2/28 → 2/27 ddl　　　2/11 – 2/27　　　14天.

1. 收集数据　　　2/11 – 2/13　　　13晚　　美东晚八.
↓
2. EDA　　　2/13 – 2/15.　　　15晚
↓
3. 正式数据　　　2/15 – 2/17
↓
4. 改善　　　2/17 – 2/19
↓
5. 确认　　　2/19 – 2/21
↓
6. Narrative　　　2/21 – 2/23
↓
7. Code submit　　　2/23 – 2/25.

template →
→
→

Python

+1 • Template
+2 • 简历

−1 • 统计包 ×.

R ✓.

+1 • TS / forecast / 包 ✓.
+2 • Rmd.

1st

Prof. Japanese Language Culture

Cultural significance of cherry & cherry blossom in Japan.

flower viewing - peak blooming time

late march ~ early April

Map 2017~2021

2nd.

phenological model

Timing

- Harvest date

global warmth → shift a month early

- 1731 Reaumur scale   0 — 100 water
  freeze. boil
  1735 GDD

① Global warm vs. local temperature

② GDD +winter chilling model

③ Empirical Function

$TSD = \exp\{\cdots \frac{T-0}{288.57}\}$

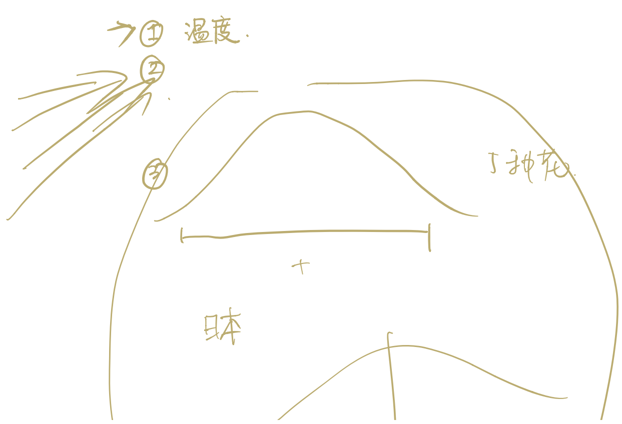
④ 花开速度

bucket model

- winter chilling
- spring warming GDD

crop models (stone fruits)
wild animals
forest ecosystems

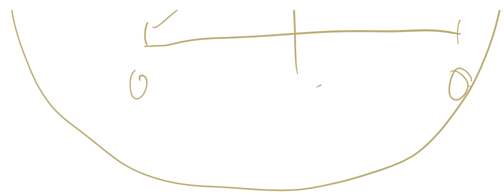
① 温度 昼夜
② 太阳 ↑↓
③ 云覆盖率
④ 湿度
⑤ 海拔



+ 下载 +学习 Rstudio.

3rd

nature's notebook



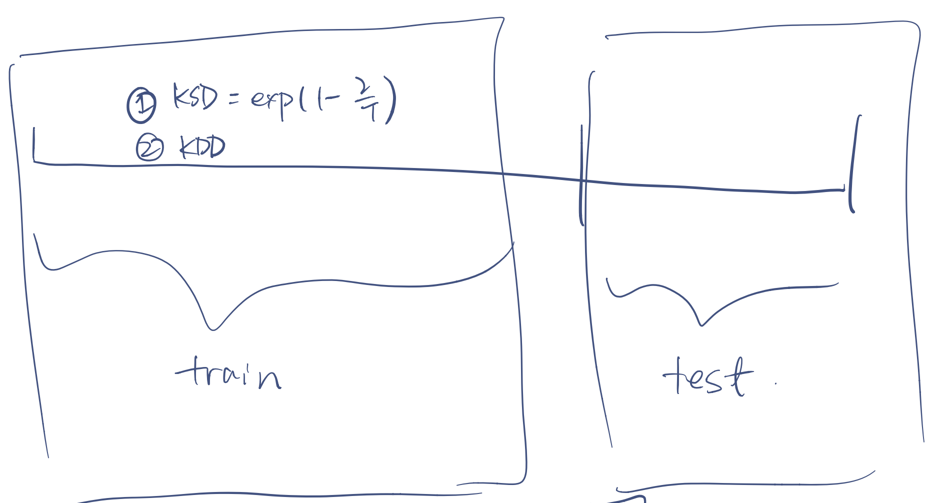
USA-NPN

3种花 2009-2021 的数据

status & intensify

individual ID → plants ID

individual phenometrics



① T → KSD . → 1天

② → 2天 .

10-fold cross validation

Kyota 801 - 2021 80%. bloom.

test 20%.

801 1999 2000 2021

V. overlap

20

10

20

20

10

non-overlap.

20

forecast horizon. = 1.

h = 10 20 1

KDD model (过去20年).

