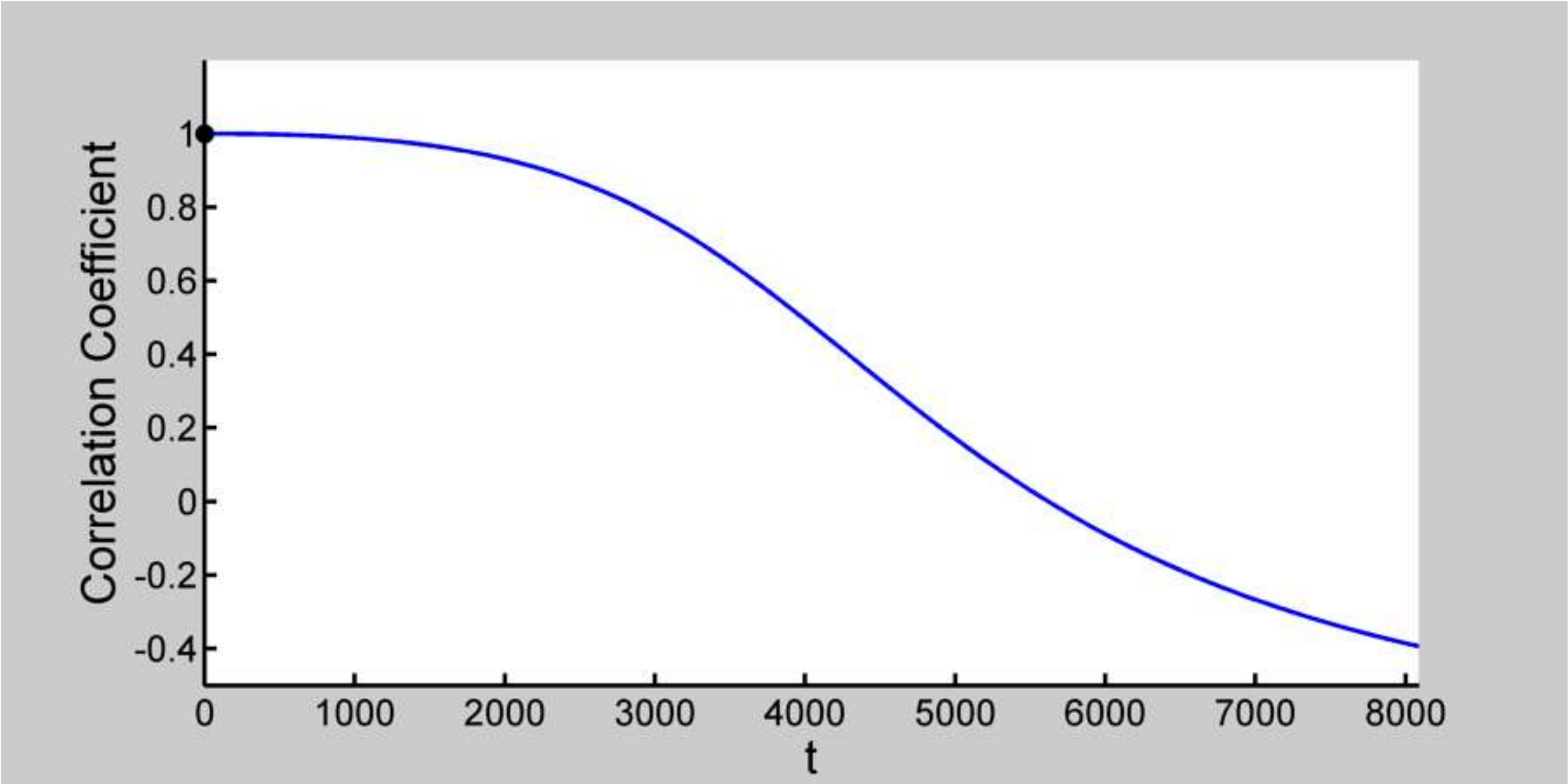
# Unknown anywhere



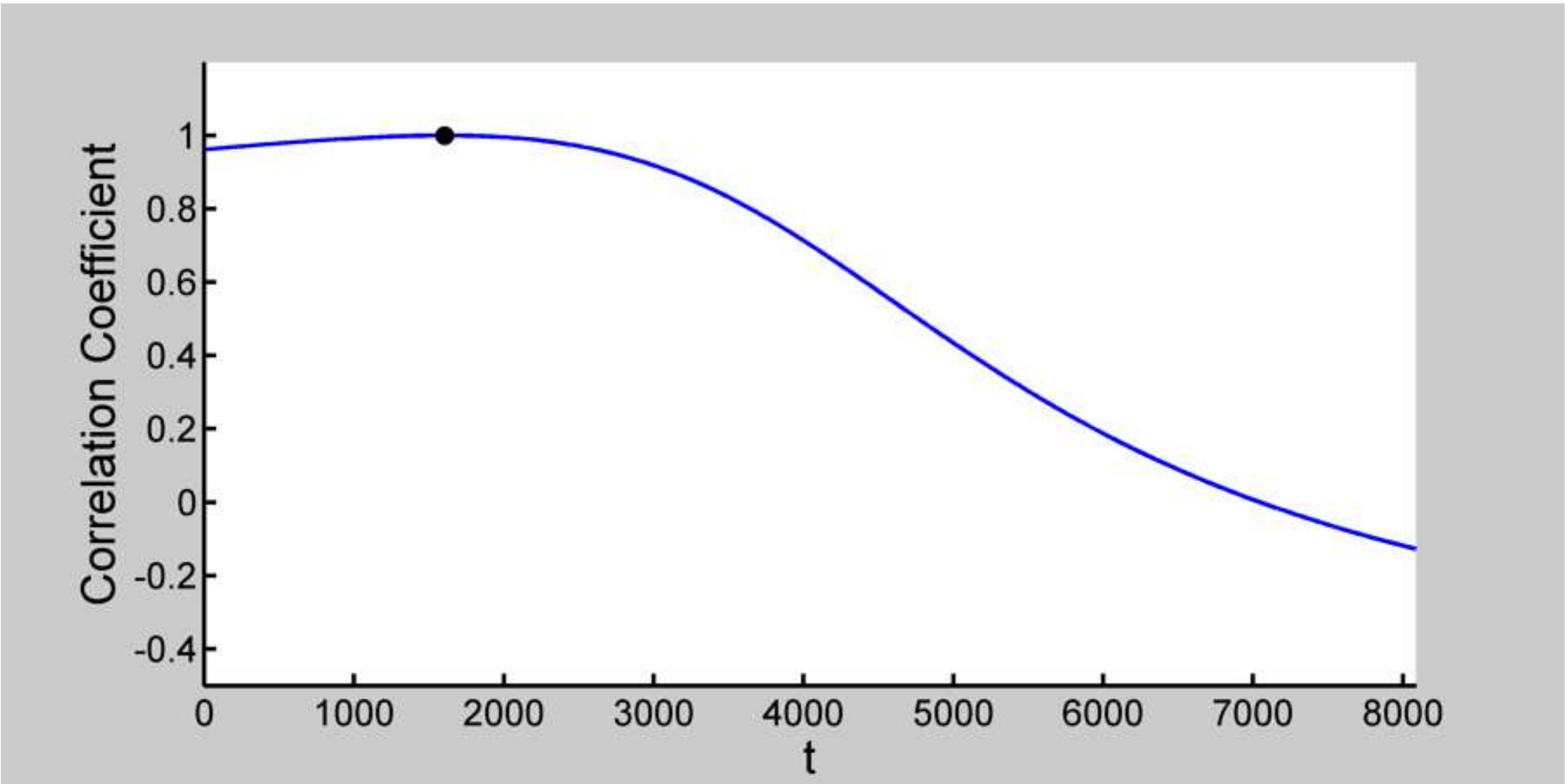
# For a line:


-2.3
-2.35
-2.4
-2.45
-2.5
-2.55
y
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
t

