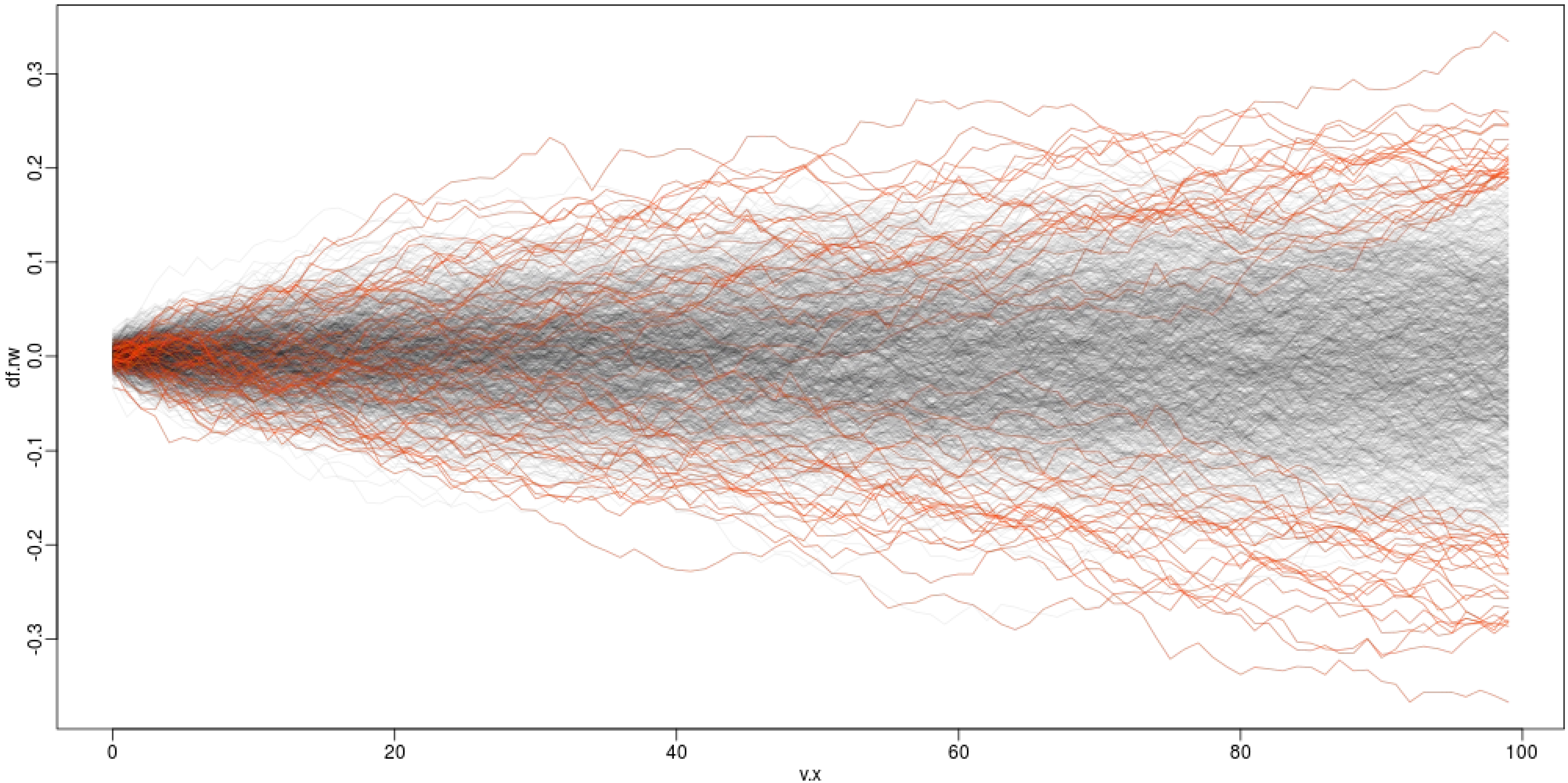
An introduction to time-series models for ecological data

久保拓弥 mailto:kubo@ees.hokudai.ac.jp



Do not apply simle GLMs

to time-series (TS) data!

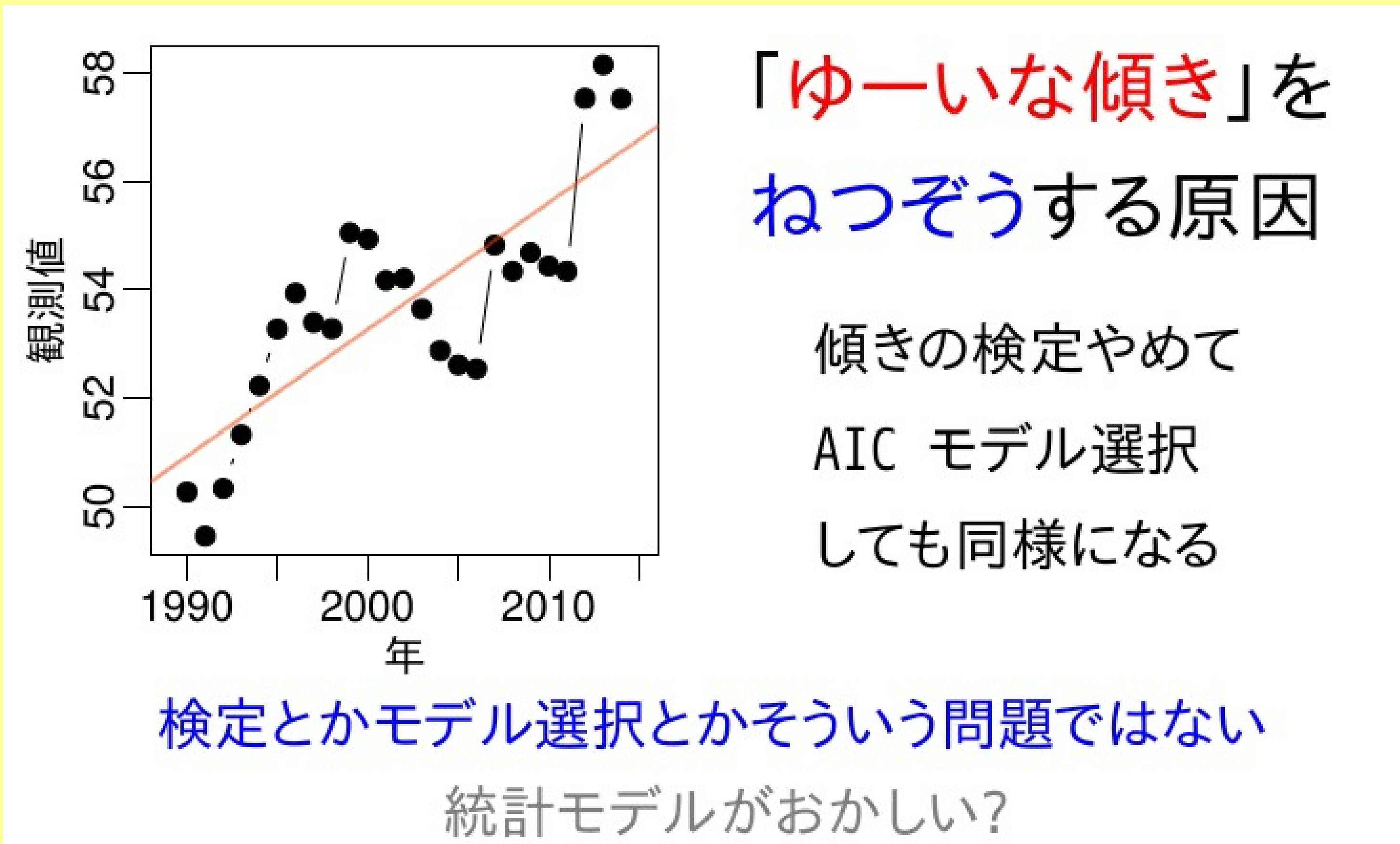
However, hierarchical Bayesian

models are

still effective to TS data

(Bad model 1) fit GLM to TS

NO!

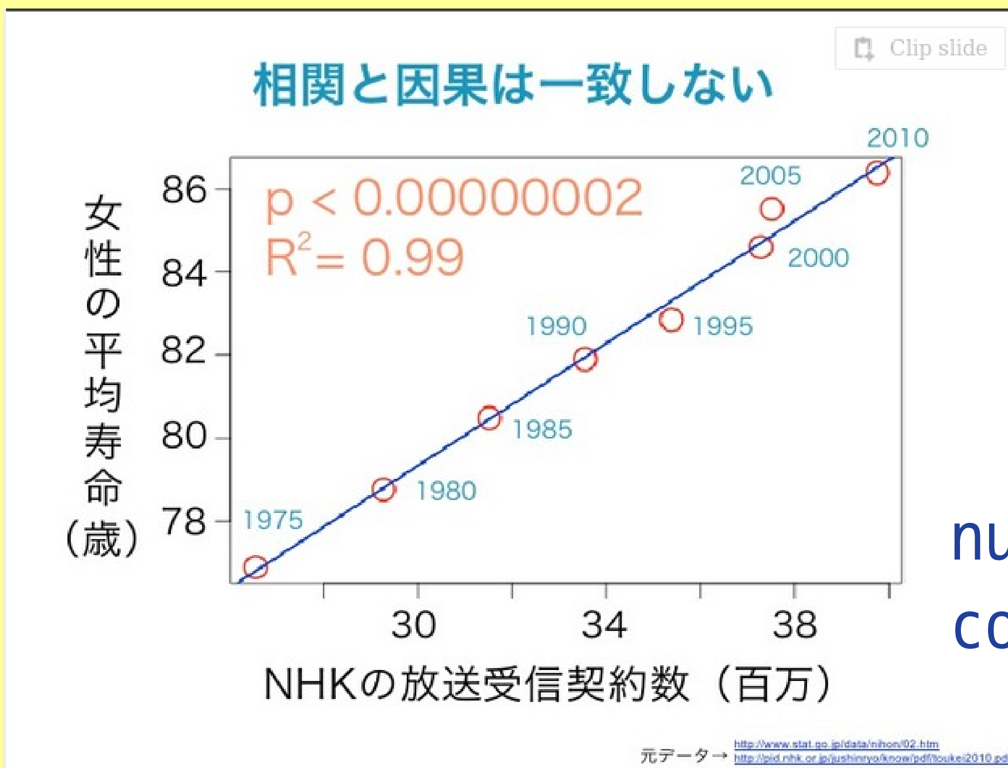


(Bad model 2) TS $Y_t \sim$ TS X_t

so called

“spurious regression”

women longevity



number of NHK contracts

http://www.slideshare.net/takehikoihayashi/ss-13441401

Statistical modeling for time series (TS) data

- Do NOT fit GLM to TS data
- A basic component: Random walk (RW) model
- RW + GLM → State Space Model (SSM)

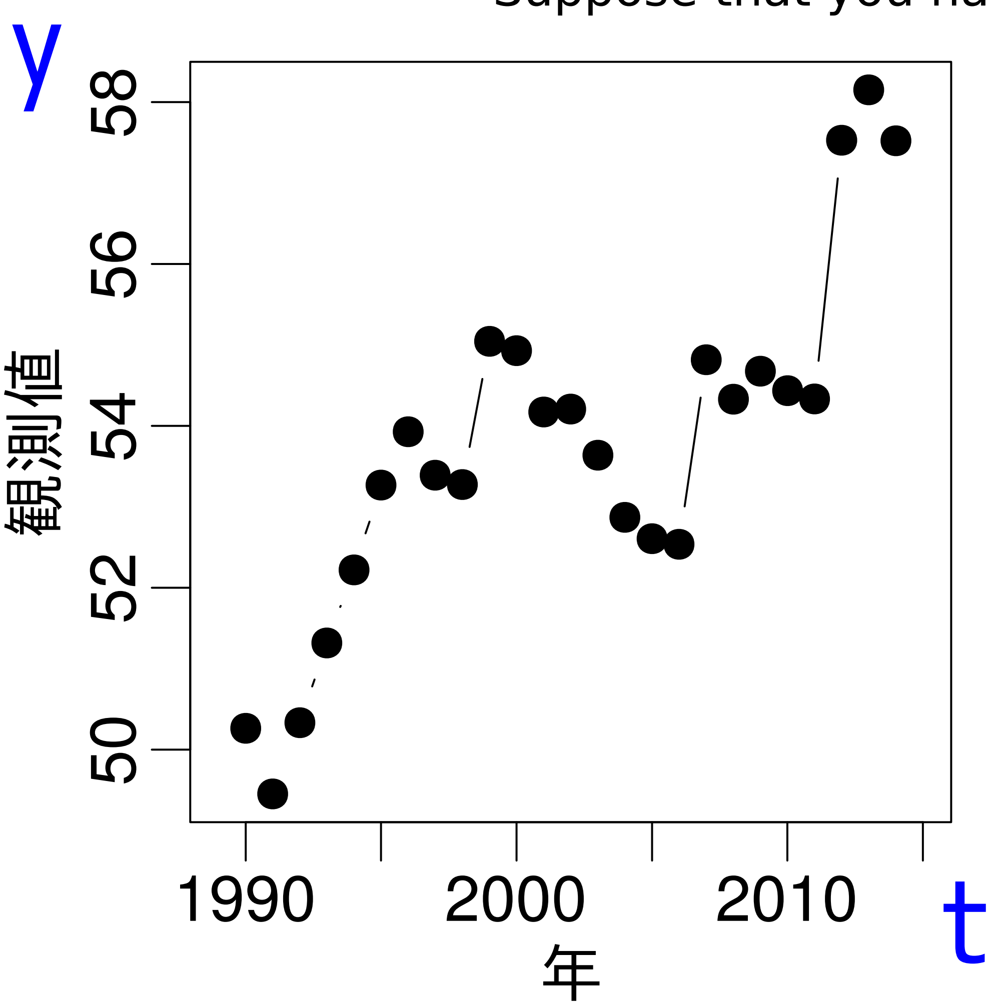
状態空間モデル

simple GLM: A bad model

for time-series (TS) data

このような時系列データがあったとしましょう

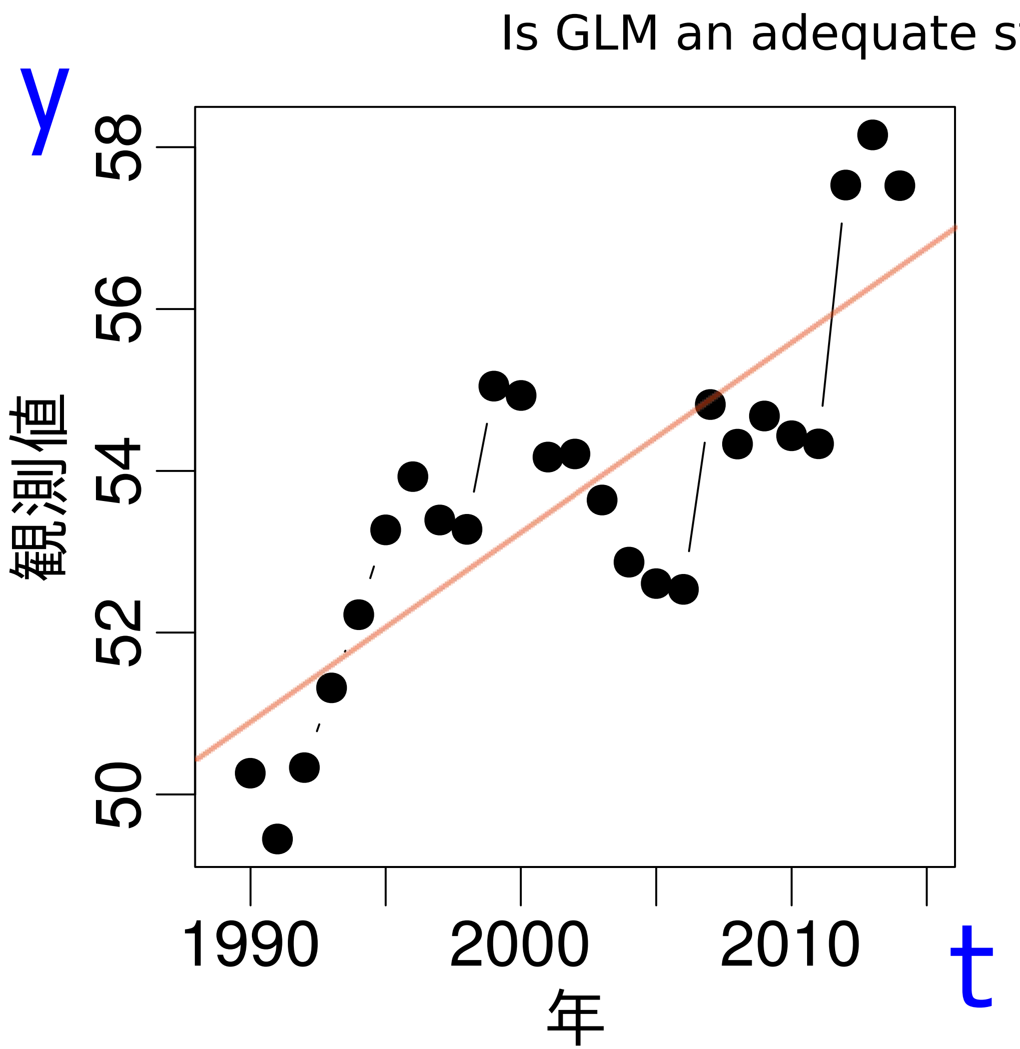
Suppose that you have a time-series data



y は何か連続値としましょう

y: some continuous measurements

時系列データの統計モデリング入門

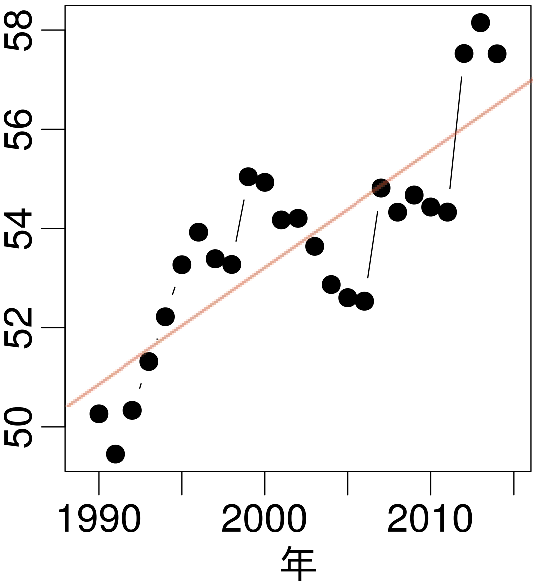


glm(y ~ t)

…とモデル
をあてはめてみた

「やったーゆーいだ!!」……??

