Modele matematyczne; czy lepsze są lepsze?

P. Wojtaszczyk [1]

Instytut Matematyczny PAN

Abstrakt Odczyt bedzie się składał z dwu części. Pierwsza filozoficzna: Komputery są coraz mocniejsze a więc się zatykają–trzeba im pomóc. Druga to konkretny opis pewnej zwariowanej metody liczenia średniej temperatury na Ziemi i nie tylko.

[1] grant UMO-2016/21/B/ST1/00241.

Oczywistość

W ostatnich latach (dziesięcioleciach) nastąpiło

1. Nastapił wielki wzrost mocy obliczeniowej komputerów
2. Nastapił wielki wzrost ilości danych–wynik rozwoju różnych sensorów (czyli też komputerów)
3. Ciągle ciągle stosujemy starą zasadę

Daj kurze grzędę, a ona: wyżej siędę

czyli chcemy te dane uwzględnić w modelu.

Tak naprawdę nie jest to nic nowego.

Przykład pseudohistoryczny

Wyobraźmy sobie, że chcemy opisać model lotu pocisku artyleryjskiego

1. W czasach wojen napoleońskich: równania Newtona, kąt nachylenia lufy i prędkość wylotu z lufy-problem jest dwuwymiarowy
2. Ale już w czasie wojny działa morskie miały nośność do 40-50 km. Dochodzi: opór powietrza–równania aerodynamiki pocisk zaczyna wirować i przestaje być kulą, cechy atmosferyczne –gęstość powietrza, wiatry
3. O rakietach balistycznych nic nie powiem, ale w oczywisty sposób sprawa się komplikuje.

Model matematyczny

Ma on najczęściej trzy składniki

1. Układ równań, najczęściej różniczkowych który opisuje naszą ogólna wiedzę o procesie

2. Dane–to co trzeba wstawić do tych równań, żeby miały rozwiazanie.

3. Wynik czyli rozwiazanie.

Układ równań $D(\mu, x) = 0$. μ to dane, elementy pewnej przestrzeni Banacha $\mathcal{V}$ i dostajemy rozwiazanie $x = x_\mu$ należące do pewnej przestrzeni Banacha $\mathcal{X}$.

Jak używamy modelu?

Model matematyczny

Ma on najczęściej trzy składniki

1. Układ równań, najczęściej różniczkowych który opisuje naszą ogólna wiedzę o procesie

2. Dane–to co trzeba wstawić do tych równań, żeby miały rozwiazanie.

3. Wynik czyli rozwiazanie.

Układ równań $D(\mu, x) = 0$. μ to dane, elementy pewnej przestrzeni Banacha $\mathcal{V}$ i dostajemy rozwiazanie $x = x_\mu$ należące do pewnej przestrzeni Banacha $\mathcal{X}$.

Jak używamy modelu?

Przekleństwo wymiaru

Standardowe twierdzenia w teorii równań różniczkowych wyglądają tak:

Theorem

Jeżeli $D(.,.)$ jest w miare przyzwoite, oraz $\mathcal{V}$ i $\mathcal{X}$ są odpowiednimi przestrzeniami funkcji gładkich to dla $\mu \in \mathcal{V}$ istnieje rozwiązanie $x_\mu \in \mathcal{X}$ i odwzorowanie $\mu \mapsto x_\mu$ jest ciagłe z $\mathcal{V}$ do $\mathcal{X}$.

Problem Jeżeli chcemy, żeby nasz model uwzględniał wiele czynników to μ a czasami i x jest funkcją bardzo wielu zmiennych.

Theorem (Novak Woźniakowski–J.Complex. 25)

Dla dowolnego układu punktów $(x_j)_{j=1}^m \subset [0,1]^d$, gdy $m = 2^{\lfloor d/2 \rfloor} - 1$, istnieje funkcja f na $[0,1]^d$ taka, że $f(x_j) = 0$ dla $j = 1, \ldots, m$ i $\|f\|_\infty = 1$, **której wszystkie pochodne istnieją i są ograniczone przez 1.**

Przekleństwo wymiaru

Standardowe twierdzenia w teorii równań różniczkowych wyglądają tak:

Theorem

Jeżeli $D(.,.)$ *jest w miare przyzwoite, oraz* $\mathcal{V}$ *i* $\mathcal{X}$ *są odpowiednimi przestrzeniami funkcji gładkich to dla* $\mu \in \mathcal{V}$ *istnieje rozwiązanie* $x_\mu \in \mathcal{X}$ *i odwzorowanie* $\mu \mapsto x_\mu$ *jest ciagłe z* $\mathcal{V}$ *do* $\mathcal{X}$*.*

Problem Jeżeli chcemy, żeby nasz model uwzględniał wiele czynników to μ a czasami i x jest funkcją bardzo wielu zmiennych.