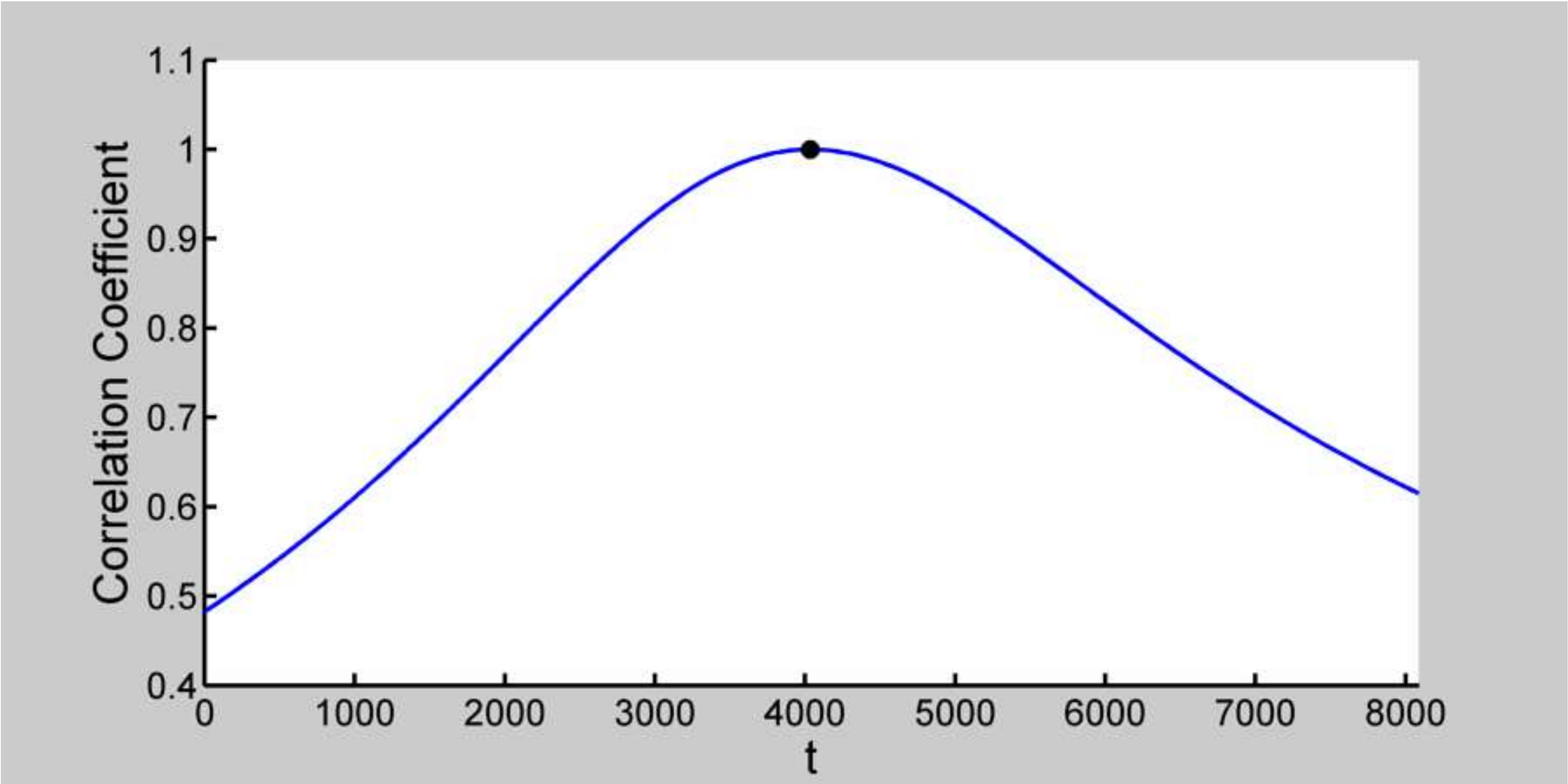
# Unknown at beginning



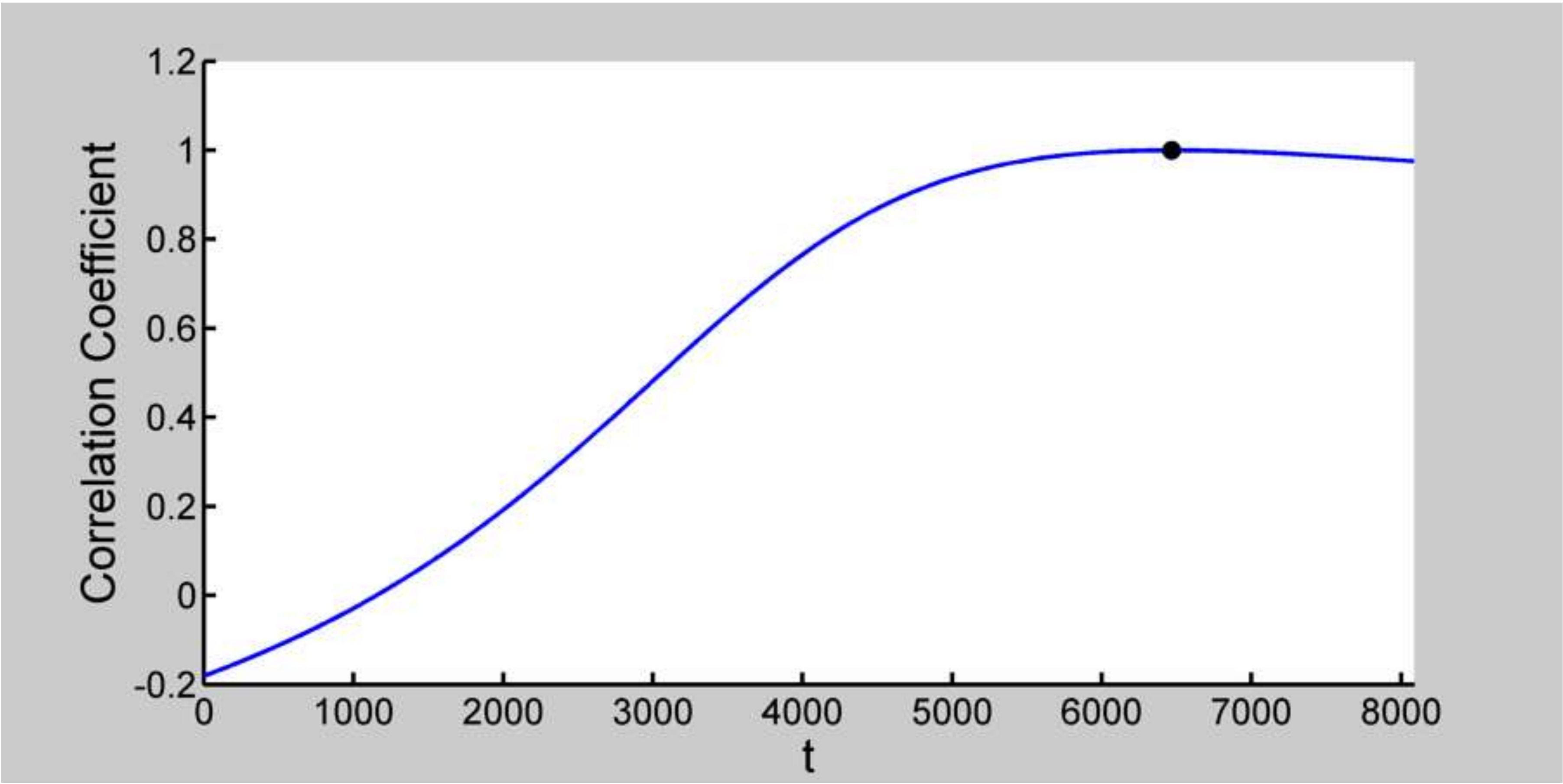