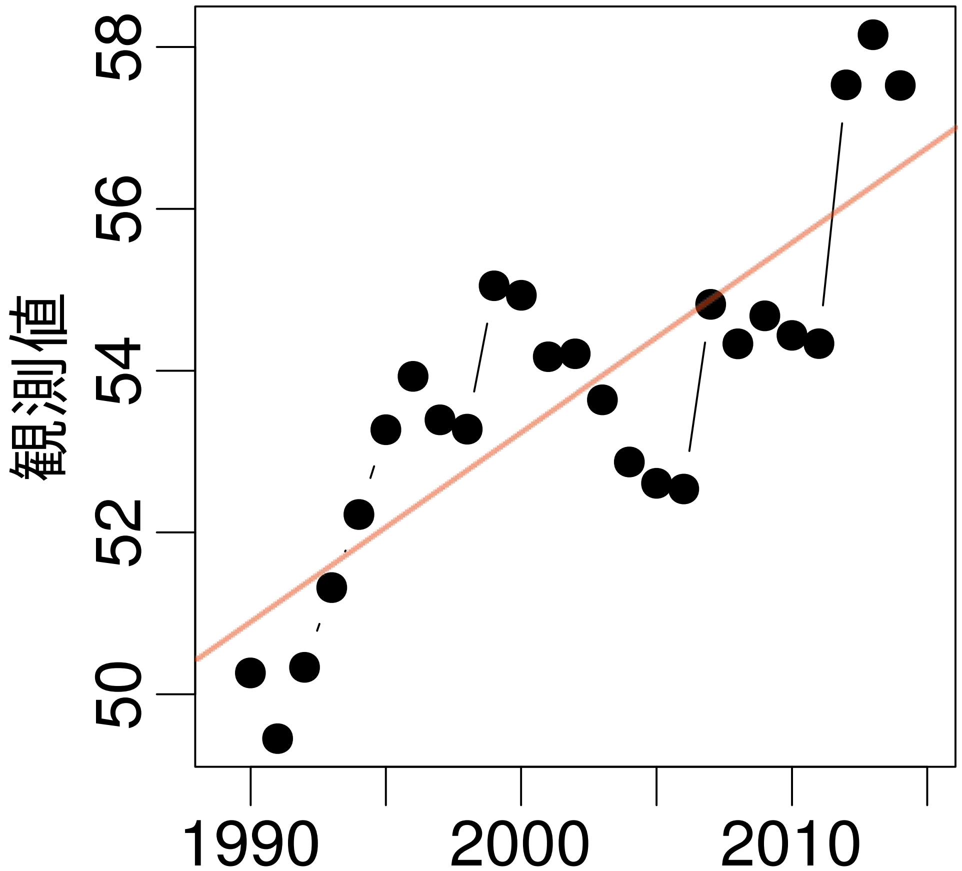
Significant ??? No!




```
> summary(glm(formula = y ~ t))

Deviance Residuals:
    Min        1Q    Median        3Q       Max
-2.1295   -1.0583   -0.0817    0.9860    2.0188

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -414.5655    71.4761   -5.80  6.6e-06
t              0.2339     0.0357    6.55  1.1e-06
```

A bad modeling: glm(時系列Y ~ 時間 t)

時系列の各点は独立ではない



each data points are

affective by previous

data point!

NOT independents!

「ゆーい」な傾き

a fake significancy!

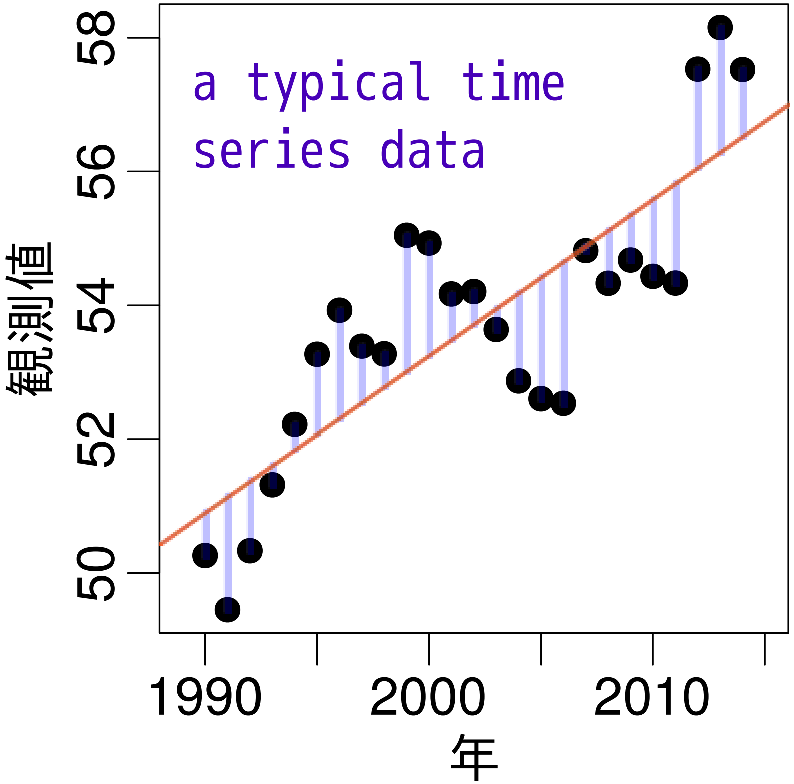
検定とかモデル選択とかそういう問題ではない

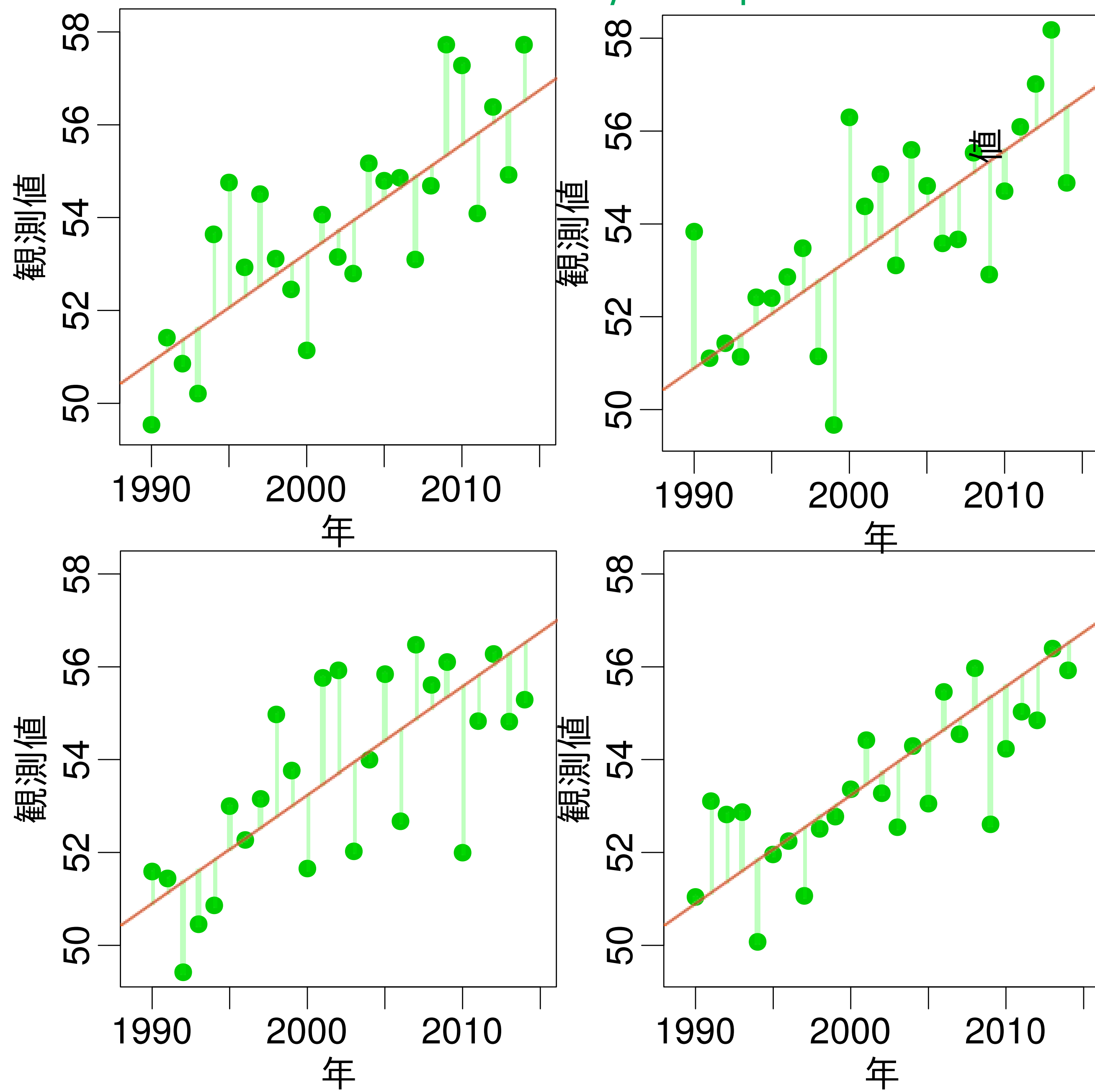
統計モデルがおかしい?

時系列の「ずれ」

GLM のずれ

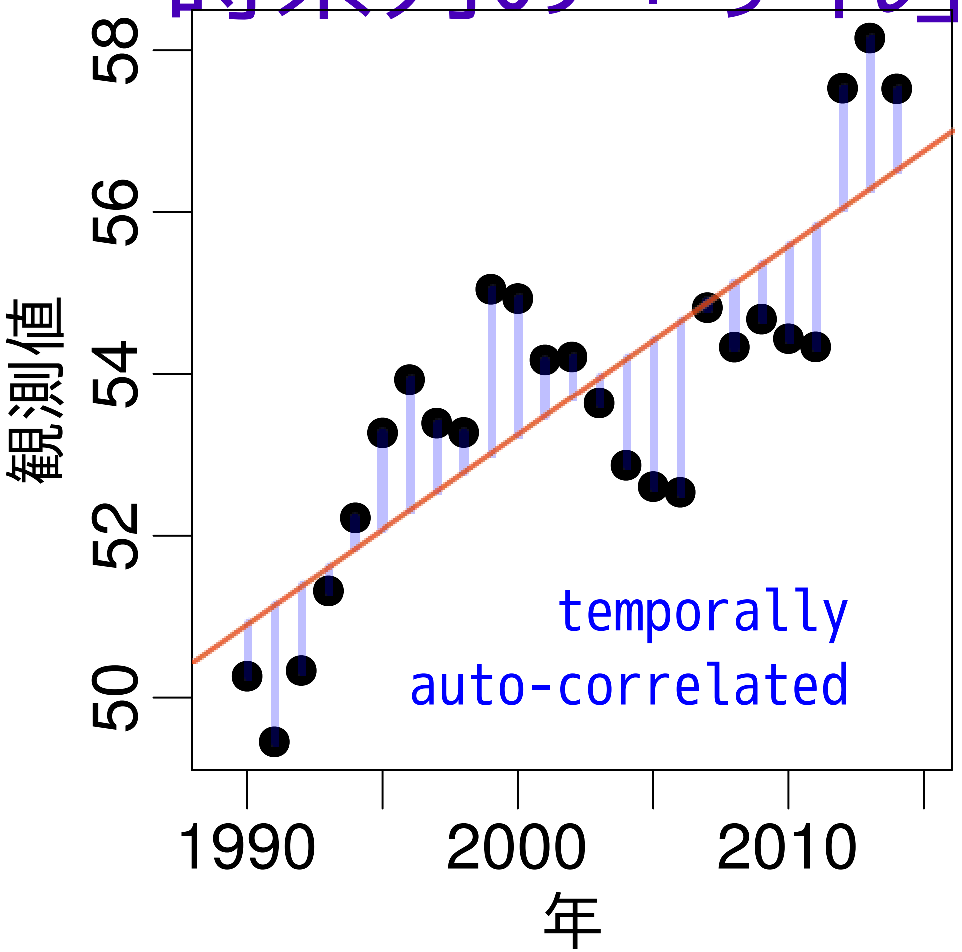
conditionally independent

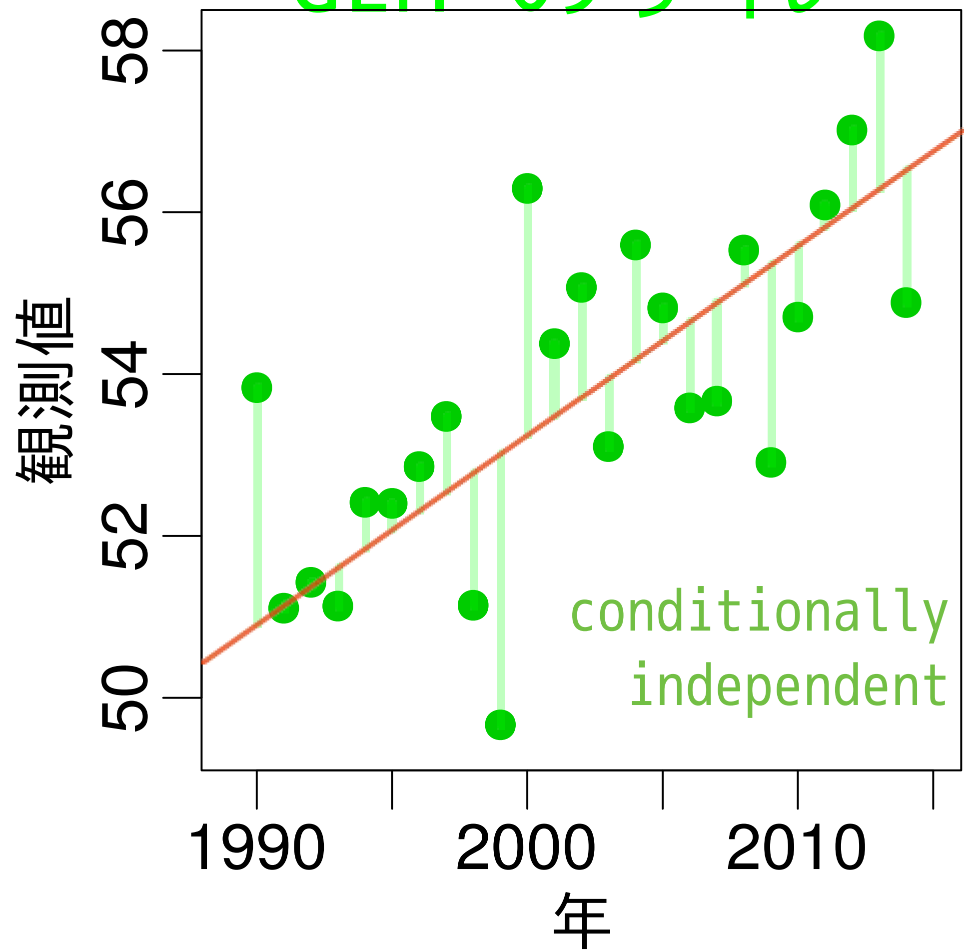




can you see
the differences?

時系列の「ずれ」　　GLM のずれ

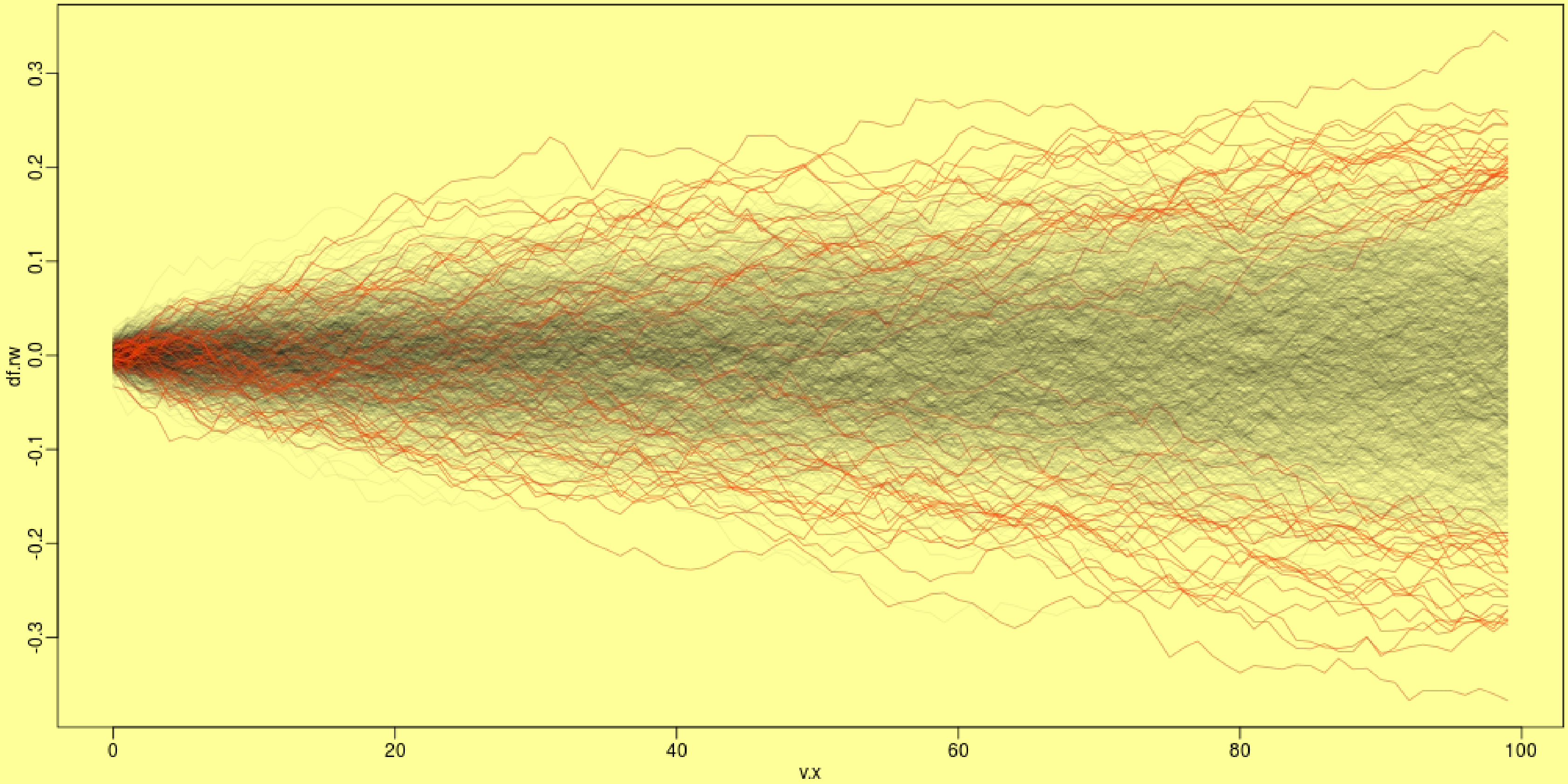




See time auto-correlation!