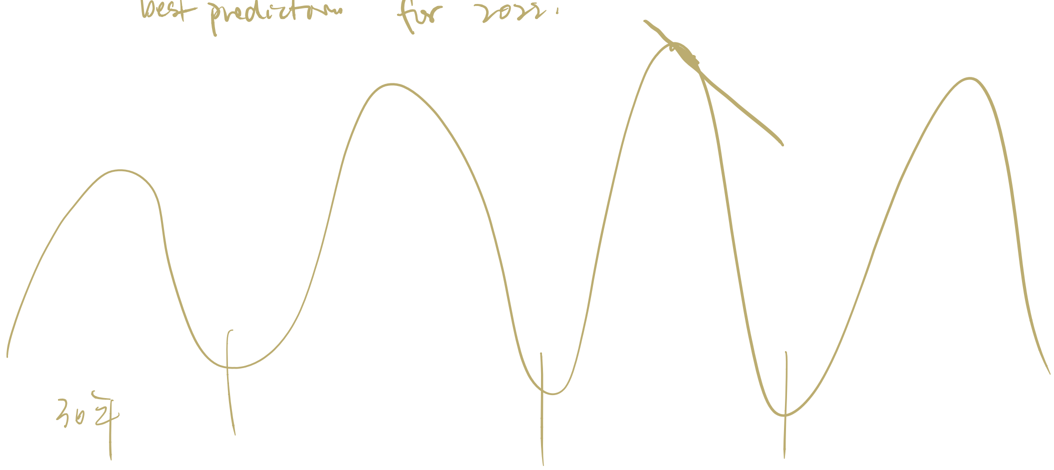
20年 10年

2022 — 2031

KDD(2001 - 2021) → 2022-2031.

KDD(1980 - 2000) → 2001-2011

best prediction for 2022.



30年

$$KSD = \exp\left\{ 1 - \frac{T - 288.5}{288.5\,T} \right\} \rightarrow \text{Japan}. \quad \chi.$$

$$\log KSD = 1 - \frac{1}{288.5} - \frac{1}{T} \approx 1 - \frac{1}{T} \quad ☆☆☆.$$

KDD. $T + S$.

2022/02/13

1. 收集数据　　2/11 - 2/13　　13晚
   ↓
2. EDA　　2/13 - 2/15 .　　14 - 15
   ↓
3. 正式数据　　2/15 - 2/17　　16-17
   ↓
4. 改善　　2/17 - 2/19　　18-19
   ↓
5. 确认　　2/19 - 2/21
   ↓
6. Narrative　　2/21 - 2/23
   ↓
7. Code submit　　2/23 - 2/25.

瑞士　　小时

日本　　附近
　　　　月为单位
　　　　pdf 非 csv.

加拿大　　气温

DC　　station

日本 ①	DC
瑞士 .	加拿大

① 日本　　T= Temperature

winter chill　　summer warm

T → GDD_WC ← V1
　　GDD_SW ← V2.

pt. — GDD → GDD f1.

compare 经验公式 with f1.

$y = \beta_0 + \beta_1 V_1 + \beta_2 V_2$ 日本.

y = logit

exp

日本. 日照.

DC. 气温

. . . . . .

f1 → 日本

Y

②.

β_0 β_1 β_2

V_3 →

V_4.

模块化: ① $l_1 = lm(y \sim x)$ data= Japan

⇒ residual plots (l_1)

data cleaning

1. feature
2. missing value (mean → Amelia / Mice)
3. weight . . .

总结.

日本. . exp($\hat{\beta}_0 + \hat{\beta}_1 V_1 + \hat{\beta}_2 V_2$)

Hochmaier 2014

$T_0 = 4.5°C$

$$GDD = \sum_{t=t_0}^{t_1} (T - T_0)$$

CDD $\boxed{\text{Dormancy time}}$

1050 ~ 1900 h	< 7°C	1982~1985
1081 ~ 1214 h	3.2–3.7°C	2000
• 1200 h	≤ 6°C at	→ 2008

$\boxed{\text{Base temperature} \quad T_0 = 4.5°C}$

4°C	Sour cherry	1980	
• 4.5°C	Iezzoni	1985	← bloom

$\boxed{\text{Chilling Unit}}$

Temp (°C)	<2.5	2.5 ~ 9.2	9.2 ~ 12.5	12.5 ~ 16	16 ~ 18	> 18
CU	1	1	0.5	0	−0.5	−1

Empirical CU 1615 (SD = 161)
1550 (SD = 118) ⇒ CU ≈ 1600 ± 150

$\boxed{\text{Degree-Day}}$

$$CDD = \sum_{t=101}^{t} (T - T_0) \Rightarrow 開花.$$

Bing	Van	Lapins	Sweetheart
805 (75)	765 (66)	852 (64)	916 (88)

1. Avg t − T_0

2. $\frac{\text{Low t} + \text{high t}}{2} - T_0$

Cumulative degree day

$$CDD_t = \sum_{t=t_0}^{t} (T_t - T_0)$$

t_0 starting day of CDD $\boxed{t_0 = 1\ ?}$

T_0 4.5°C

T_t Average daily temperature of day t

CDD_{th} Threshold CDD_{th} at which cherry blossom satisfy the requirement.