# Unknown near beginning



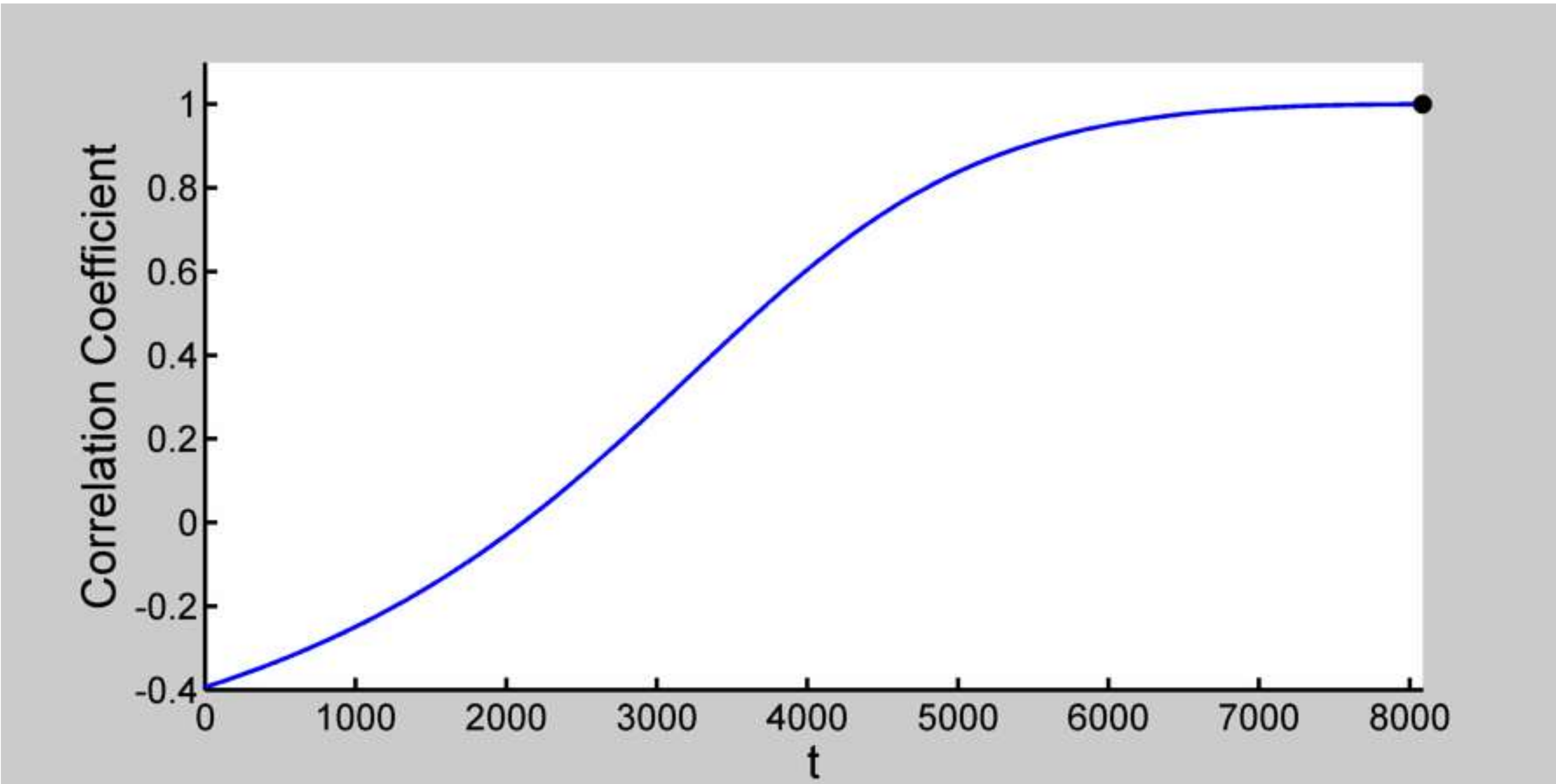
# Unknown at middle



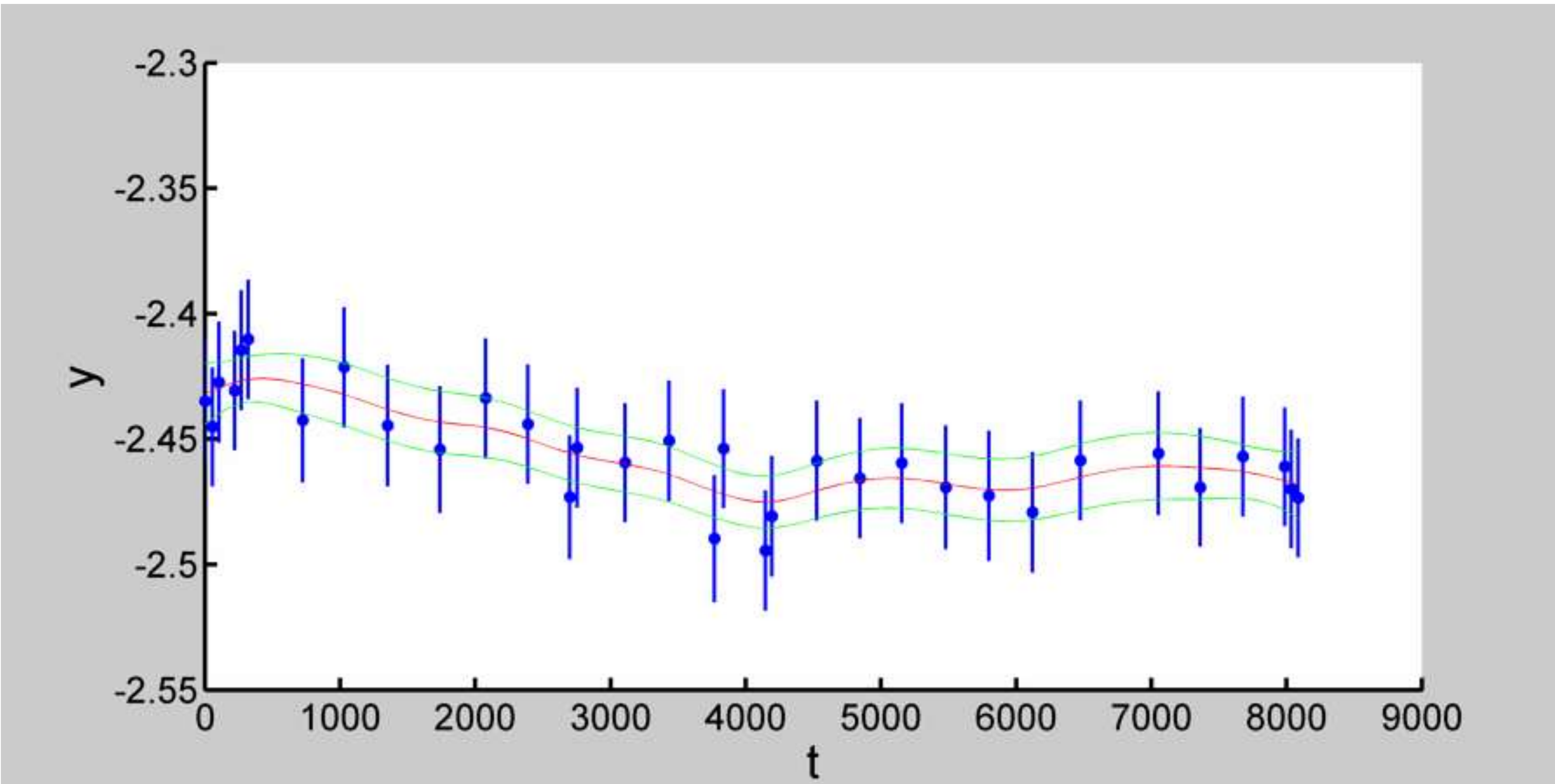