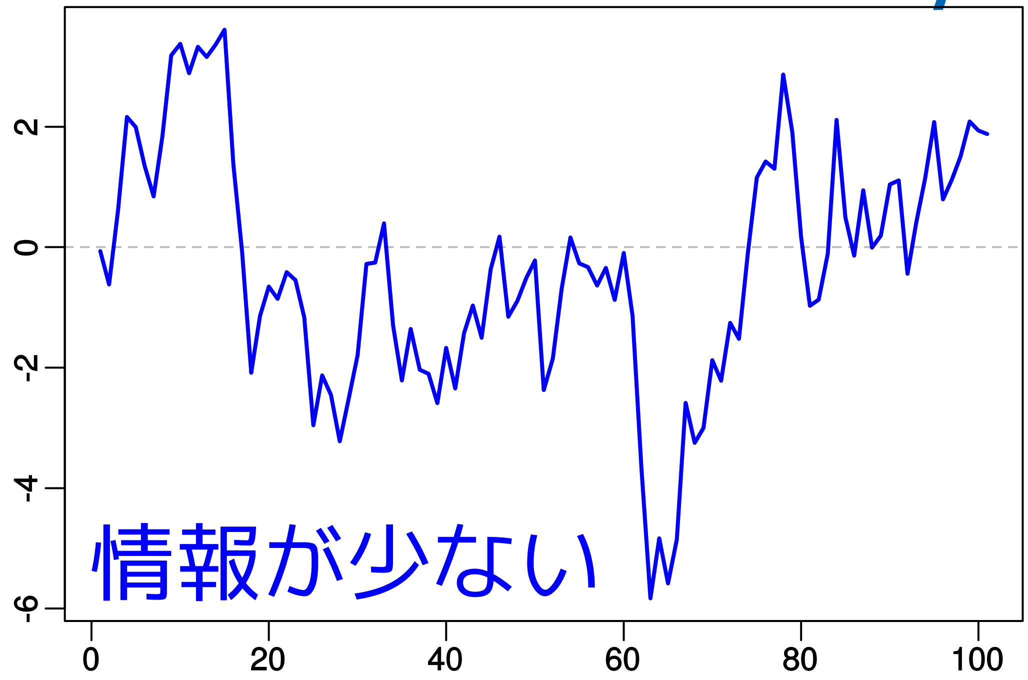
時間的自己相関がある　　時間的自己相関がない

Time series model as Random walk (乱歩)

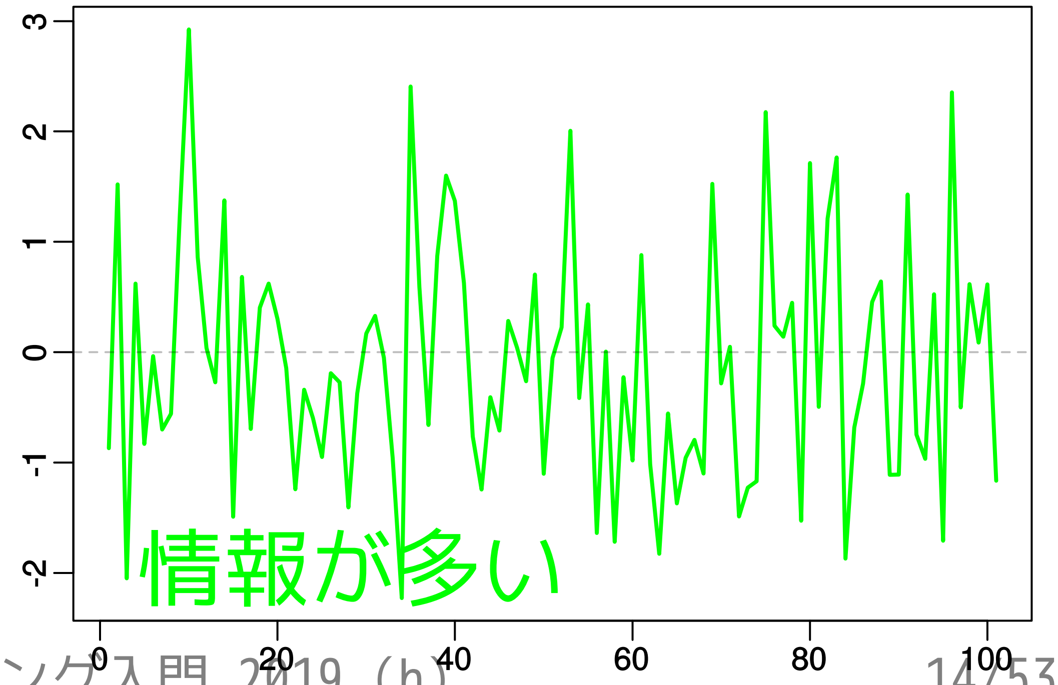


non-stationary

Depending on previous data point



Independent at each time



temporal auto-correlation

function (ACF)

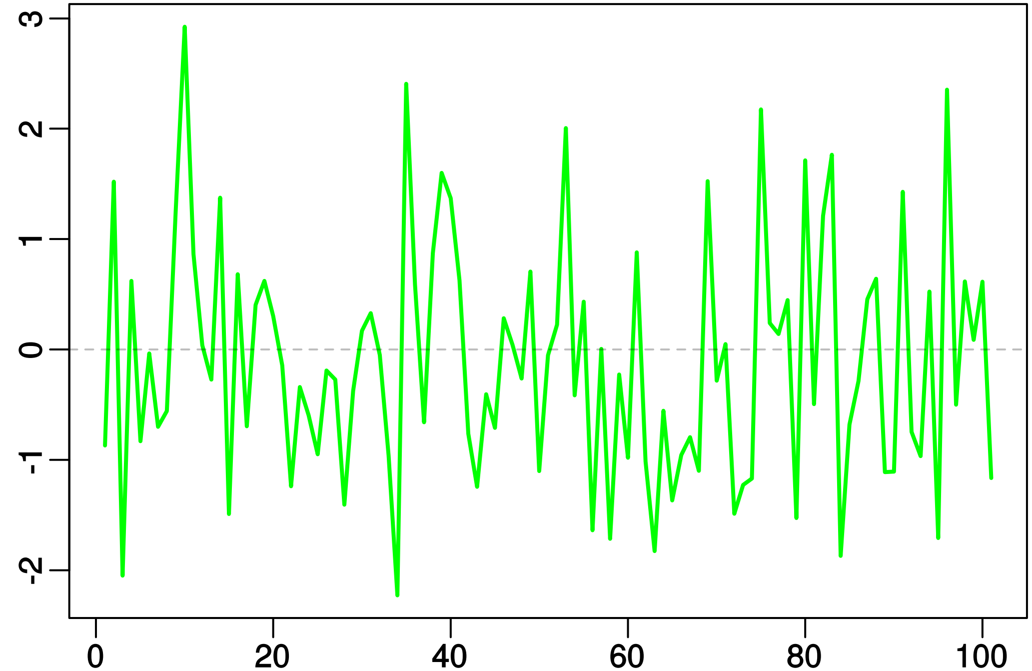
(略称:自己相関，時間相関)

を調べたらいいの?

$$\rho_k = \frac{\mathrm{Cov}(y_t, y_{t-k})}{\sqrt{\mathrm{Var}(y_t) \cdot \mathrm{Var}(y_{t-1})}}$$

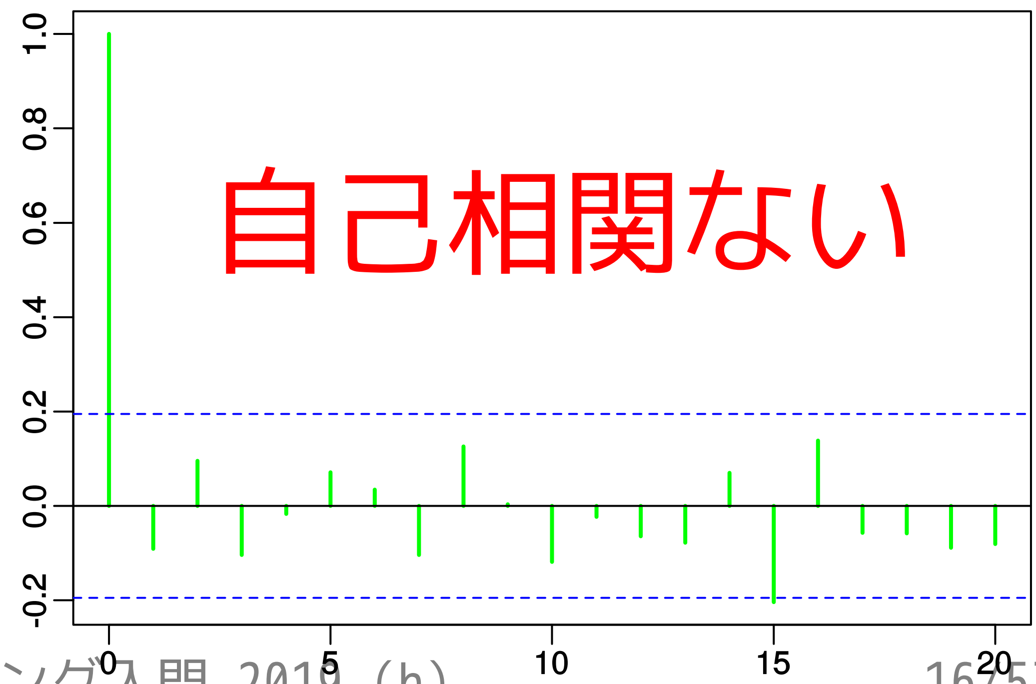
R の ts クラス: 時系列をあつかう

plot(ts(Y))

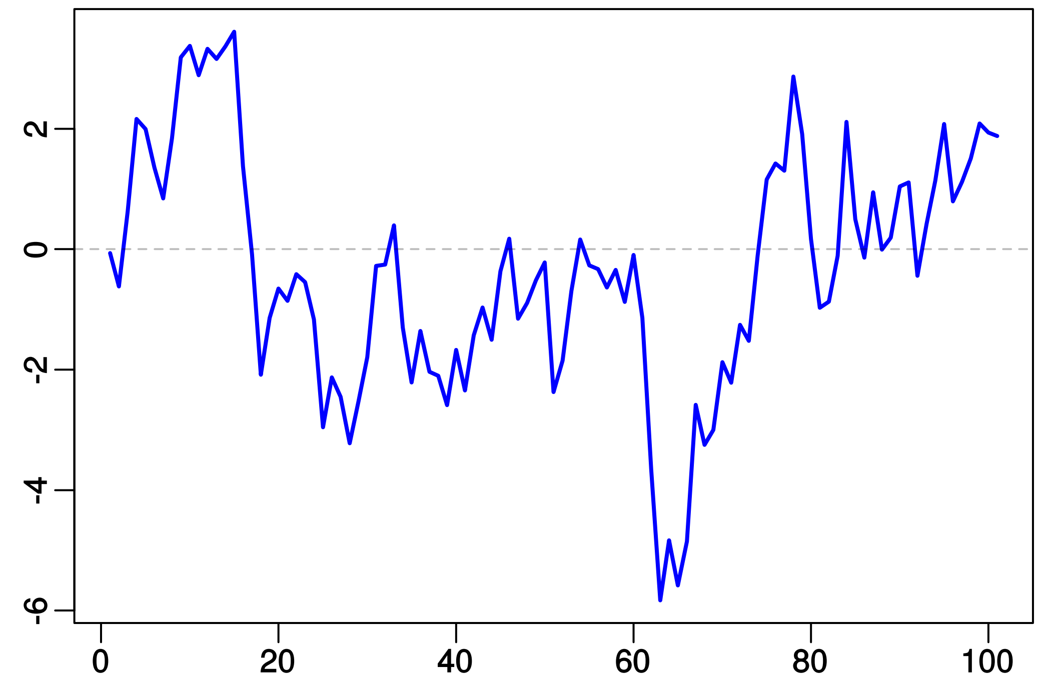


これはたんなる
100 個の正規乱数

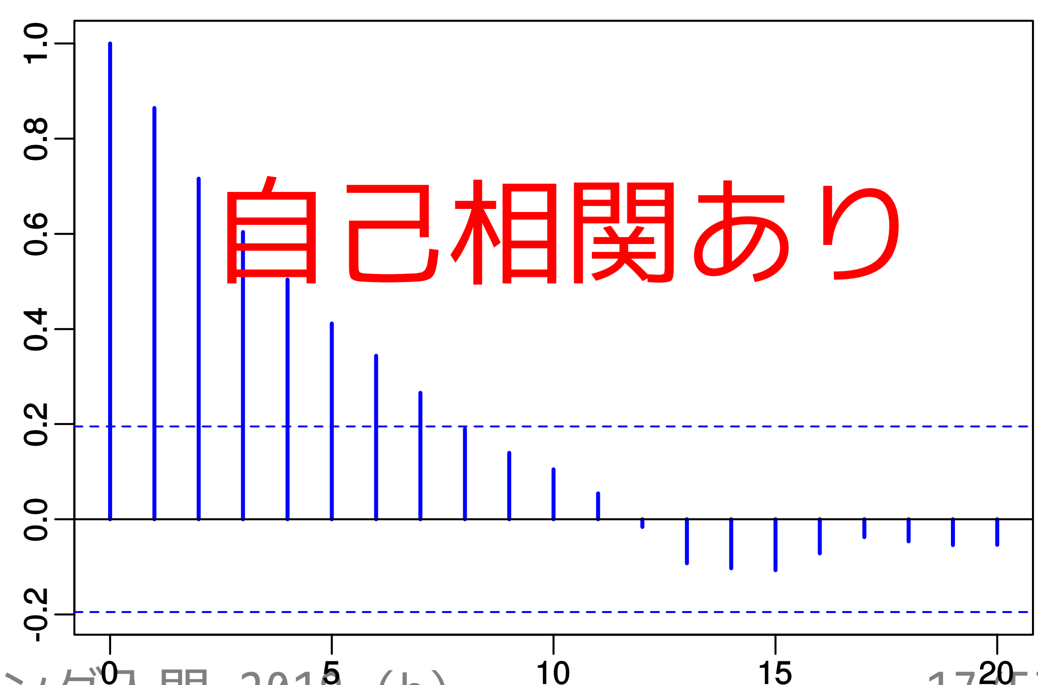
plot(acf(ts(Y)))



`plot(ts(Y))`



`plot(acf(ts(Y)))`



自己相関減衰

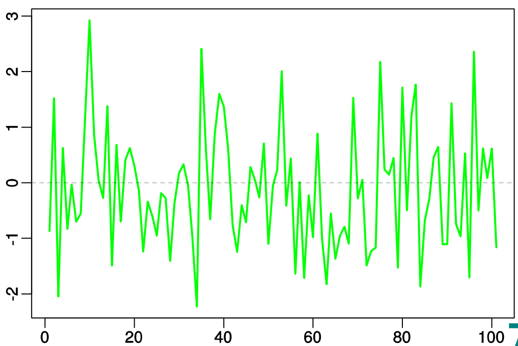
時間的自己相関

いつも役にたつわけではない?

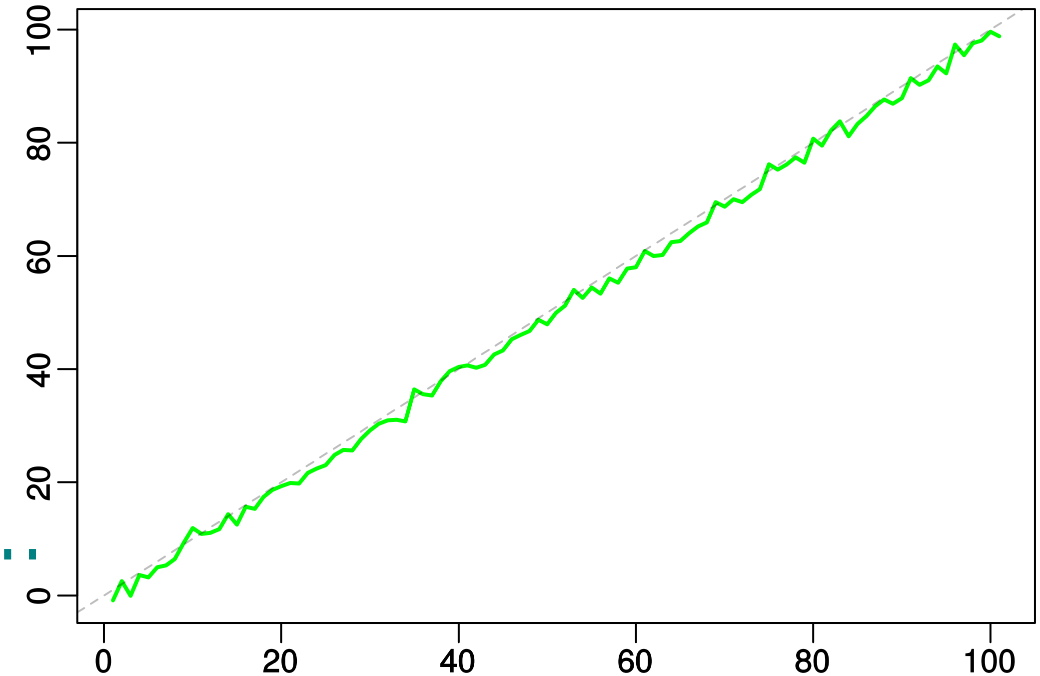
$$\rho_k = \frac{\mathrm{Cov}(y_t, y_{t-k})}{\sqrt{\mathrm{Var}(y_t) \cdot \mathrm{Var}(y_{t-1})}}$$

各点独立のデータをナナメにすると?

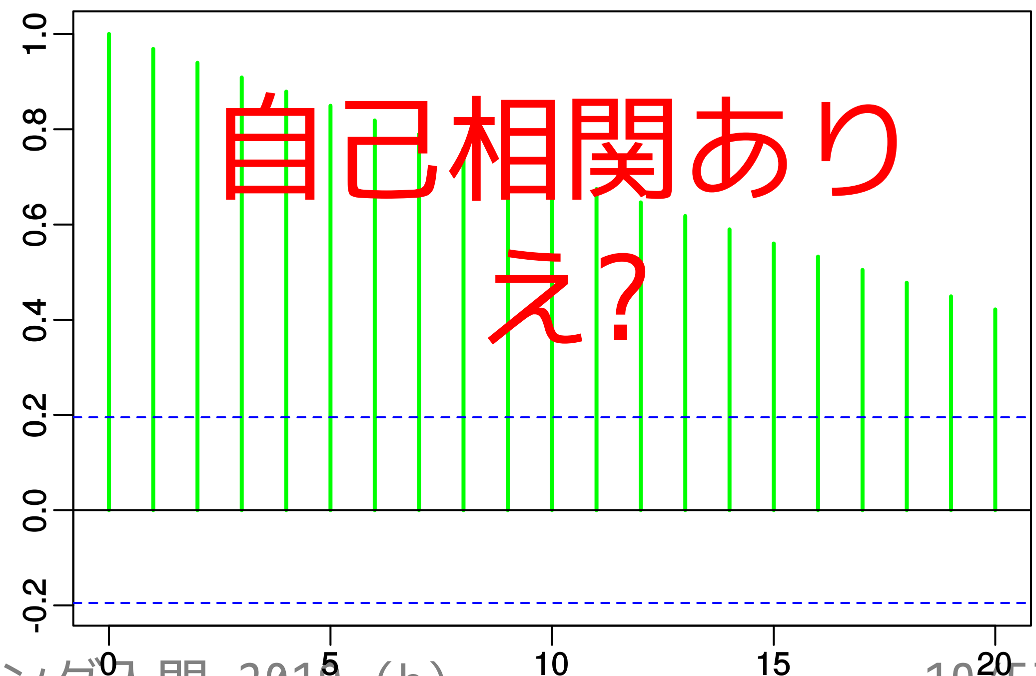
`plot(ts(Y))`



これを
ナナメに
したもの
なんだけど…



`plot(acf(ts(Y)))`



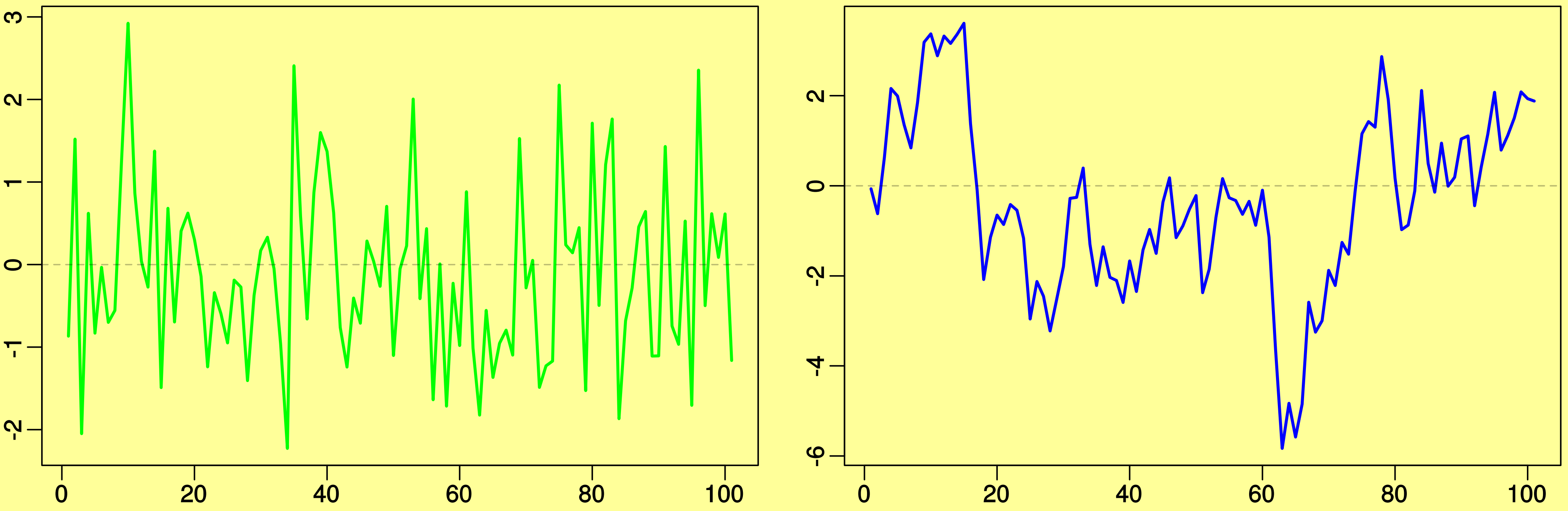
自己相関あり
え?