When $CDD_t \geq CDD_{th} \Rightarrow t_b$ blossom time

$$t_b = \underset{t}{\operatorname{argmin}} (CDD_t > CDD_{th})$$

Time series model for temperature estimation T_t

$\Leftrightarrow$ - - - - - - - - - - - - - - CDD_t

Assume T_t follows $AR(p)$

$$p_t = P(CDD_t > CDD_{th})$$

$$\log\left(\frac{p_t}{1-p_t}\right) = \beta_0 + \beta_1 T_t$$

数据.

$z_t = 1$ If $CDD_t \geq CDD_{th}$

$z_t = 0$ If $CDD_t < CDD_{th}$

Cumulative degree day

$$CDD_t = \sum_{t=t_0}^{t} (T_t - T_0)$$

t_0 starting day of CDD $\boxed{t_0 = 1\,?}$

T_0 $4.5°C$

T_t Average daily temperature of day t

CDD_{th} Threshold CDD_{th} at which cherry blossom satisfy the requirement.

When $CDD_t \geqslant CDD_{th} \Rightarrow t_b$ blossom time

$$t_b = \underset{t}{\operatorname{argmin}} (CDD_t > CDD_{th})$$

Time series model for temperature estimation T_t

$\Leftrightarrow$ - - - - - - - - - - - - - - CDD_t

Assume T_t follows $AR(p)$

$$p_t = P(CDD_t > CDD_{th})$$

$$\log\left(\frac{p_t}{1-p_t}\right) = \beta_0 + \beta_1 T_t$$

数据.

$z_t = 1$ If $CDD_t \geqslant CDD_{th}$

$z_t = 0$ If $CDD_t < CDD_{th}$

location = l = koyota

Cumulative degree day

$$CDD_t = \sum_{t=t_0}^{t} (T_t - T_0)$$

$t_0 = 1$ starting day of CDD $\boxed{t_0 = 1\,?}$ ①

T_0 4.5°C

T_t Average daily temperature of day t

CDD_{th} Threshold CDD_{th} at which cherry blossom satisfy the requirement. ②

800

When $CDD_t \geq CDD_{th} \Rightarrow t_b$ blossom time

T_t.

$$t_b = \underset{t}{\text{argmin}}\,(CDD_t > CDD_{th}) \quad (1)$$

Time series model for temperature estimation $\boxed{T_t}$

⟺ - - - - - - - - - - - - - - CDD_t

Assume T_t follows AR(p) ③ ZXY.

$$\boxed{P_t = P(CDD_t > CDD_{th})}$$

$P_t \in (0,1)$ $\frac{(0,1)}{(0,1)}$

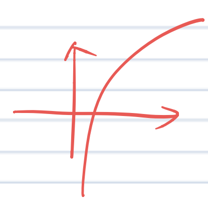
$\frac{P_t}{1-P_t} \in (0, +\infty)$

☆ $$\text{logit}(P_t) = \log\left(\frac{P_t}{1-P_t}\right) = \underbrace{\beta_0 + \beta_1 T_t}_{(-\infty,+\infty)} \quad (2)$$

数据.

$Z_t = 1$ If $CDD_t \geq CDD_{th}$

$Z_t = 0$ If $CDD_t < CDD_{th}$



$\log\left(\frac{P_t}{1-P_t}\right) \in (-\infty, +\infty)$

$P_t \in (0,1)$

$$\text{logit}(P_t) = \underbrace{\beta_0}_{\text{intercept}} + \underbrace{\beta_1}_{\text{slope}} T_t$$

train.

T 温度

t 时间

K1 Train   K1 Test   test.

2022/2/28

T_t T_{train} = observed T_{test} (AR(p) <ii>)

<ii> $T_t \sim AR(p)$

<i> Z_t. = T_t → $\hat{\beta}_0$ $\hat{\beta}_1$

K1Train Z_t, T_t. → $\hat{\beta}_0$ $\hat{\beta}_1$

K1Test $\cancel{Z_t}$. $\hat{Z}_t = \hat{\beta}_0 + \hat{\beta}_1 T_t$

Z_t $\hat{Z}_t$ → Accuracy.

M_B. Best model = $\arg\min_{\beta} Err(Z_t, \hat{Z}_t)$

<iii> M_B.

T_{test}

$\hat{Z}_{test}$ = $M_B \cdot T_{test}$ →

glm(Zt ~ Tt, link = "binomial")

(a) CDD_t

(b)

1   T_{1trd} T_{1test}

2.

K. $T_{k\,train}$ $T_{k\,test}$

1953

1954

1954

(c) Z_t = $CDD_t \geqslant CDD_{th}$

Z_t, T_t → model (2) logit.

k上 train $\hat{\beta}_k$ $\rightarrow$ k上 test $\hat{z}t_k$

$\hat{p}_t$