# Unknown near end



# Unknown at end

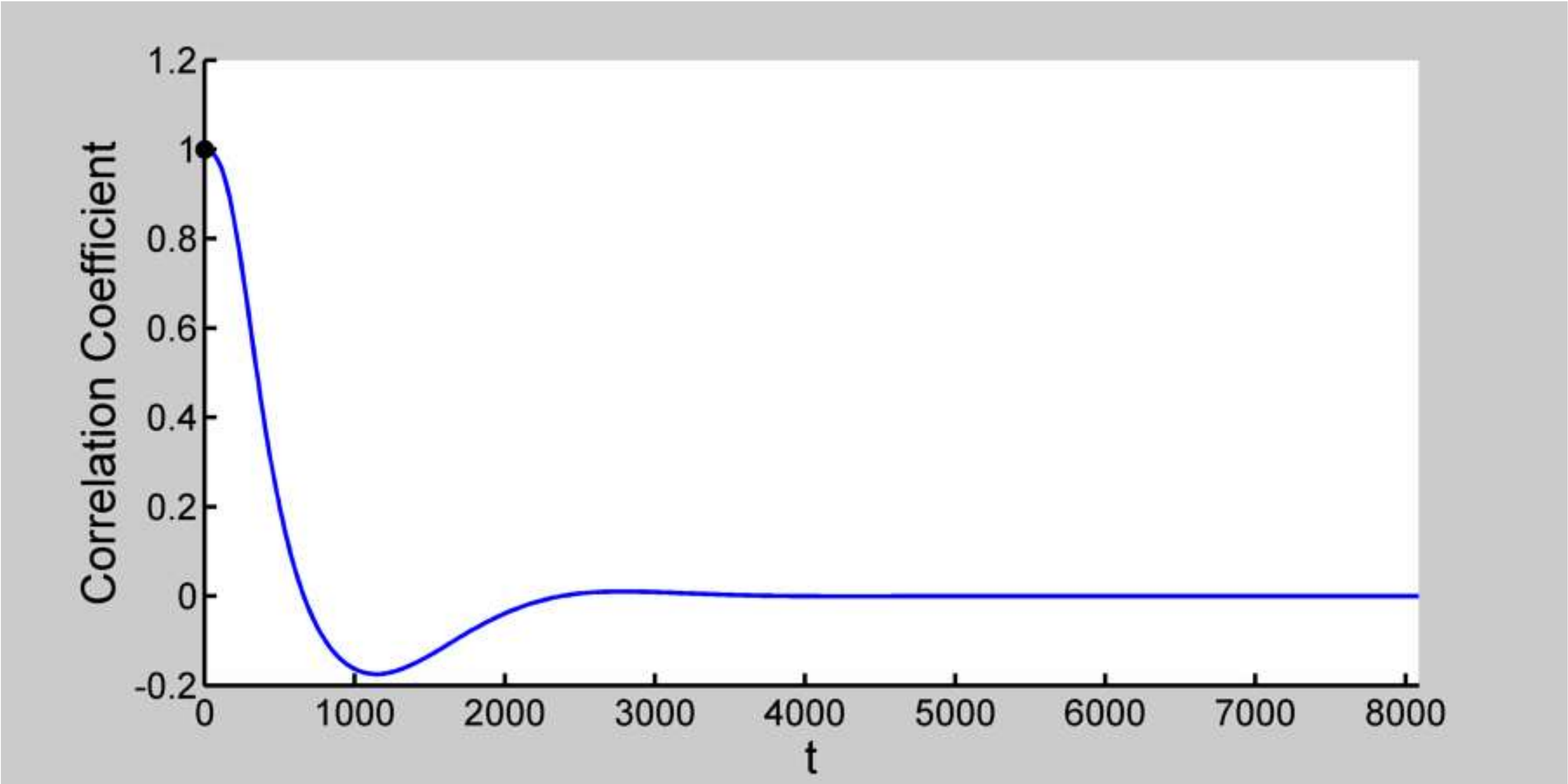


# For a spline:

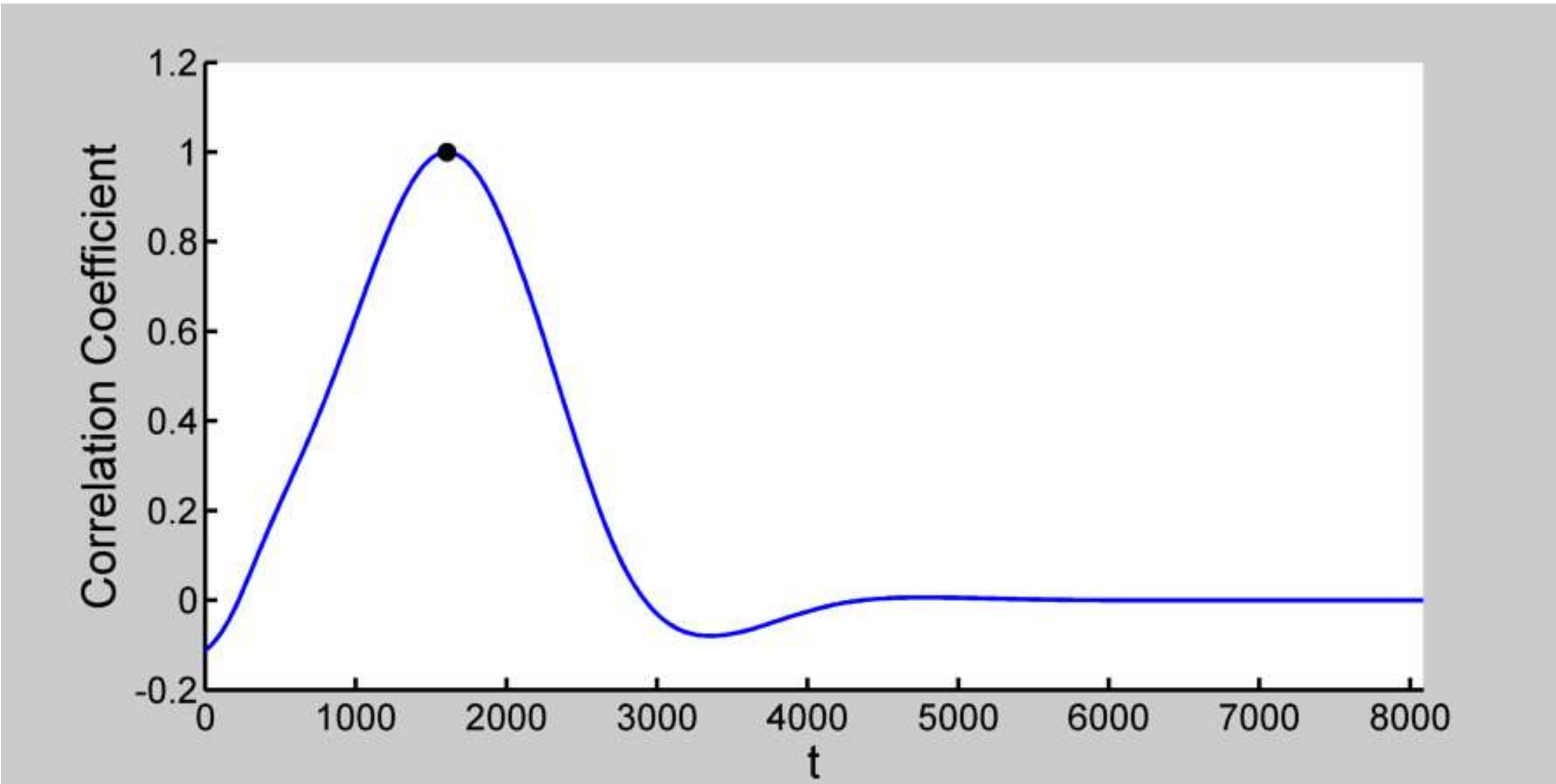


-2.3
-2.35
-2.4
-2.45
-2.5
-2.55
y
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
t