状態空間モデルでたちむかう

時系列データ解析

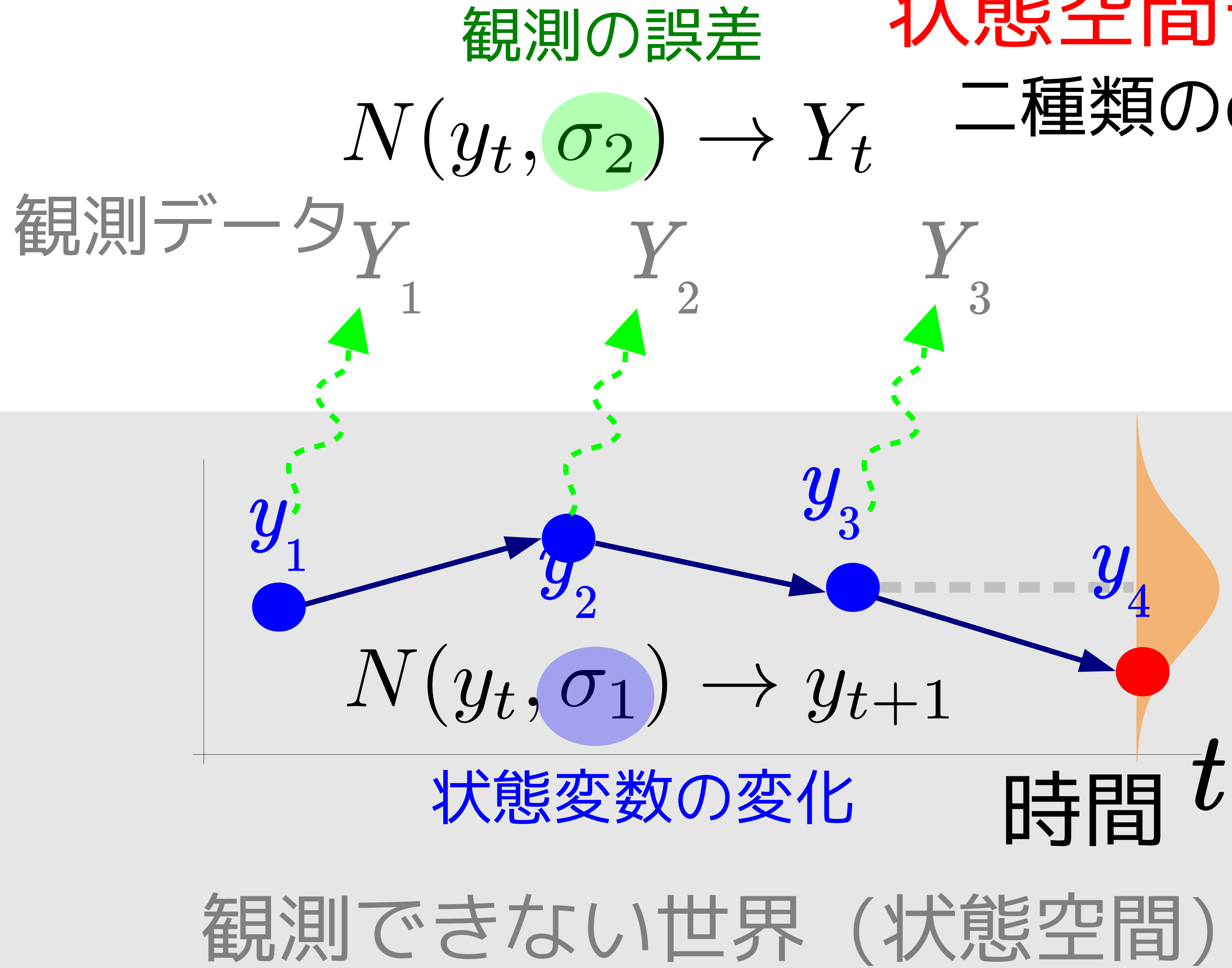
The State-space model,

a unified time-series model



状態空間モデル

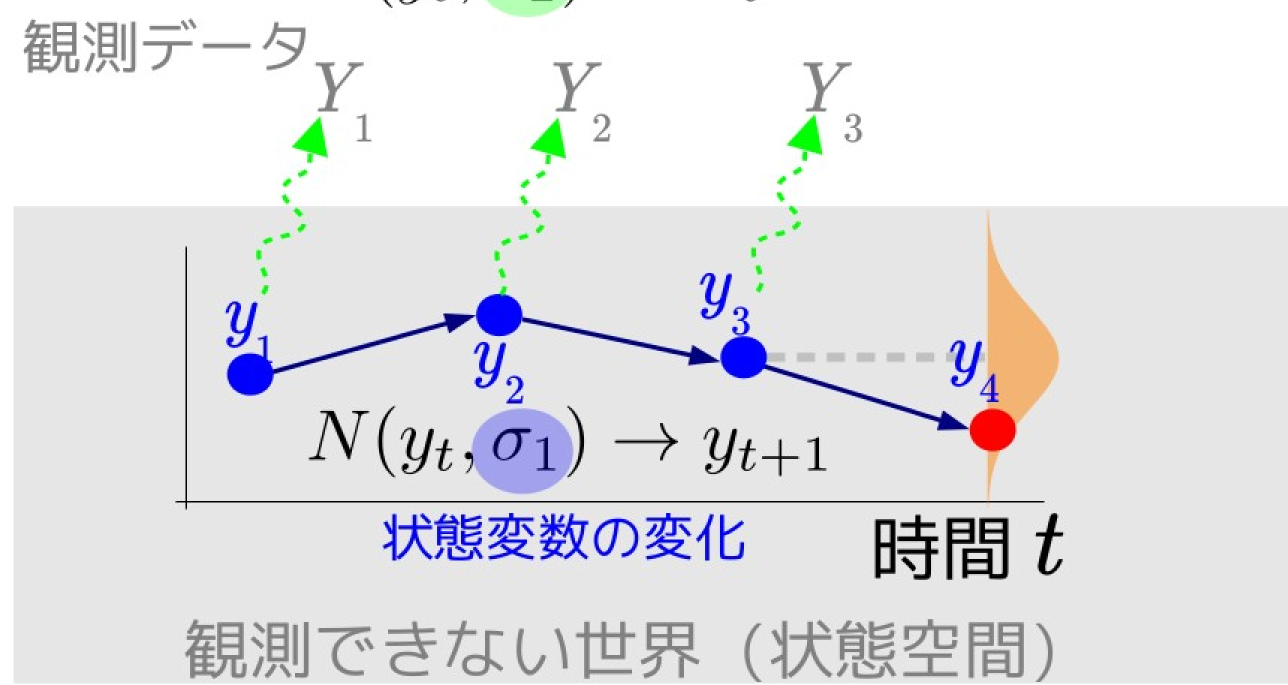
二種類のσをもつ



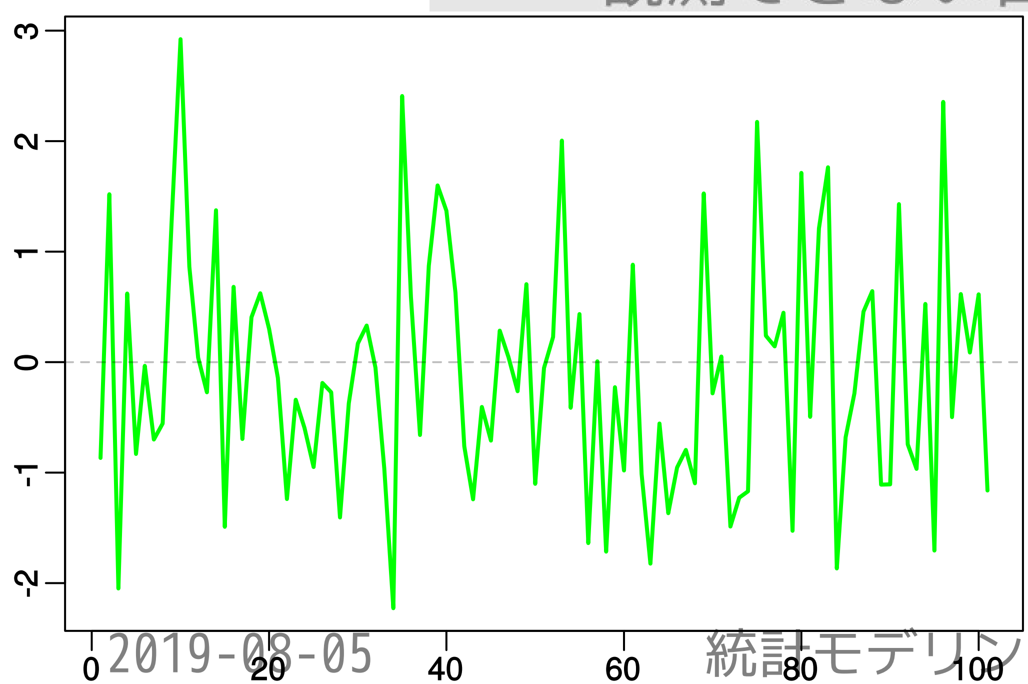
状態空間モデル

観測の誤差

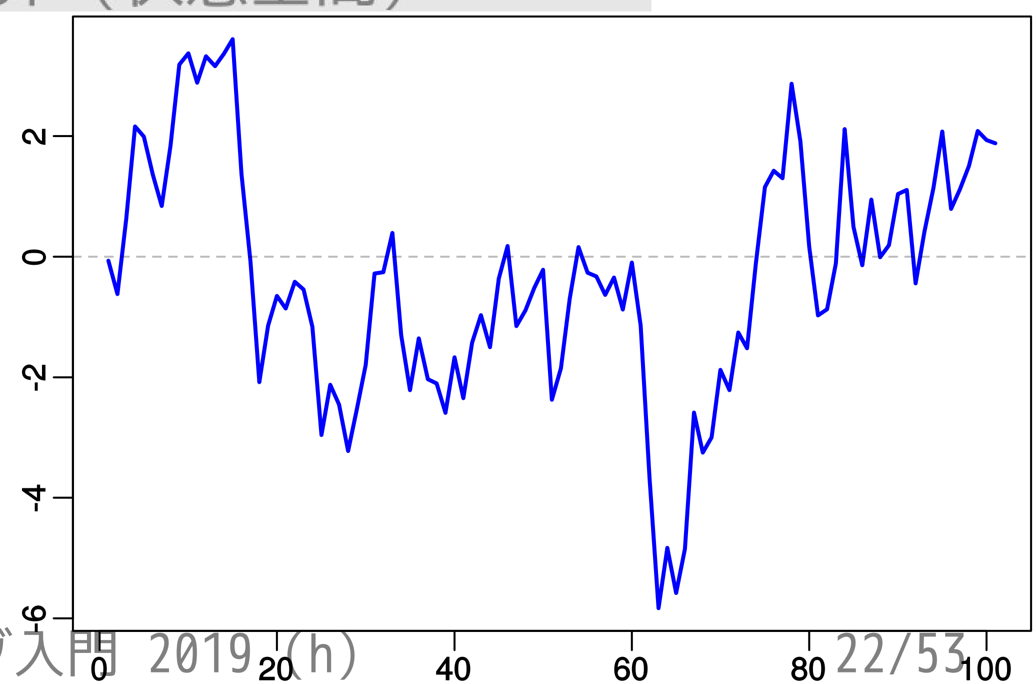
$N(y_t, \sigma_2) \rightarrow Y_t$　二種類の σ をもつ