Theorem (Novak Woźniakowski–J.Complex. 25)

Dla dowolnego układu punktów $(x_j)_{j=1}^m \subset [0,1]^d$*, gdy* $m = 2^{\lfloor d/2 \rfloor} - 1$*, istnieje funkcja* f *na* $[0,1]^d$ *taka, że* $f(x_j) = 0$ *dla* $j = 1, \ldots, m$ *i* $\|f\|_\infty = 1$*,* **której wszystkie pochodne istnieją i są ograniczone przez** $\mathbf{1}$**.**

Świat nie jest taki zły

ale opis matematyczny bardzo zachłanny.

- Jest faktem eksperymentalnym, że daleko nie każda funkcja $\mu \in \mathcal{V}$ pojawia się jako sensowne dane, nie każda funkcja $x \in \mathcal{X}$ pojawi się jako rozwiązanie (przy sensownym μ).
- Cała chmara problemów, żeby sensownie praktycznie i matematycznie wprowadzić "małe" zbiory $V \subset \mathcal{V}$ oraz $X \subset \mathcal{X}$, że wystarczy rozwiązywać problem $D(\mu, x) =$ dla $\mu \in V$ oraz x prawie należy do X. *Jest to redukcja modelu.*
- Przykład – reduced basis method; znajdujemy X to m-wymiarowa podprzestrzeń liniowa $\mathcal{X}$.

Morał: Zwiekszenie mocy obliczeniowej zmusza nas do wymyślania "mniejszych" modeli.

Koniec filozofii.

Średnia temperatura powierzchni Ziemii

To co teraz nastąpi opera sie o dwie prace

S. Foucart, M. Hielsberg, G. Mullendore, G. Petrova P.W. Optimal Algorithms for Computing Average Temperatures, MATHEMATICS OF CLIMATE AND WEATHER FORECASTING *(2019)*

R. DeVore, S. Foucart, G. Petrova, P.W. Computing a quantity of interest from observational data, CONSTRUCTIVE APPROXIMATION (2018)

Średnia temperatura powierzchni Ziemi–teoria

Powierzchnia Ziemi $\mathcal{S}$ to w zasadzie sfera. $f(x,t)$ to temperatura w punkcie $x \in \mathcal{S}$ w czasie t. No to średnia w momencie t to całka

$$\int_{\mathcal{S}} f(x,t)\, d\sigma(x)$$

gdzie σ to unormowana miara powierzchniowa na $\mathcal{S}$. Średnia w okresie $[a,b]$ to

$$\frac{1}{b-a}\int_a^b \int_{\mathcal{S}} f(x,t)\, d\sigma(x)\, dt$$

Średnia temperatura powierzchni Ziemi–życie

Interesują nas również dane historyczne–kilkadziesiąt lat i więcej.

1. Tradycyjnie mówimy o temperaturze powietrza 1.5 metra nad ziemią.
2. Temperature mierzą stacje meteorologiczne, kiedyś kilka razy dziennie.

Tego się trzymamy. Mamy więc stacje w miejscach $(x_j)_{j\in L} \subset \mathcal{S}$ czyli mierzymy wartości $f(x_j, t_i^j)$. Z tych danych łatwo dostać średnie dzienne, tygodniowe itp w *każdej stacji osobno*.