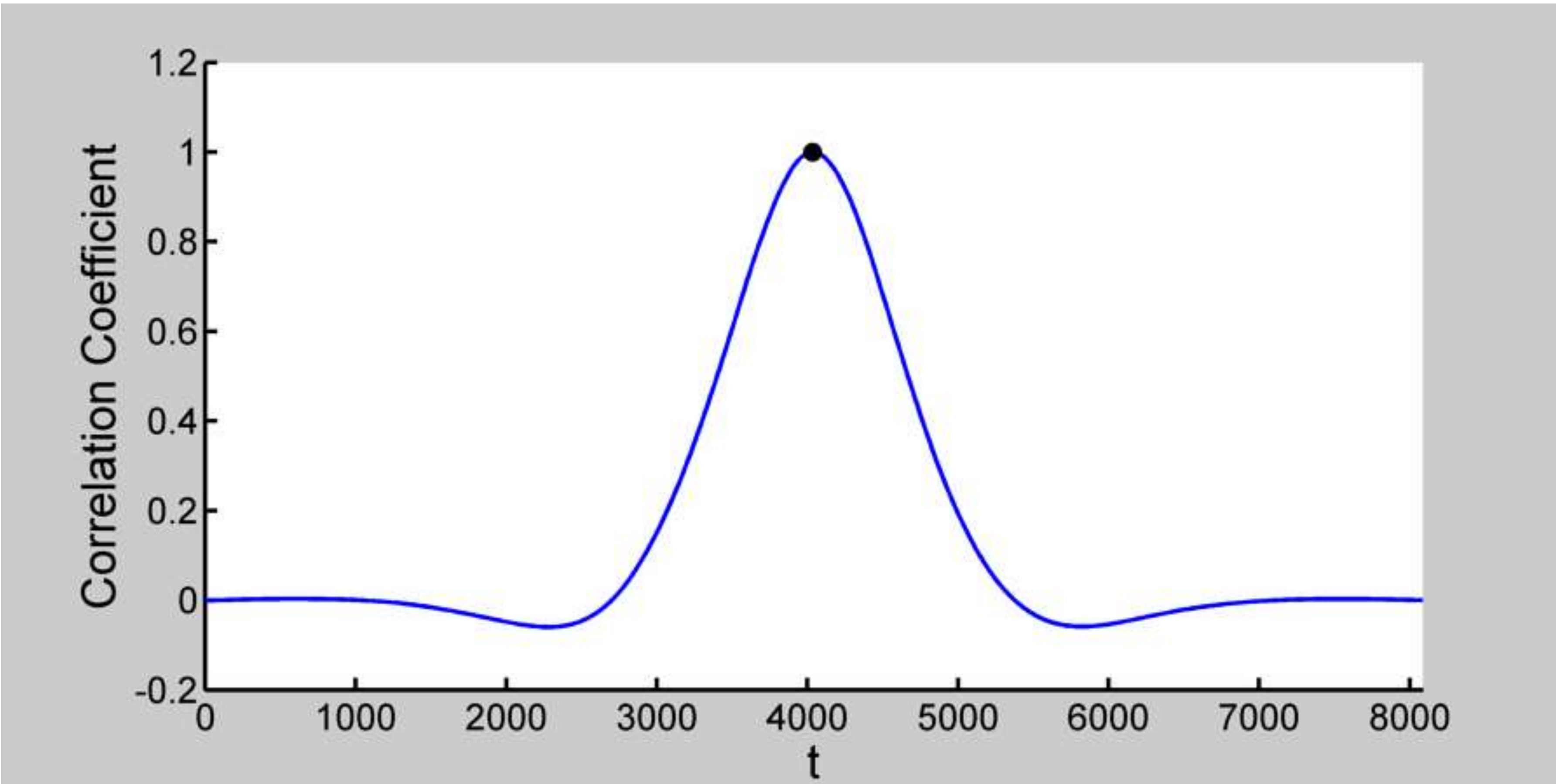
# Unknown at beginning



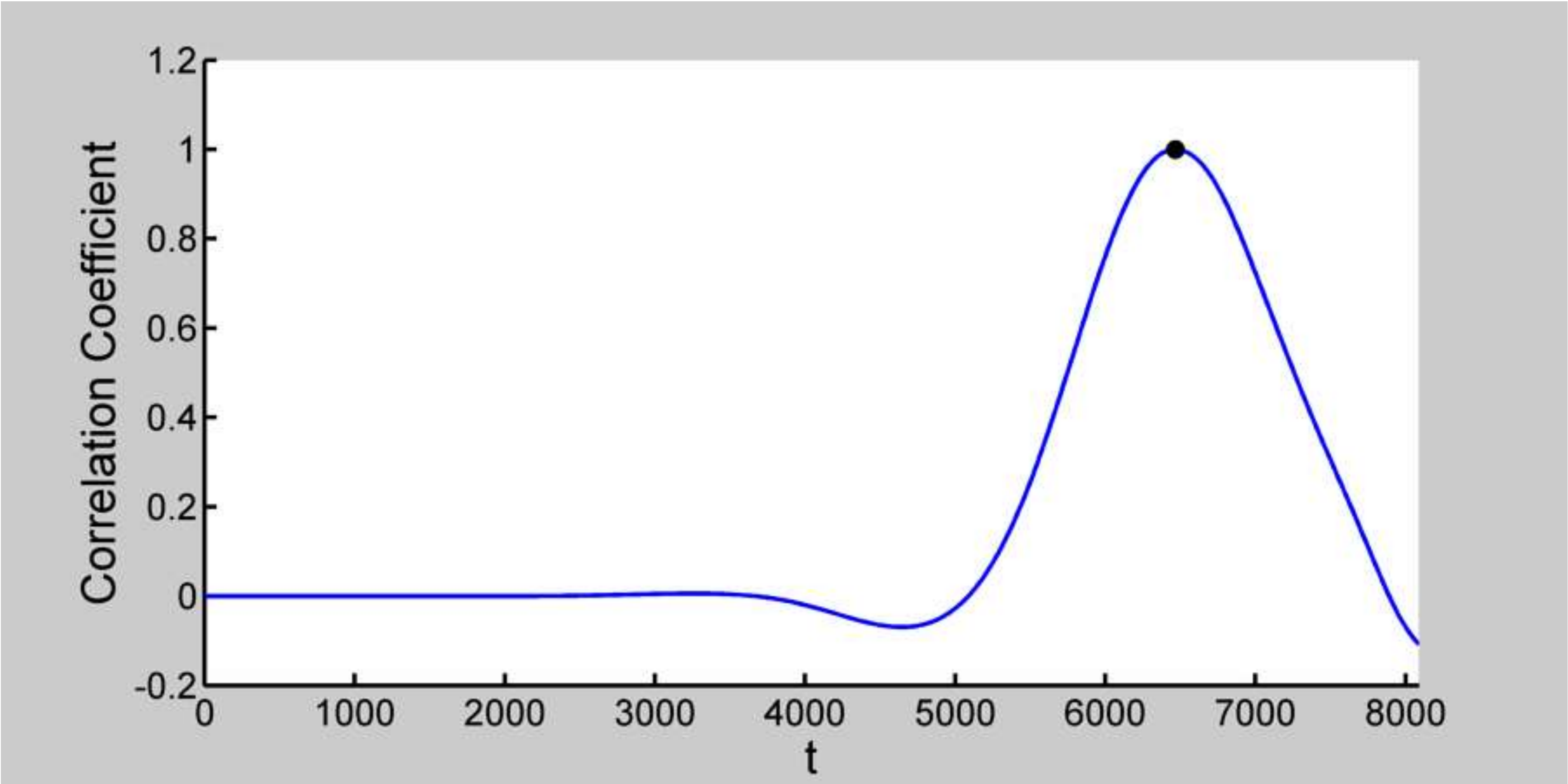
# Unknown near beginning



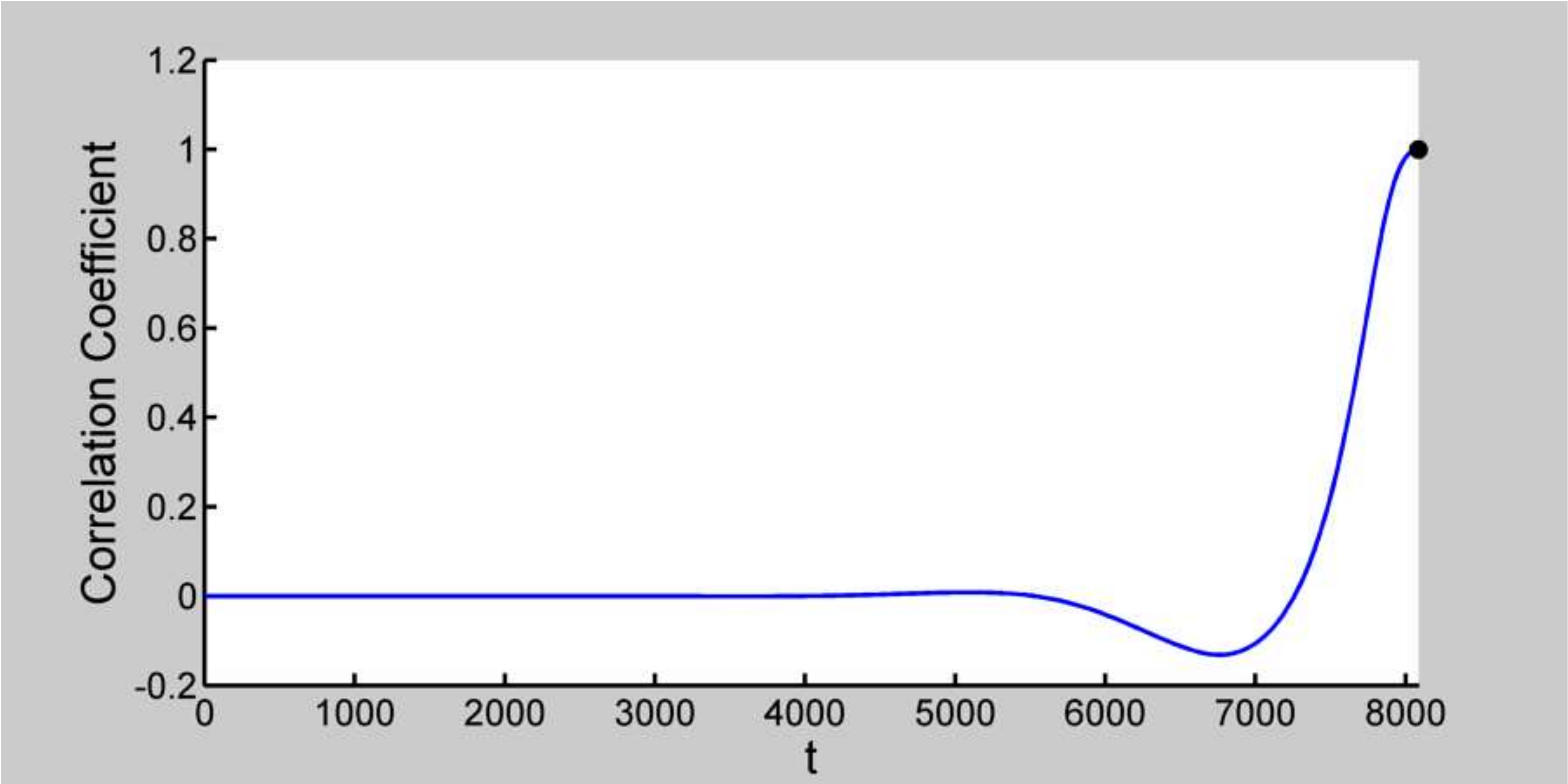
# Unknown at middle



# Unknown near end



# Unknown at end



# Conclusions

- Systematic uncertainties are uncertainty correlations
- Generalized weighted means have the ability to propagate random and systematic uncertainties correctly
- Sample-standard bracketing creates correlation between unknowns
- The degree of correlation depends on the fit