σ_2 大
σ_1 小

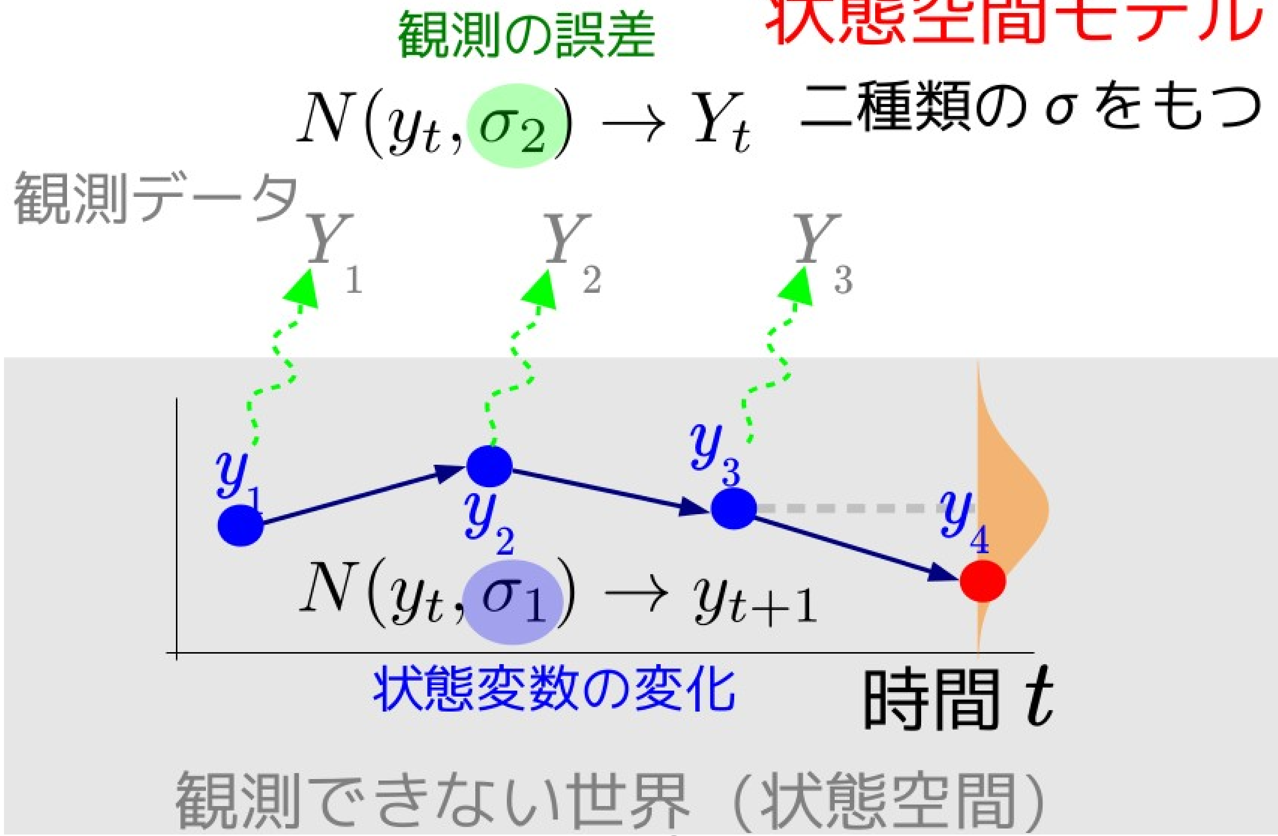


σ_2 小
σ_1 大



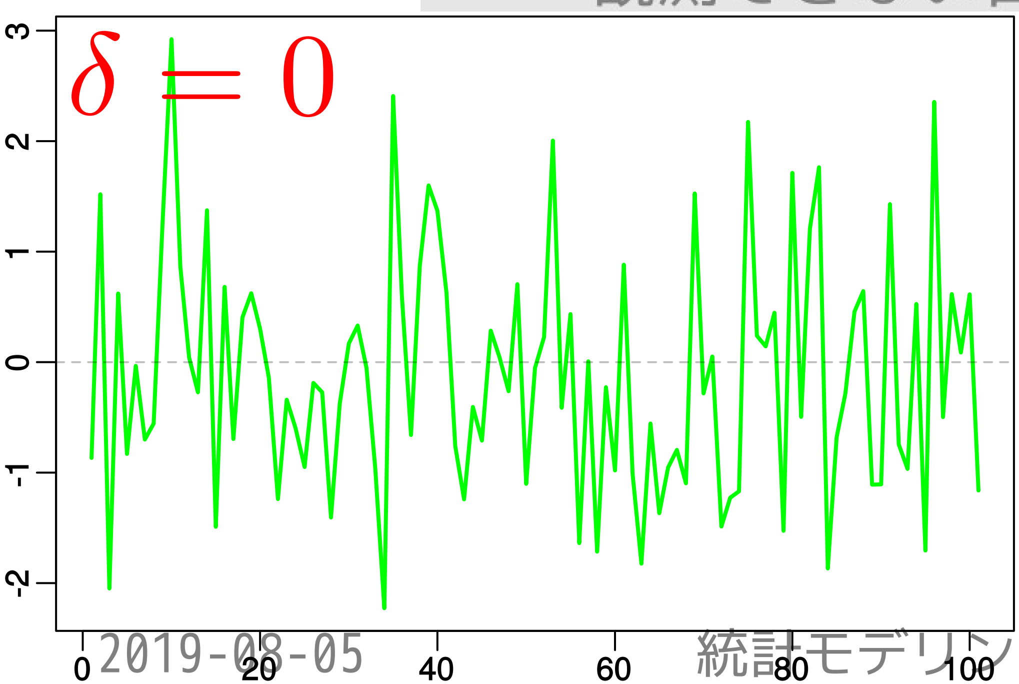
状態空間モデル

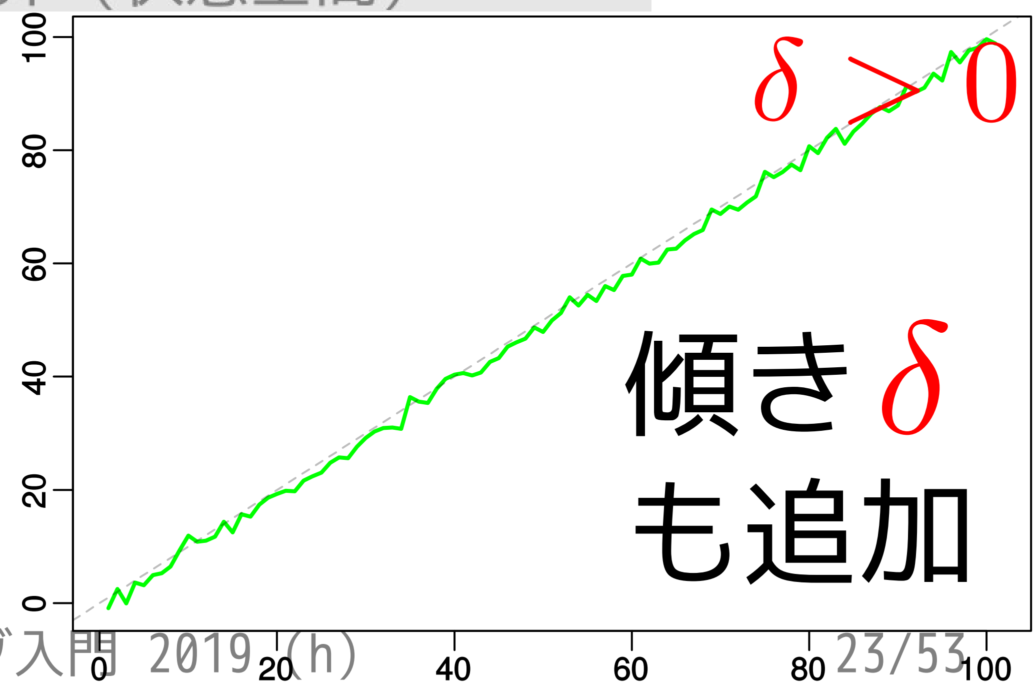
二種類の σ をもつ



σ_2 大

σ_1 小

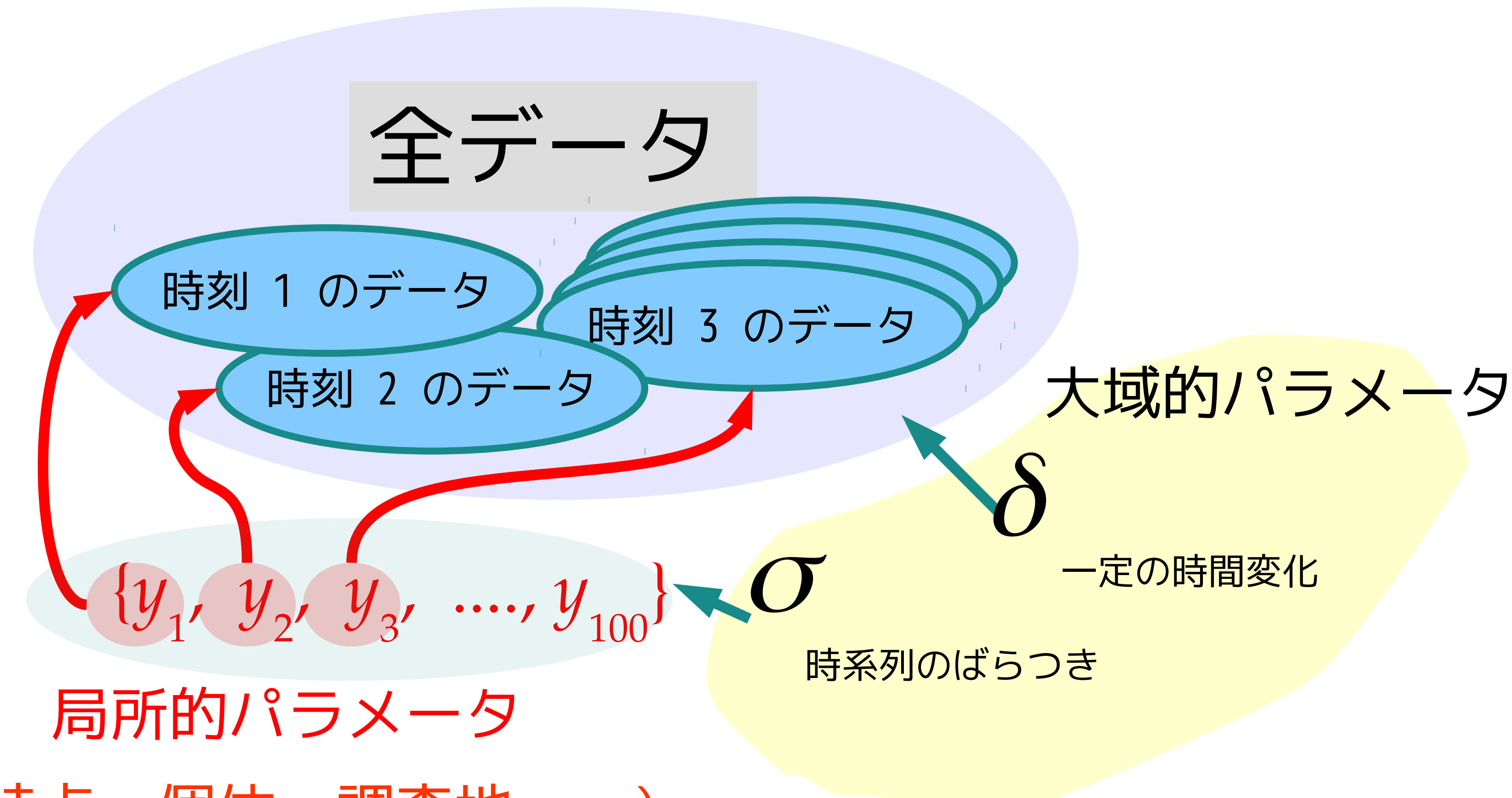




傾き δ も追加

状態空間モデルは階層ベイズモデル

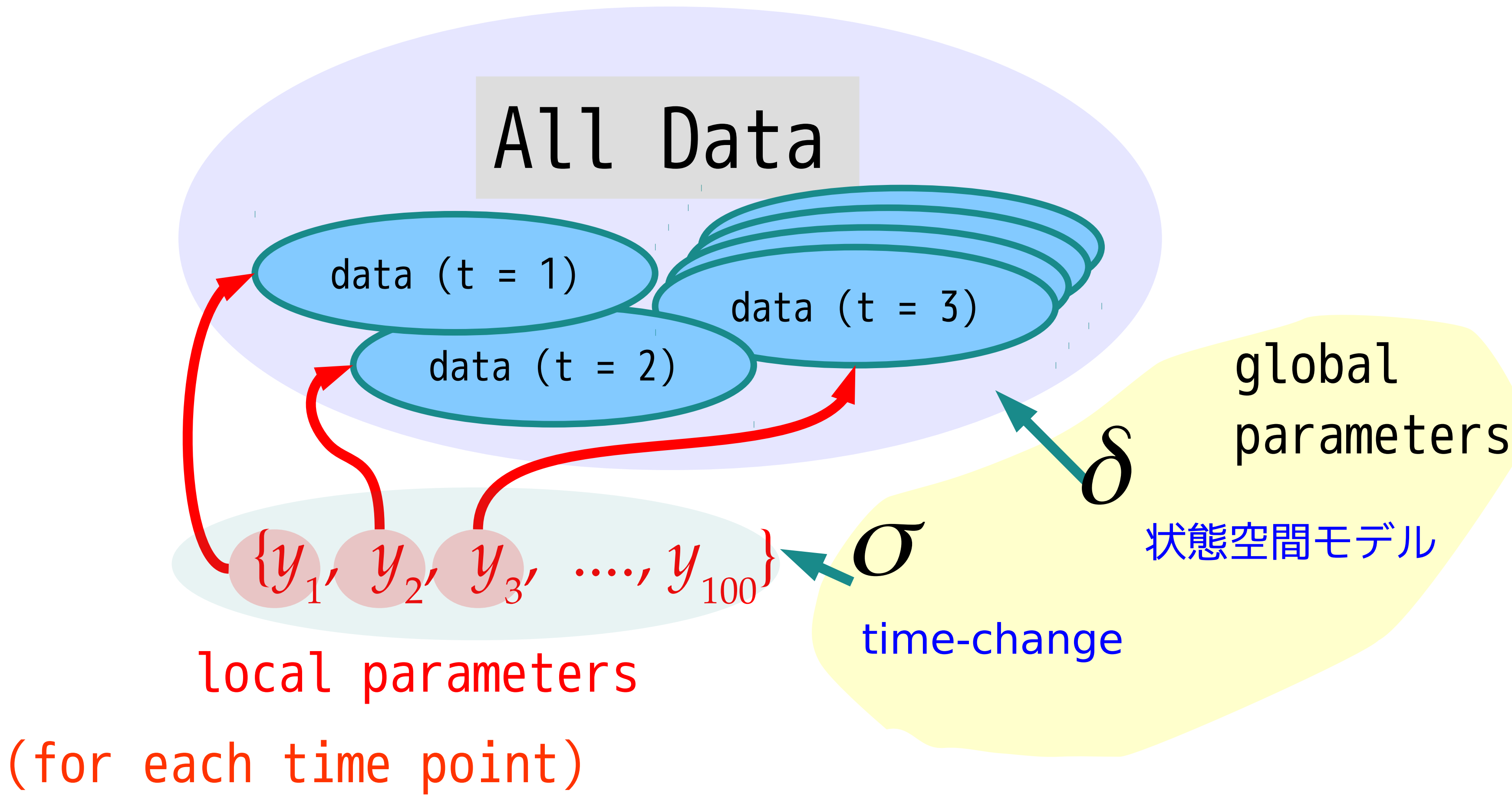
多数の「似たようなパラメーター」たちに
「適切」な制約を加えて推定できる



局所的パラメータ

(たくさんの時点・個体・調査地……)

State space model for TS data, a hiererachical Bayesian model!



状態空間モデルを使う利点

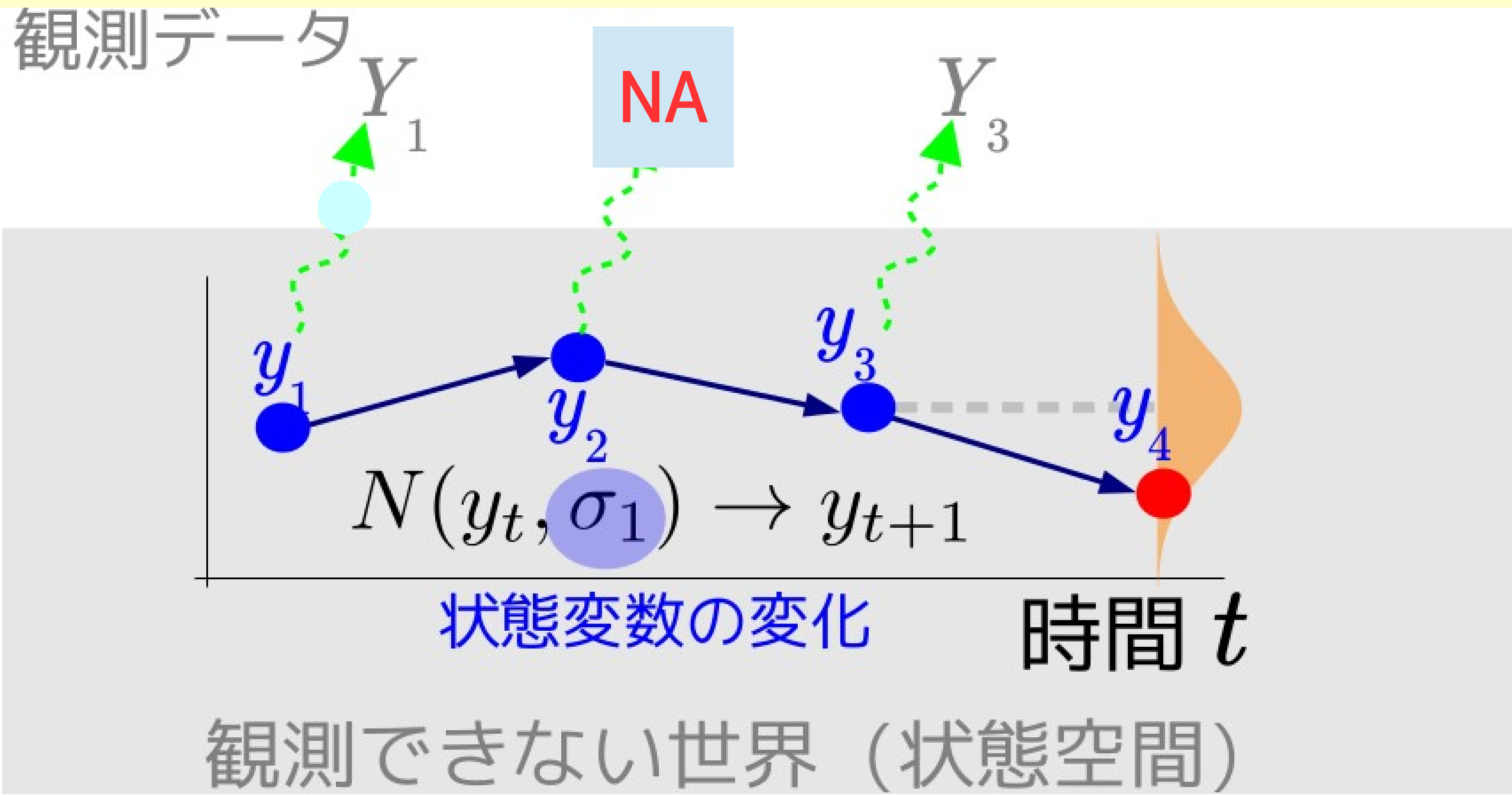
欠測とか不等間隔とか

missing data and
heterogenious time-point

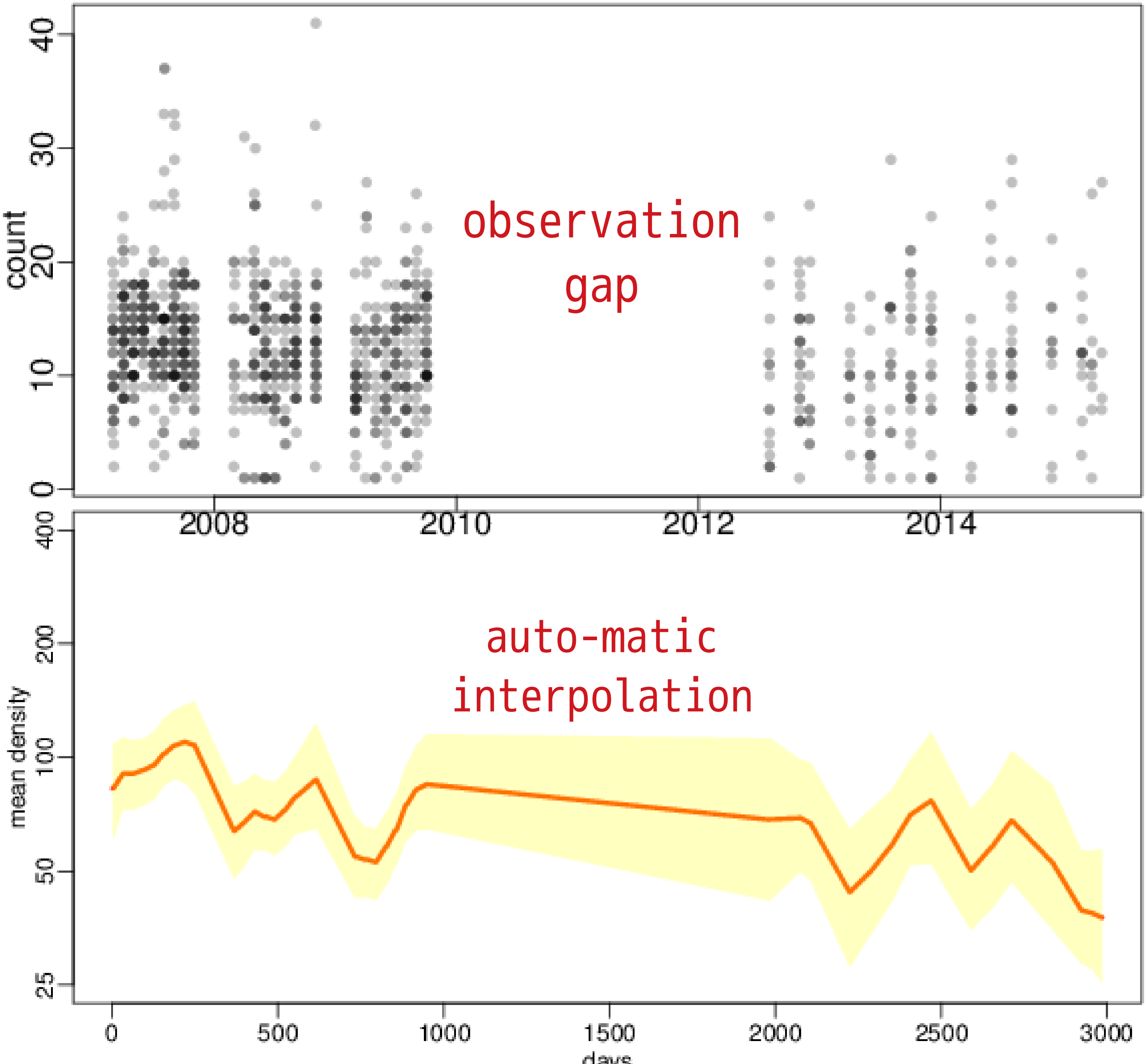
Use state-space model !

状態空間モデル + 観測モデル

欠測があっても問題ない
「補完」の必要なし！



不等間隔データでも何とかなります!



2019-08-05

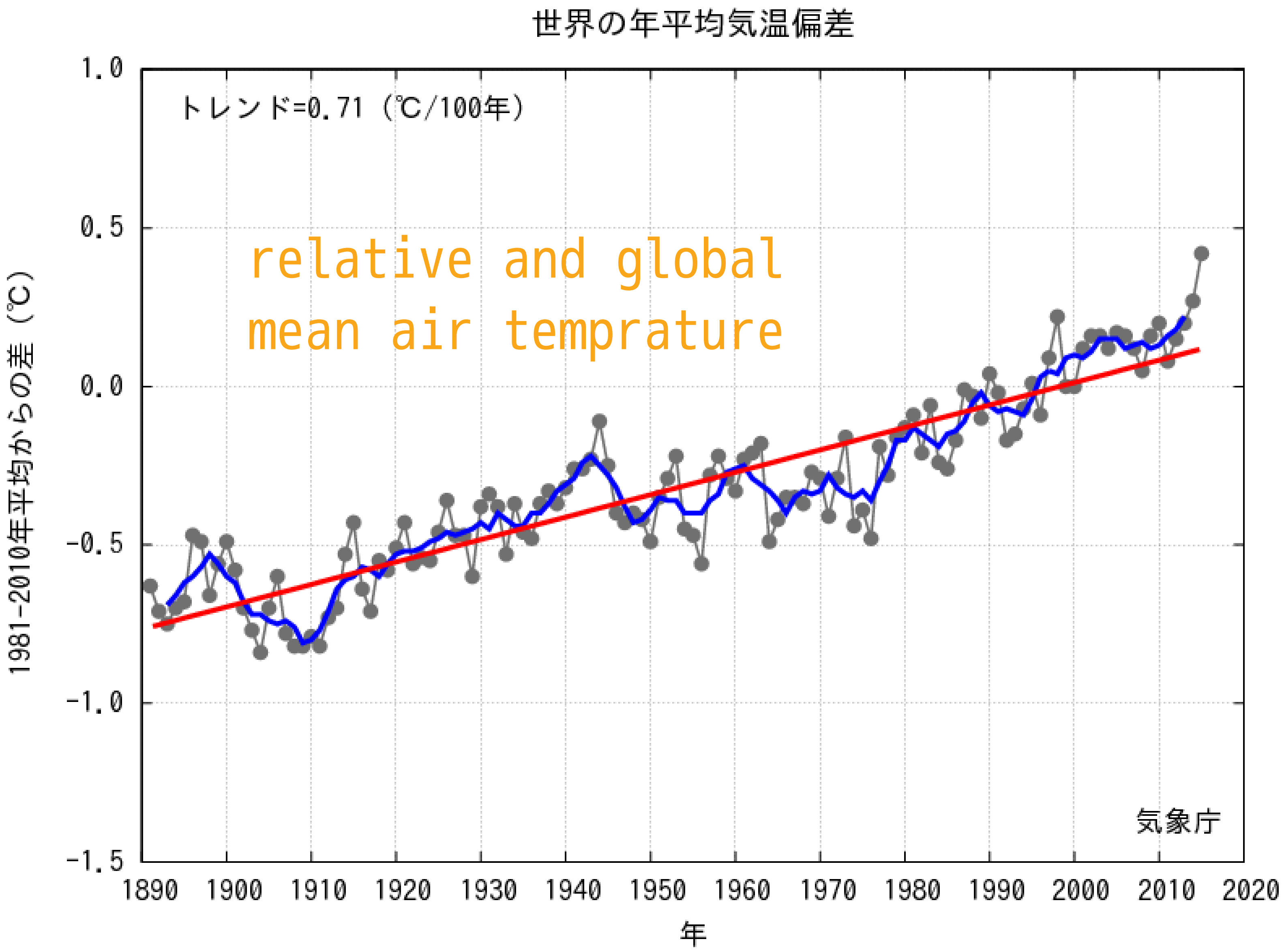
Use State-Space Model!