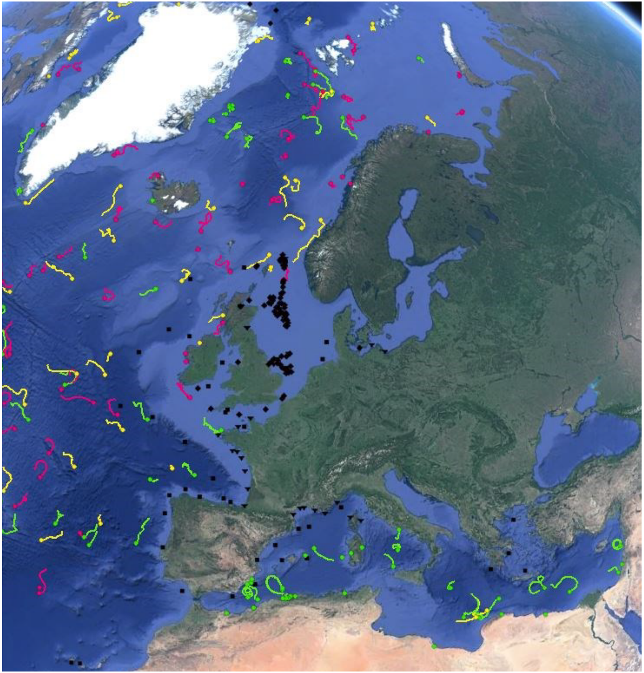
Problemy

- *Dane pochodzą z różnych żródeł, szczególnie morskie.*
- *Stacje są gdzie są a nie tam gdzie byśmy chcieli.*
- *Nie bardzo wiemy jak może wygladać f.*

Co robią klimatolodzy?

Na ogół rozpatrują t.zw. "anomalie". Ustalają okres n.p. 1951-1980 i liczą średnią w tym okresie dla każdej stacji–"normalna" temperatura dla stacji. Anomalie to odchylenie od tej "normalnej"

1. Ujednolicaja dane z różnych źródeł.
2. Wprowadzają poprawki n.p. na miasto, na strefę czasową, na metodologie pomiarów, szczególnie pomiarów na morzu.
3. Obliczają (różne algorytmy) temperaturę dla "dokładnej siatki" – 8000 punktów i wtedy jakoś liczą.
4. To co żeśmy sprawdzali to w ostatecznym rachunku algorytmy sprowadzają się do znalezienia wag $(a_j)_{j\in L}$ wtedy średnia to $\sum_{j\in L} a_j f(x_j)$. Ale waga a_j może zależyć od wartości $\{f(x_i)\}_{i\in L}$.



Boje dryfujące i stałe (również wieże wiertnicze), 29-12 września 2017, praca Poli et.al. *Automated surface marine observations from European data buoys*

Matematyka

Oznaczenia: $C(\mathcal{S})$ *to przestrzeń liniowa wszystkich funkcji ciągłych na* $\mathcal{S}$ *i norma to* $\|f\|_\infty := \sup_{x\in\mathcal{S}} |f(x)|$.
Zadane punkty $(x_j)_{j\in L} \subset \mathcal{S}$ i mamy m wymiarową przestrzeń V funkcji (ciągłych) na $\mathcal{S}$. Wtedy istnieje (prosty) algorytm który produkuje

1. m elementowy podzbiór $B \subset L$
2. liczby $(a_j)_{j\in B}$
3. stałą C_L

takie, że

$$\left| \int_{\mathcal{S}} f(x)d\sigma(x) - \sum_{j\in B} a_j f(x_j) \right| \leq C_L \inf\{\|f - v\|_\infty \ : \ v \in V\}.$$

Matematycznie jest to optymalne oszacowanie.

Wracamy na Ziemię

$\mathcal{S}$ to powierzchnia Ziemii, temperatura $f(x,t)$ to funkcja ciągła, x_j lokalizacje stacji pomiarowych. Trzeba mieć V–bierzemy z sufitu V to przestrzeń wielomianów harmonicznych stopnia ≤ 9. Wtedy wymiar V równa sie 100.