「ばらばら解析」の回避
気象庁のデータ解析？

An example:

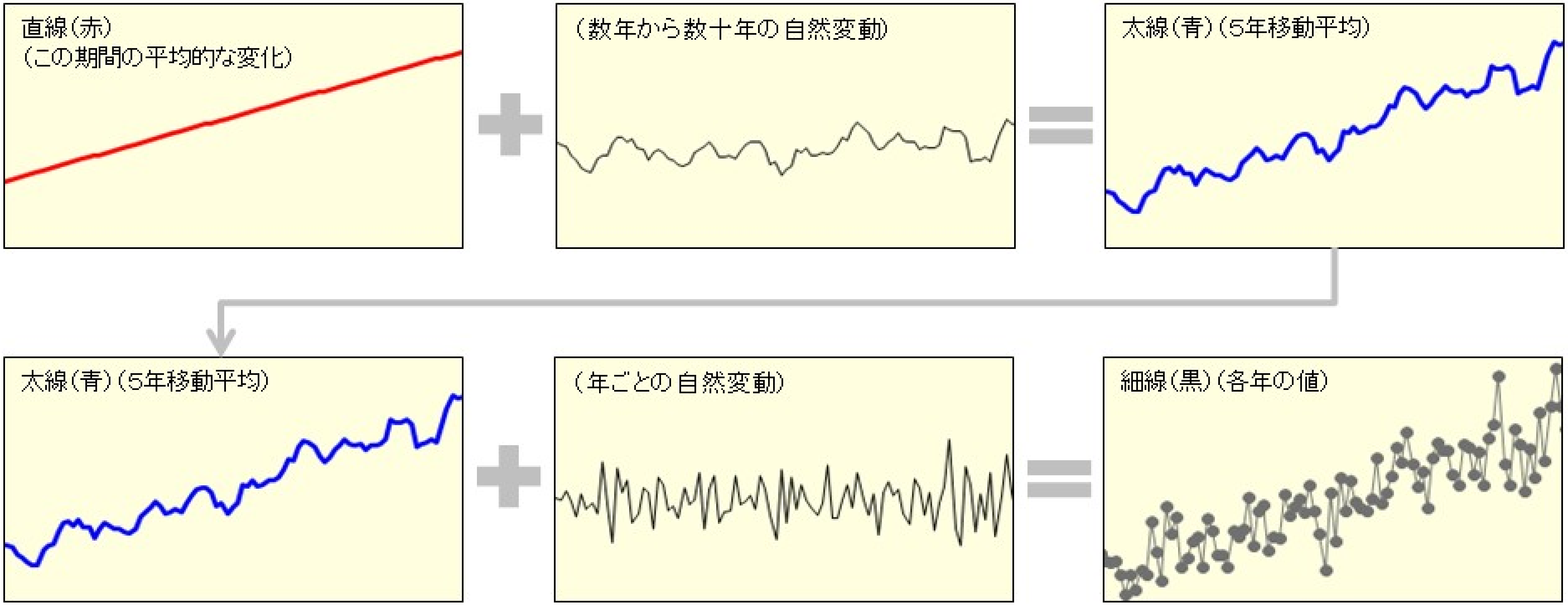
a data set of time series
data: “Is it global warming?”

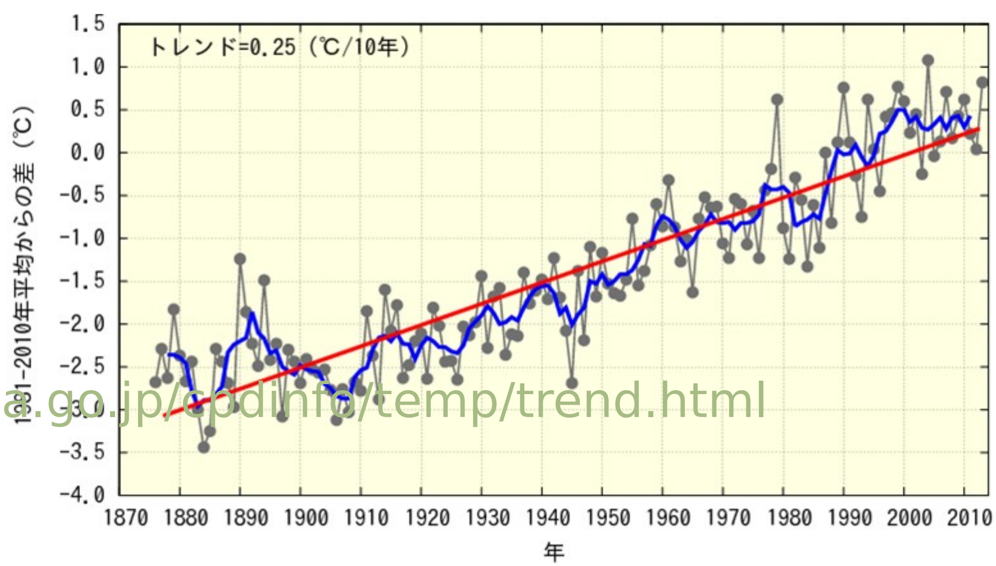
気象庁の長期変化傾向（トレンド）の解説



http://www.data.jma.go.jp/cpdinfo/temp/an_wld.html

気象庁の長期変化傾向（トレンド）の解説

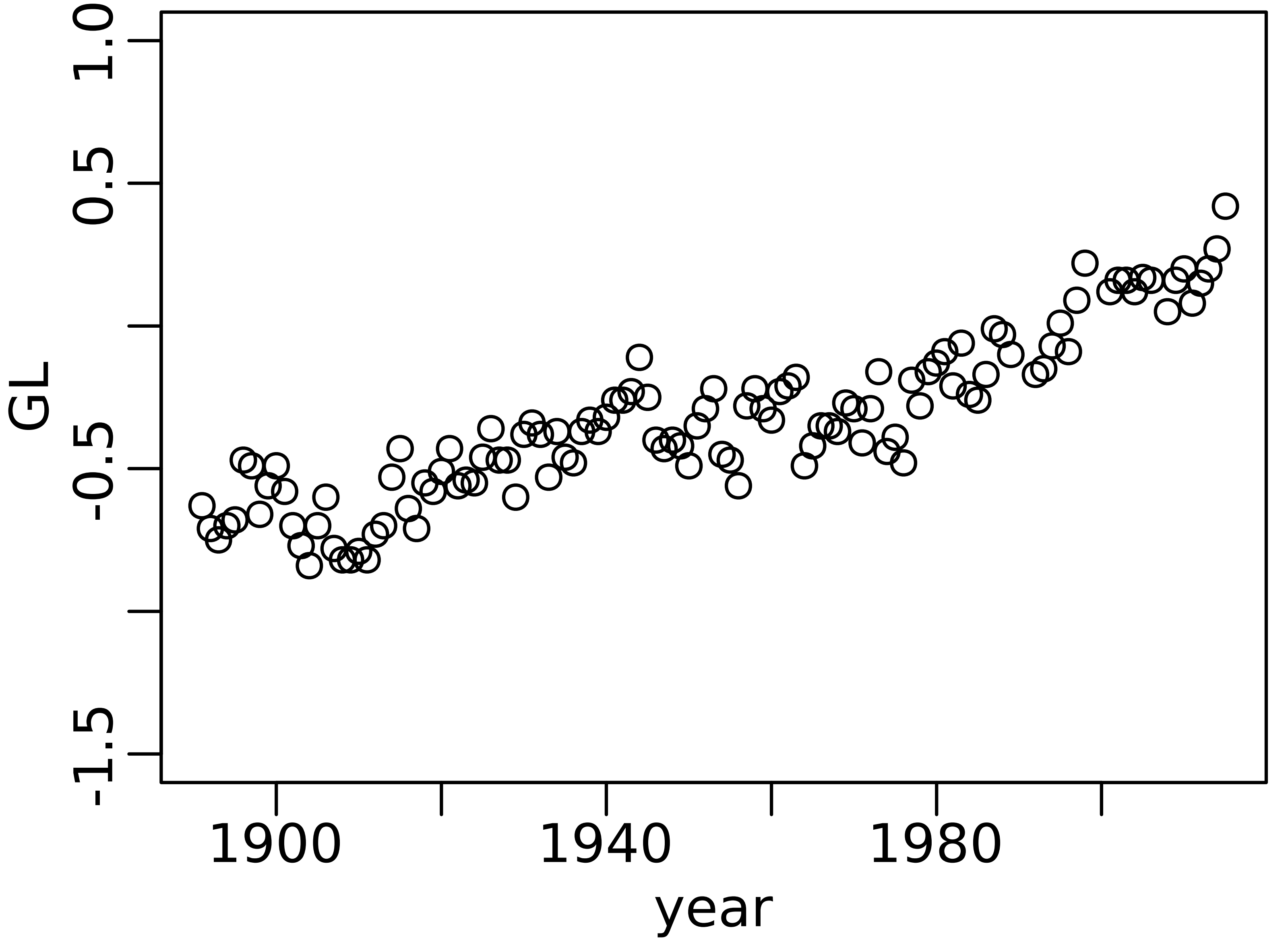




http://www.data.jma.go.jp/cpdinfo/temp/trend.html

2019-08-05

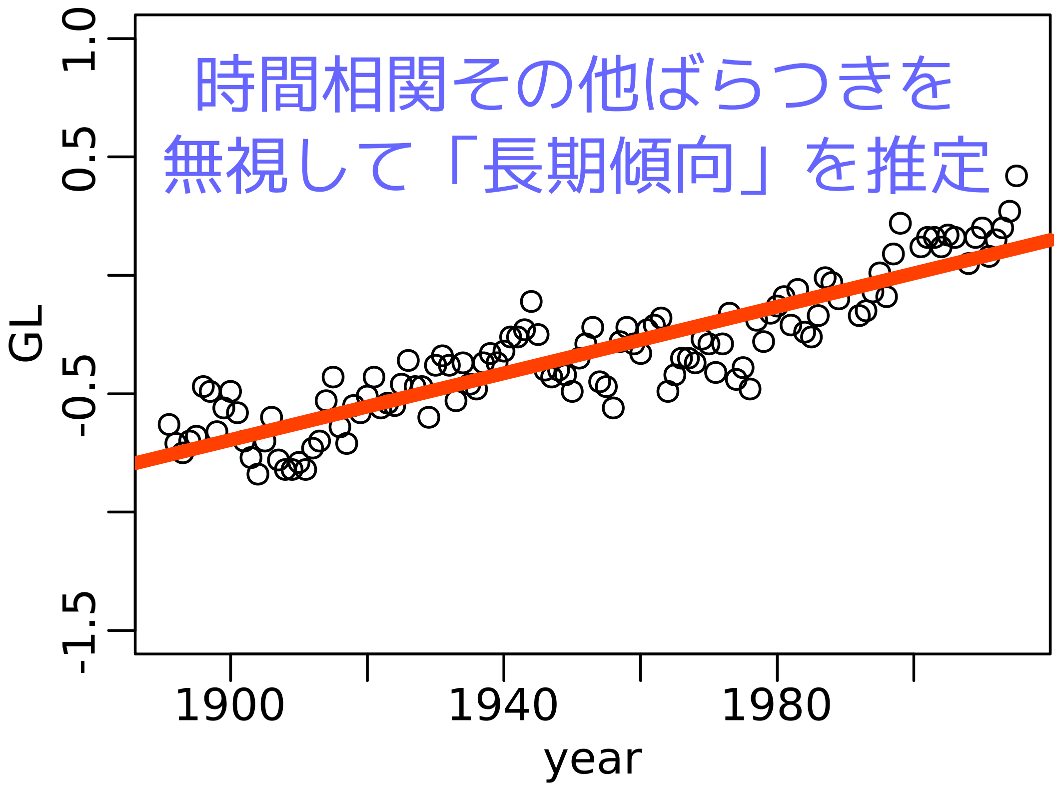
Global warming data



GLM: is it OK? too small p-value!

```
> summary(glm(GL ~ year, data = d))

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.41e+01   6.21e-01   -22.6   <2e-16
year         7.03e-03   3.18e-04    22.1   <2e-16
```

100年
あたり
0.70℃

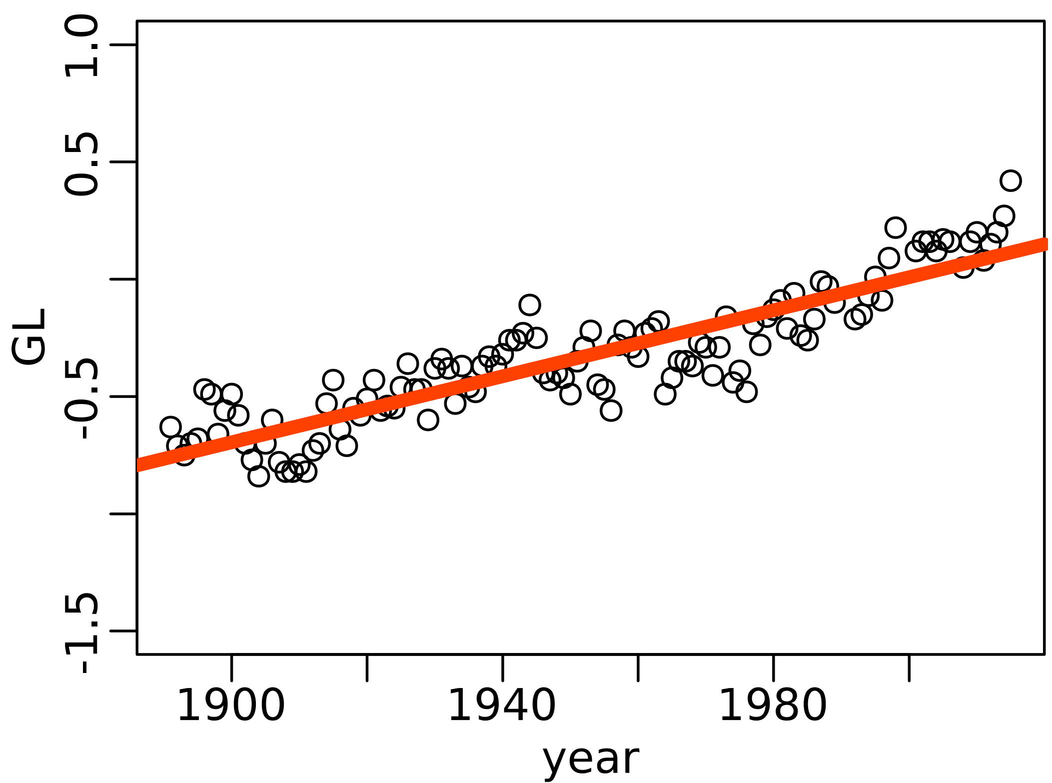


確率１京ぶんの
2?

直線あてはめ（GLM）が予測した「温暖化」

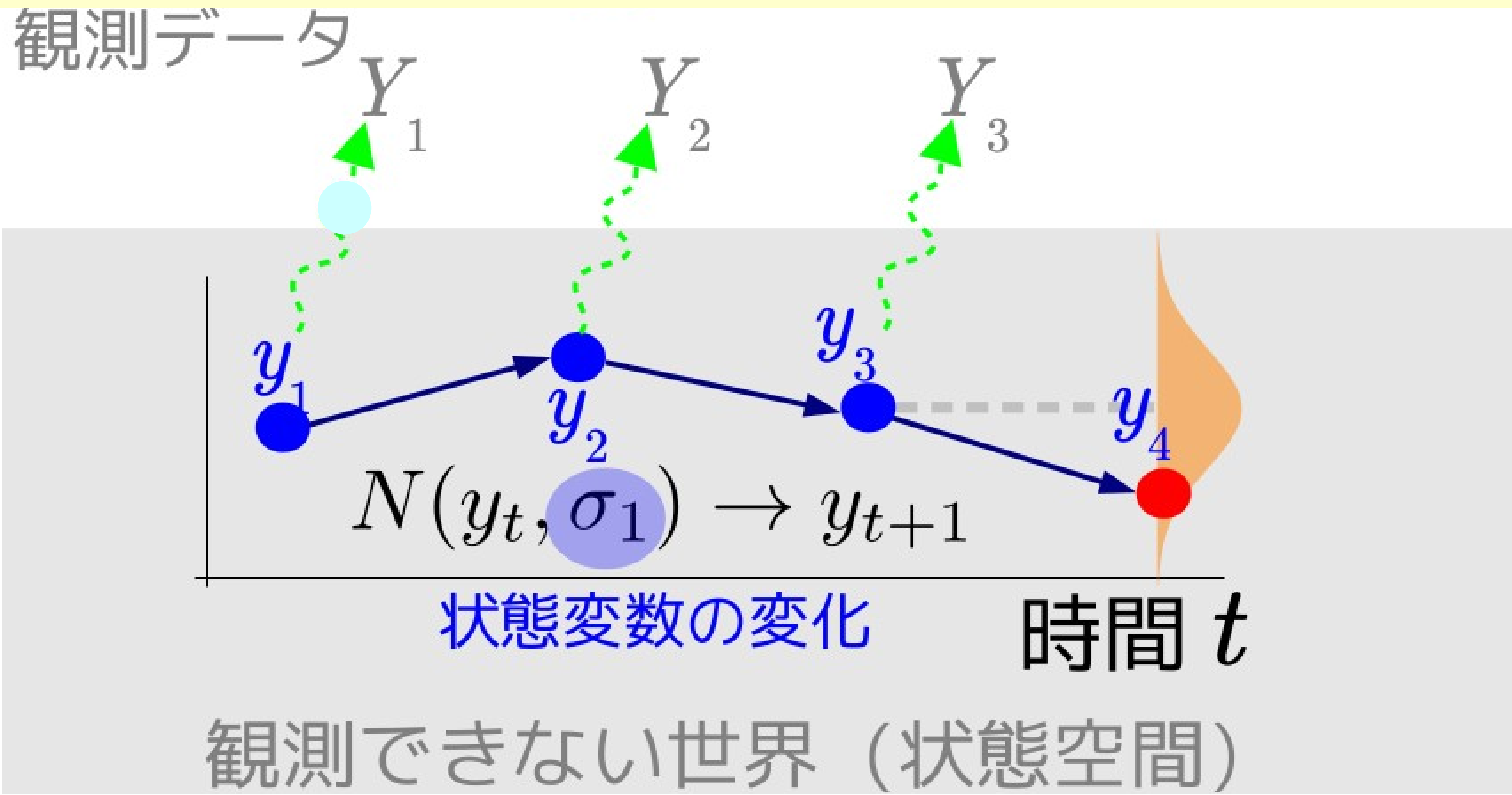
```
> summary(glm(GL ~ year, data = d))

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.41e+01   6.21e-01   -22.6   <2e-16
year         7.03e-03   3.18e-04    22.1   <2e-16
```

100年
あたり
0.70℃



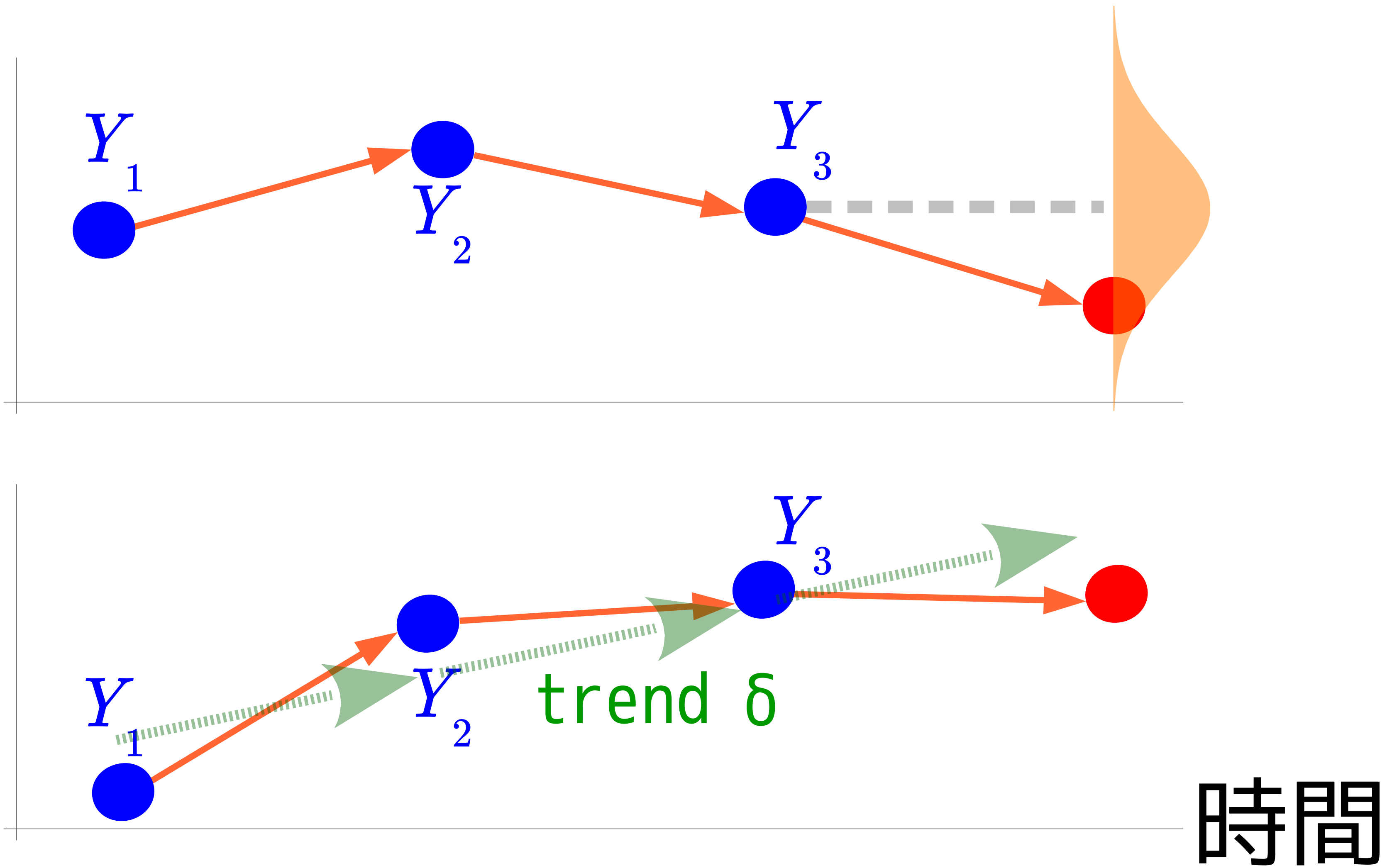
Apply State Space Model (SSM)!

ランダムウォーク+各年独立なノイズ



SSM: Random walk + noise + trend

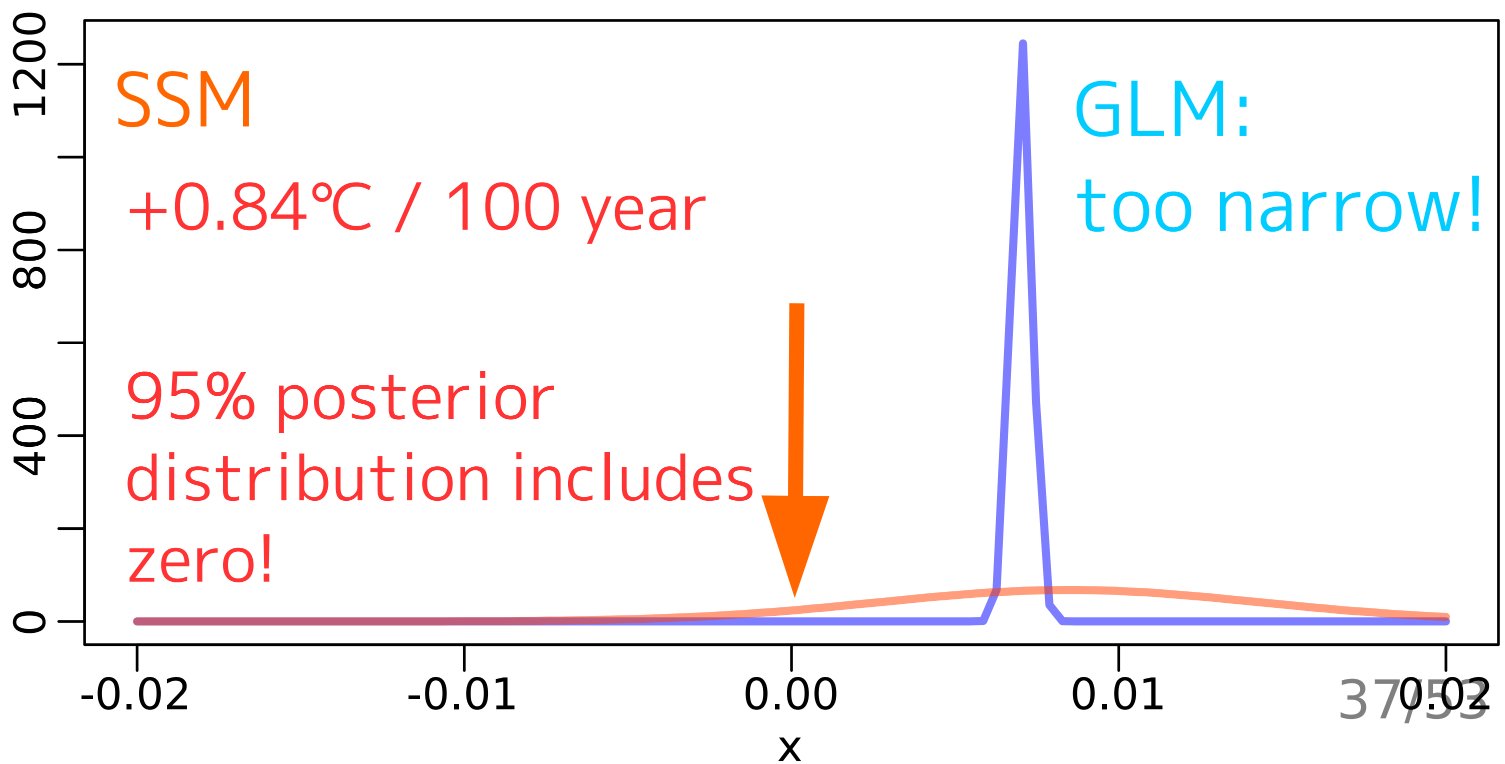
ランダムウォーク+各年独立なノイズ



SSM reveals "uncertainness in global warming"

```
> summary(glm(GL ~ year, data = d))

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.41e+01   6.21e-01   -22.6   <2e-16
year         7.03e-03   3.18e-04    22.1   <2e-16
```

+0.70℃ / 100 year



疑わしい回帰
spurious regression

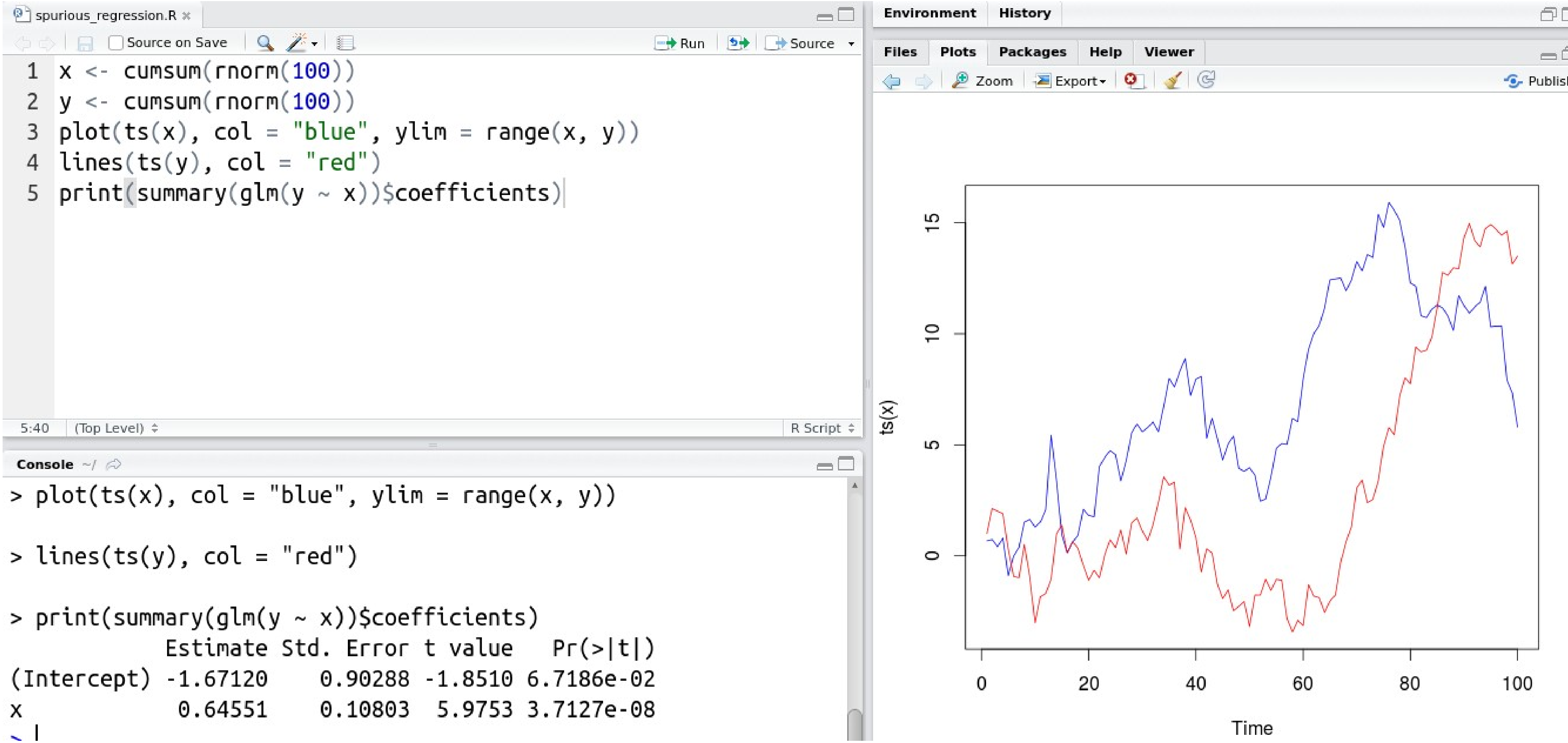
時系列どうしの回帰

time series Y ~ time series X

TS modeling: NOT to do ...

- GLM: Y(t) ~ t and Y(t) ~ X(t)
- combine measurements
- residual analysis
- ... and so on ...

「見せかけの回帰」spurious regression



ちょっとだけ実演してみます

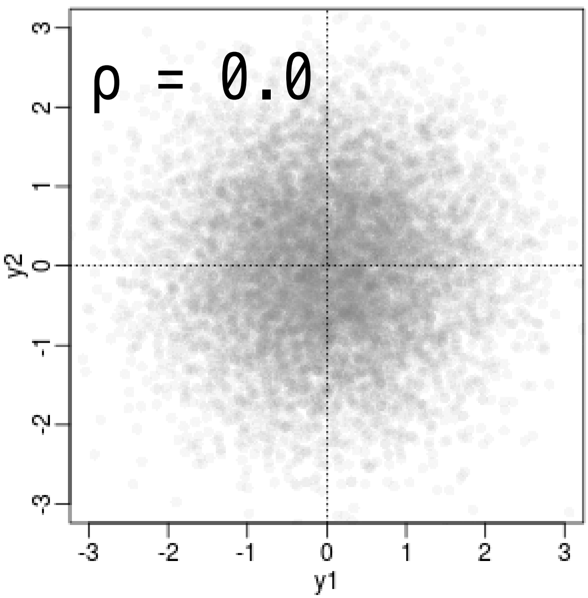
疑わしい回帰
spurious regression

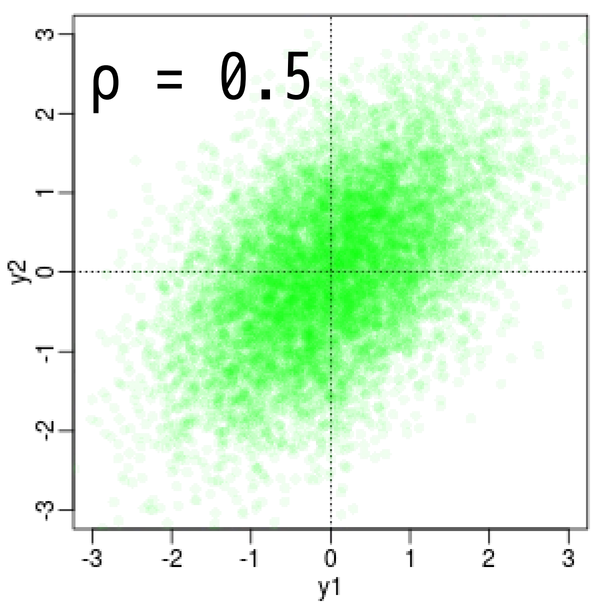
How about fitting

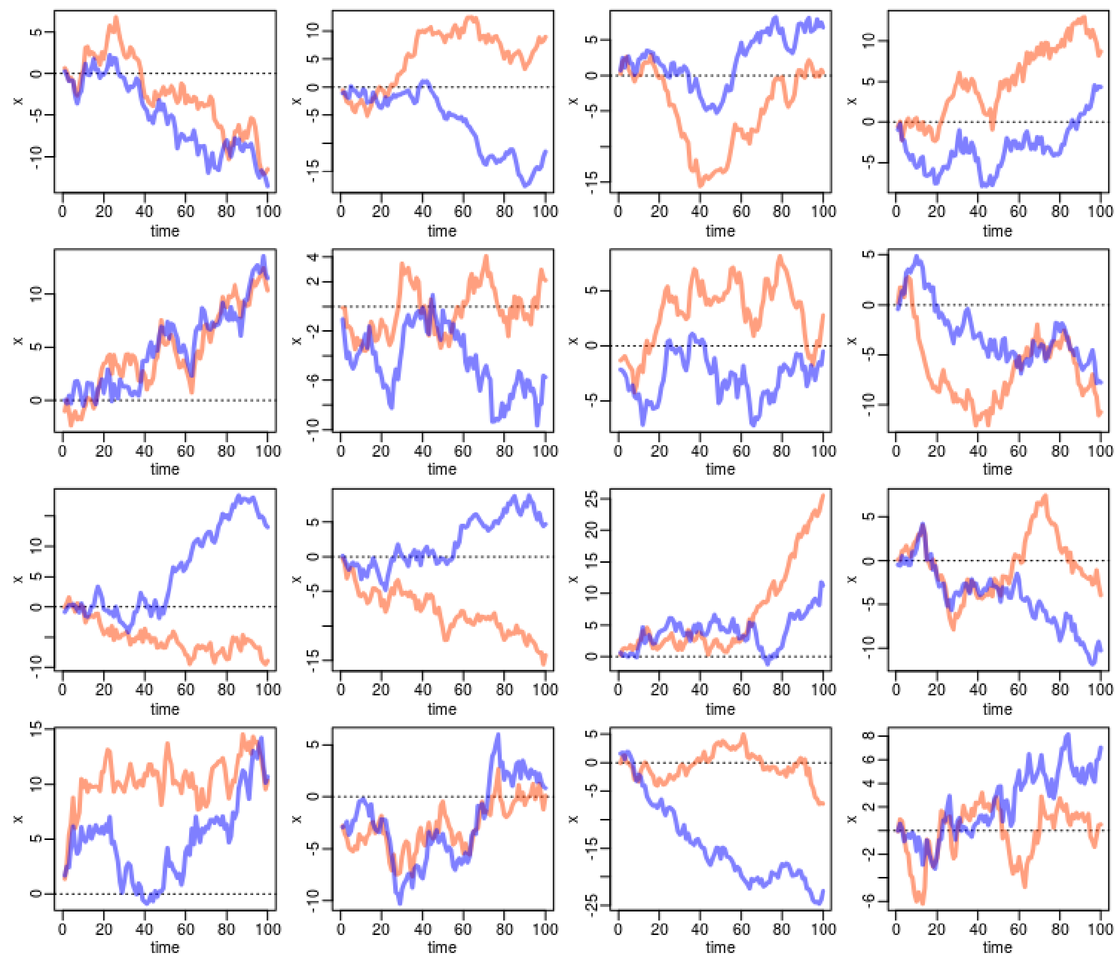
state-space model to estimate

correlations betwee two set of TS

二変量正規分布とランダムウォーク







二変量正規分布を部品とする状態空間モデル

apply bivariate state-space models
including variance-covariance matrix

```
for (i in 1:N.Y) {
    Y[i, 1:2] ~ dmnorm(mu[1:2], Omega[1:2, 1:2])
}
mu[1] ~ dunif(-1.0E+4, 1.0E+4)
mu[2] ~ dunif(-1.0E+4, 1.0E+4)
Omega[1:2, 1:2] <- inverse(VarCov[1:2, 1:2])
VarCov[1, 1] <- sigma[1] * sigma[1]
VarCov[1, 2] <- sigma[1] * sigma[2] * rho
VarCov[2, 1] <- sigma[2] * sigma[1] * rho
VarCov[2, 2] <- sigma[2] * sigma[2]
sigma[1] ~ dunif(0.0, 1.0E+4)
sigma[2] ~ dunif(0.0, 1.0E+4)
rho ~ dunif(-1.0, 1.0)
```

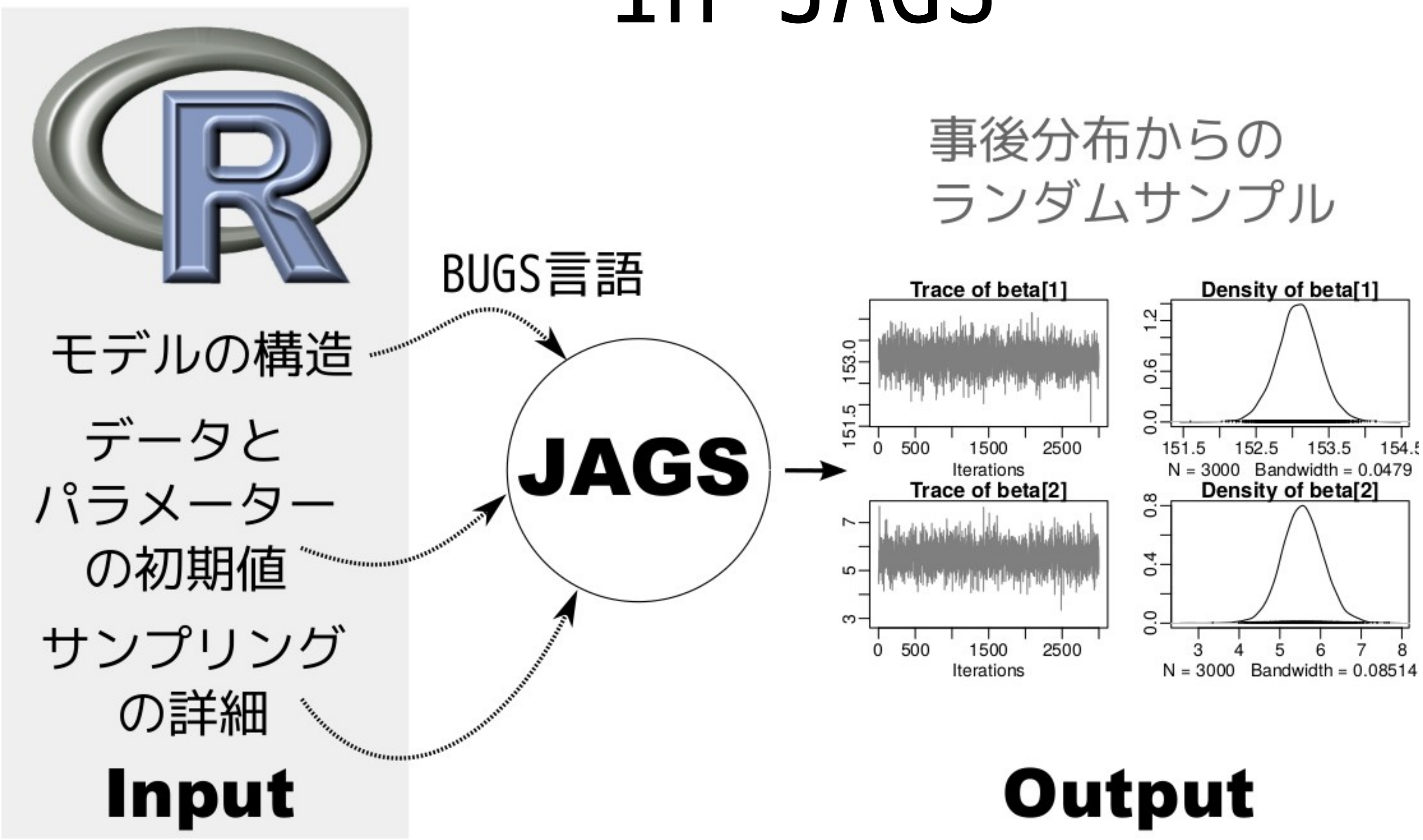
bivariate state space model estimates the posterior of variance and covariance matrix

```
 3 chains, each with 5200 iterations (first 200 discarded)
 n.sims = 15000 iterations saved
            mean    sd   2.5%    25%    50%    75% 97.5%  Rhat n.eff
mu[1]     -0.122 0.110 -0.342 -0.195 -0.120 -0.048 0.090 1.001  6000
mu[2]     -0.157 0.100 -0.355 -0.224 -0.157 -0.091 0.041 1.002  1500
sigma[1]   1.091 0.079  0.949  1.036  1.086  1.142 1.261 1.001  6100
sigma[2]   0.993 0.074  0.864  0.941  0.987  1.039 1.151 1.001  4100
rho        0.568 0.070  0.420  0.523  0.573  0.617 0.693 1.001 11000
```

ふたつの時系列データの変動が
相関しているかどうかを特定できる

MCMC parameter estimation

Hierarchical model written in JAGS



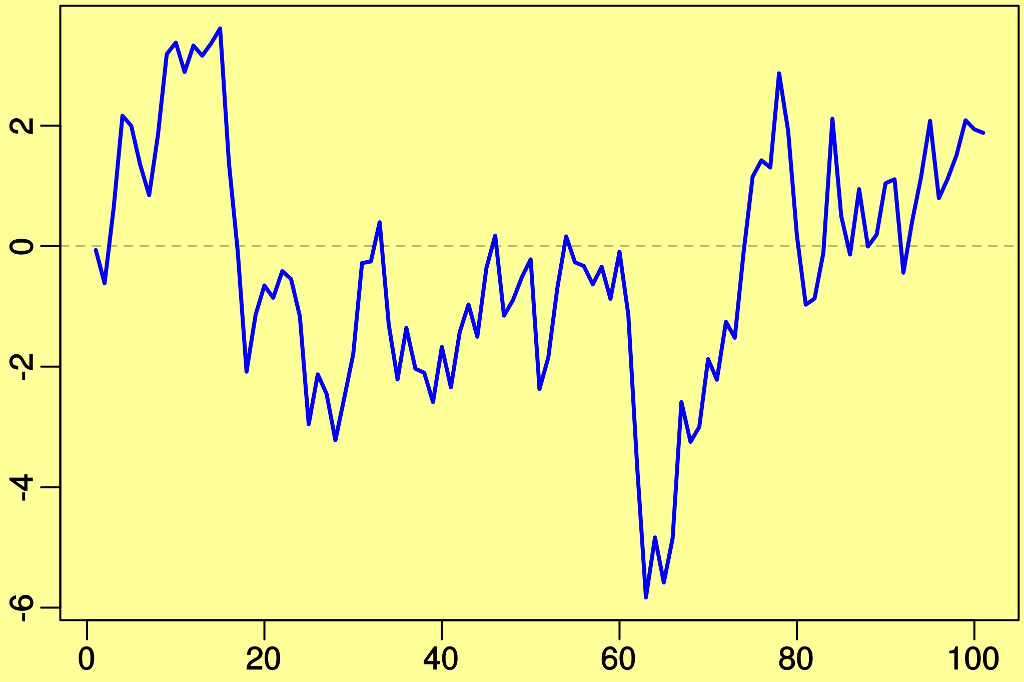
```
model
{
    Tau.Noninformative <- 0.0001
    Y[1] ~ dnorm(y[1], tau[2])
    y[1] ~ dnorm(0, Tau.Noninformative)
    for (t in 2:N.Y) {
        Y[t] ~ dnorm(y[t], tau[2])
        y[t] ~ dnorm(m[t], tau[1])
        m[t] <- delta + y[t - 1]
    }
    delta ~ dnorm(0, Tau.Noninformative)
    for (k in 1:2) {
        tau[k] <- 1 / (s[k] * s[k])
        s[k] ~ dunif(0, 10000)
    }
}
```

Apply hierarchical models to time-series (TS) data! i.e. State Space Models

σ_2 小

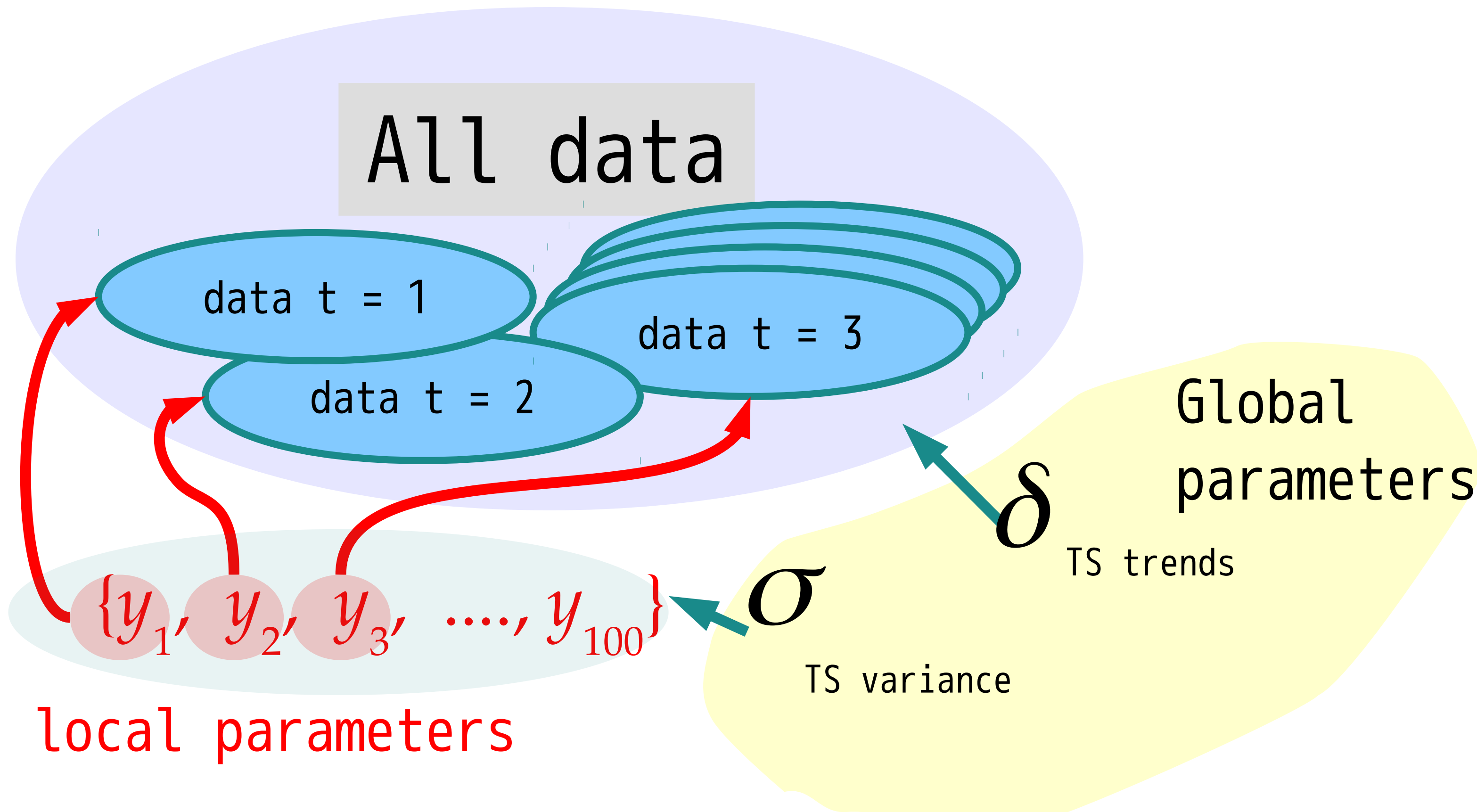
σ_1 大

$\delta = 0$



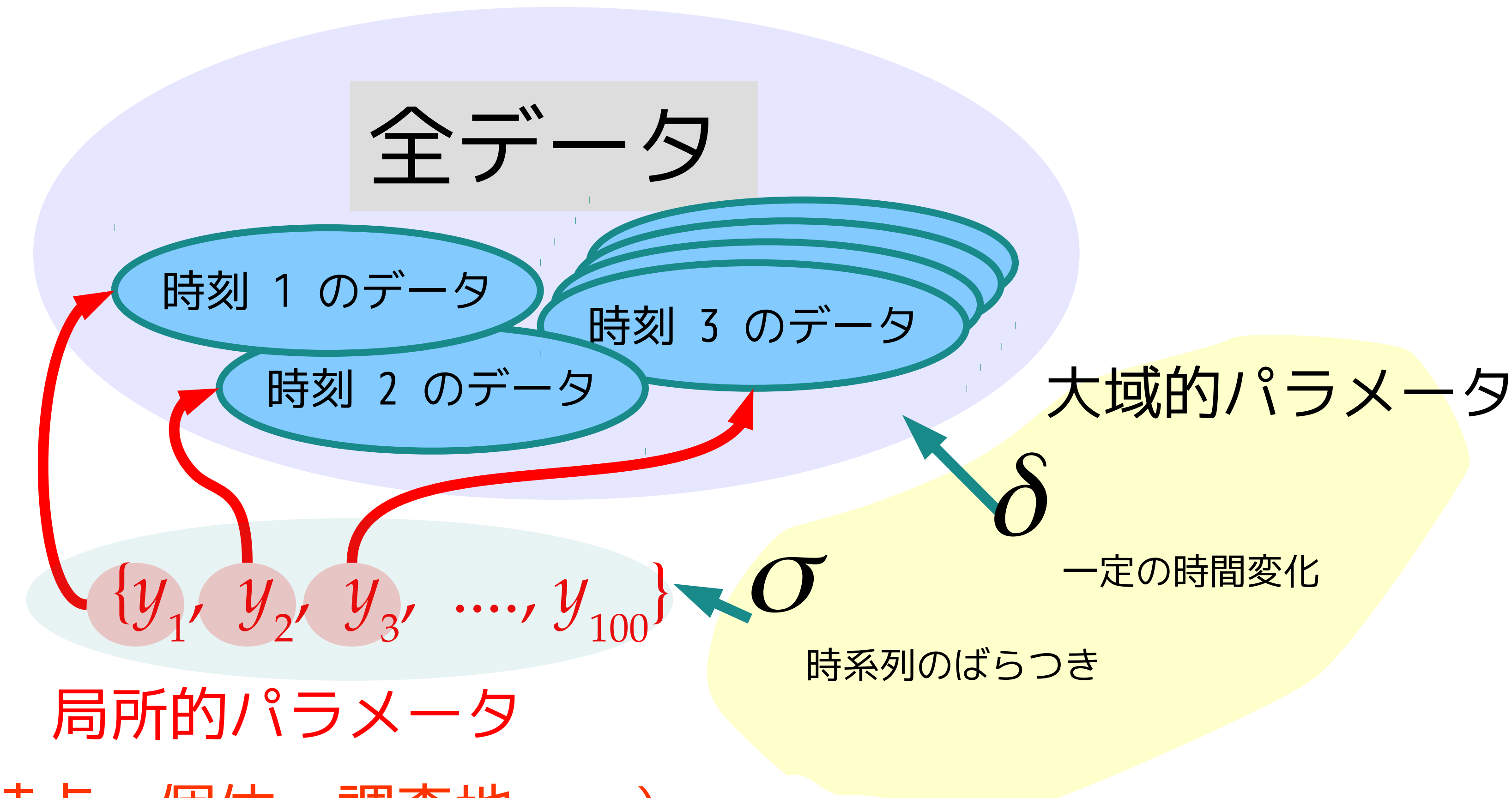
Hierarchical Model is powerful!

While GLM can not model TS data,
Hierearchical model is effective!



階層ベイズモデルとは?

多数の「似たようなパラメーター」たちに
「適切」な制約を加えて推定できる



(たくさんの時点・個体・調査地……)

How do you fit TS model to data?

R packages to estimate parameters in stat-space models

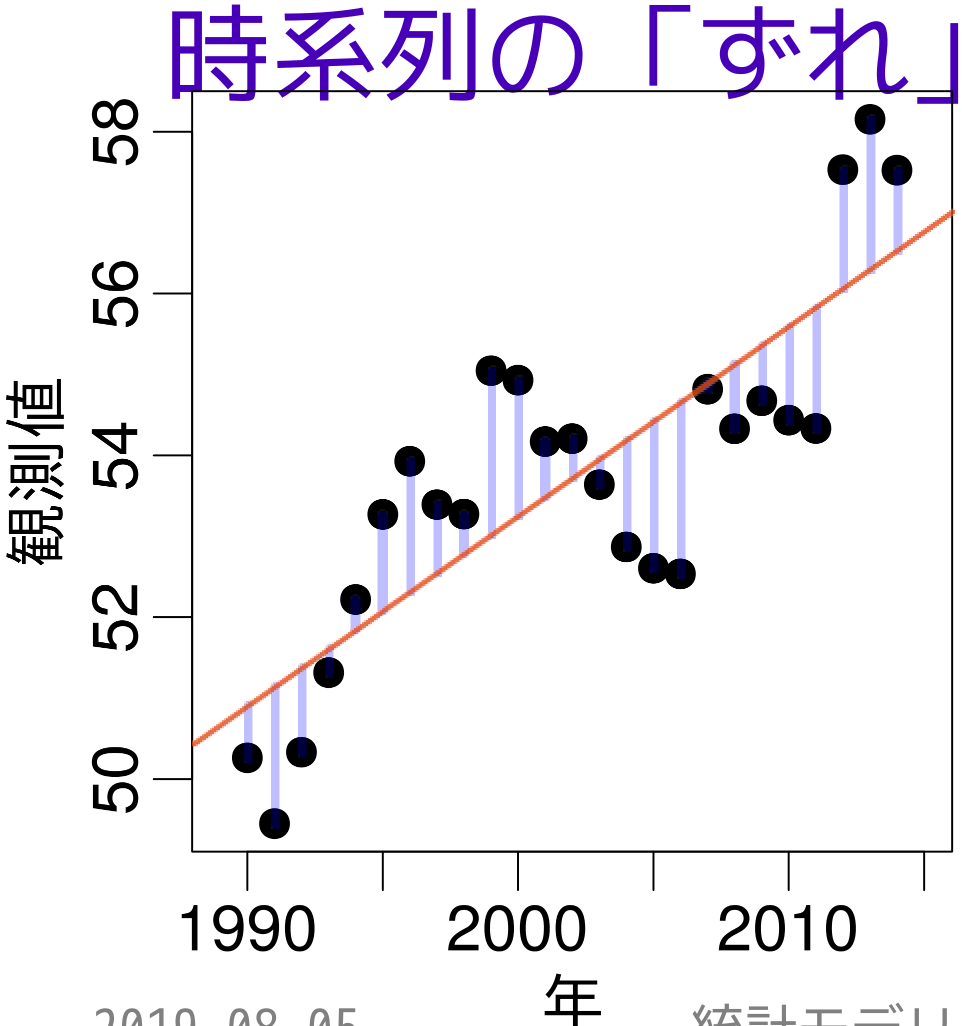
```
library(dlm)
library(KFAS)
```

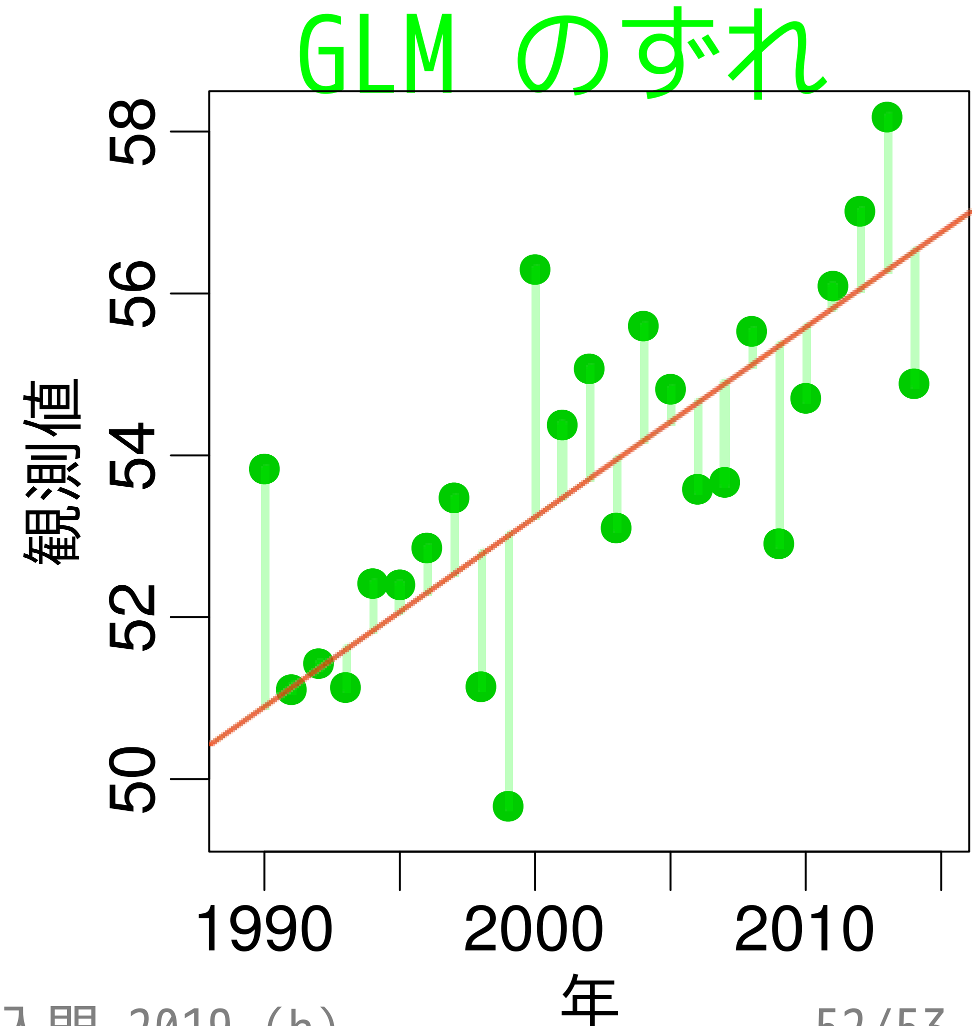
Also we have more powerful models to analyze TS data...

おわりに

take-out message

Modeling time series data can be a hierarchical modeling





The End: have a nice statistical modeling!

The Evolution of Linear Models

Parameter Estimation

MCMC

Hierarchical Bayesian Model (HBM)

↑

Generalized Linear Mixed Model (GLMM)

MLE

↑

Generalized Linear Model (GLM)

MSE

↑

Linear Model

データ解析は

階層ベイズモデルで！

2019-08-05