Eksperymenty numeryczne

Używamy dwu zbiorów danych

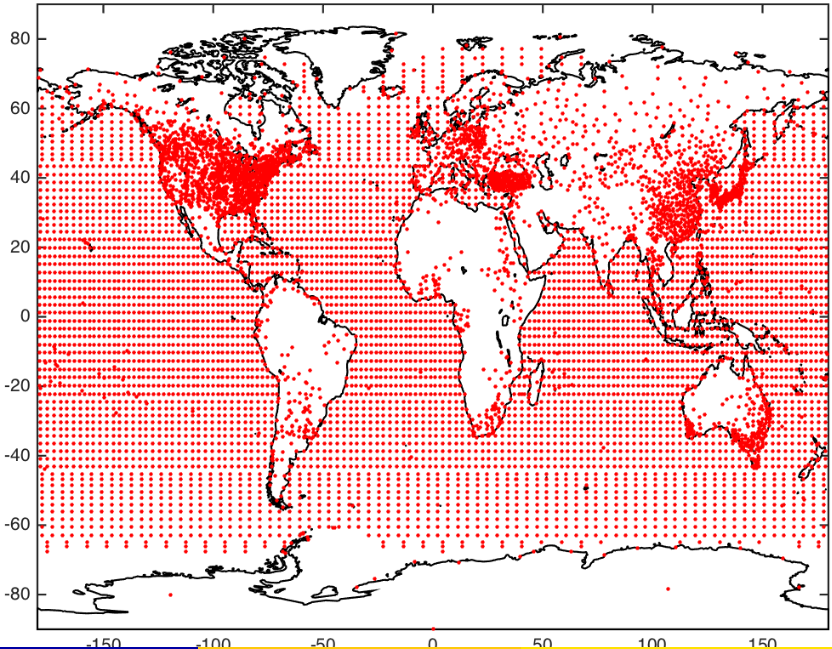
- Zbiór 1; jest to standardowy zbiór danych używany przez NASA Goddard Institut for Space Studies==GISTEMP oraz National Oceanic and Atmospheric Administration=NOAA. Są to dane przeliczone do “dokładnej siatki”.
- Zbiór 2; Na lądzie są to dane nieprzeliczone a na morzu przeliczone do “dokładnej siatki”

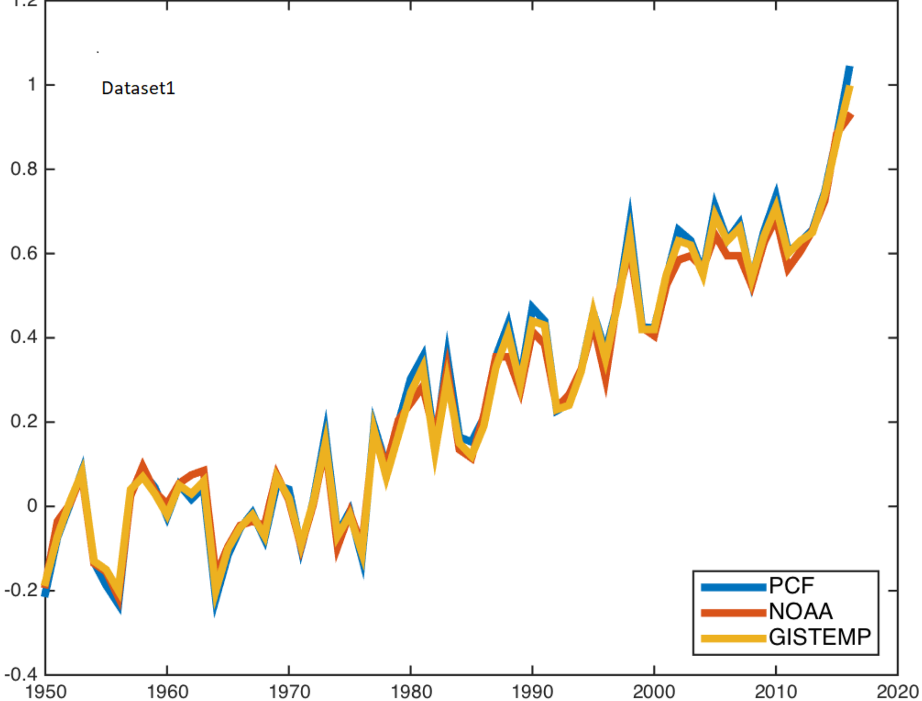
Używamy następujących przestrzenie V:

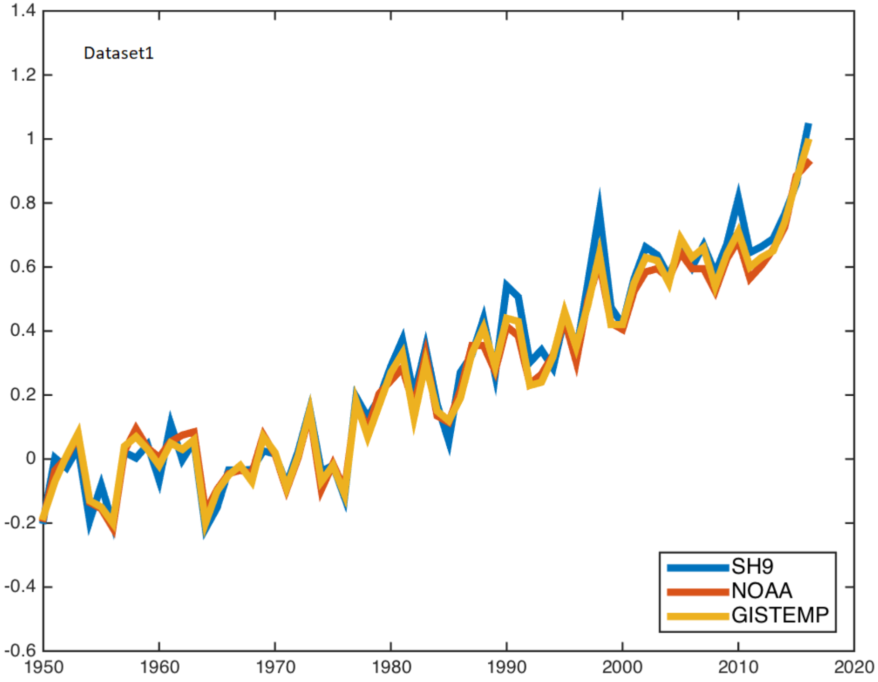
1. Przestrzeń SH9 wielomianów harmonicznych na sferze stopnia ≤ 9. Wtedy wymiar V równa sie 100.
2. Przestrzenie funkcji kawalkami stałych na 80 ‘kwadratach’ – PCC oraz na 8000 – PCF. Są to standardowe podziały.

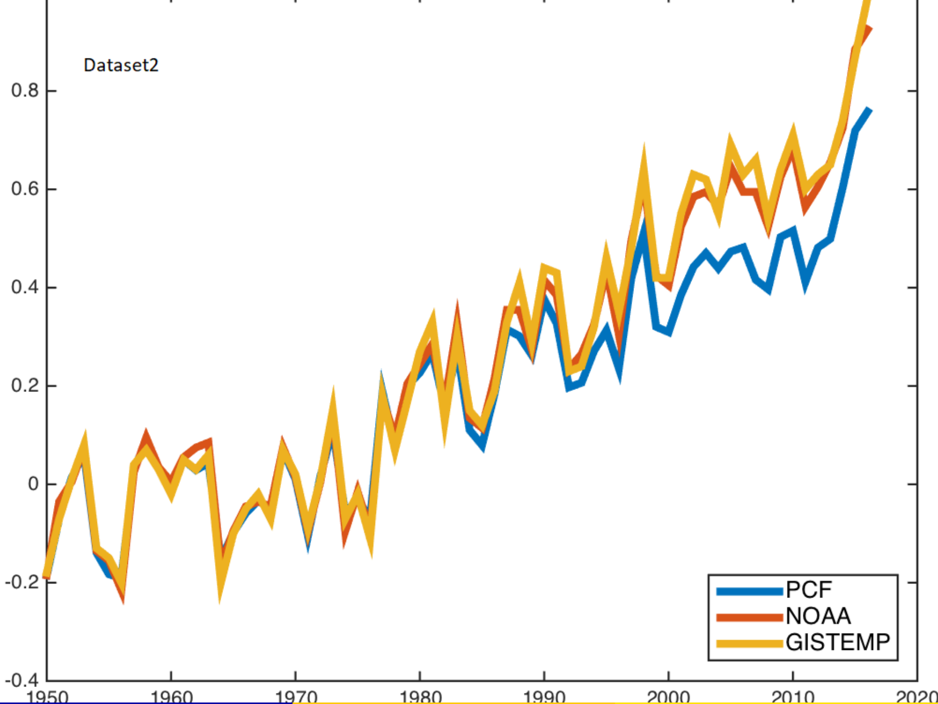
Porównujemy nasze wyniki z wynikami podanymi przez NASA oraz NOAA.

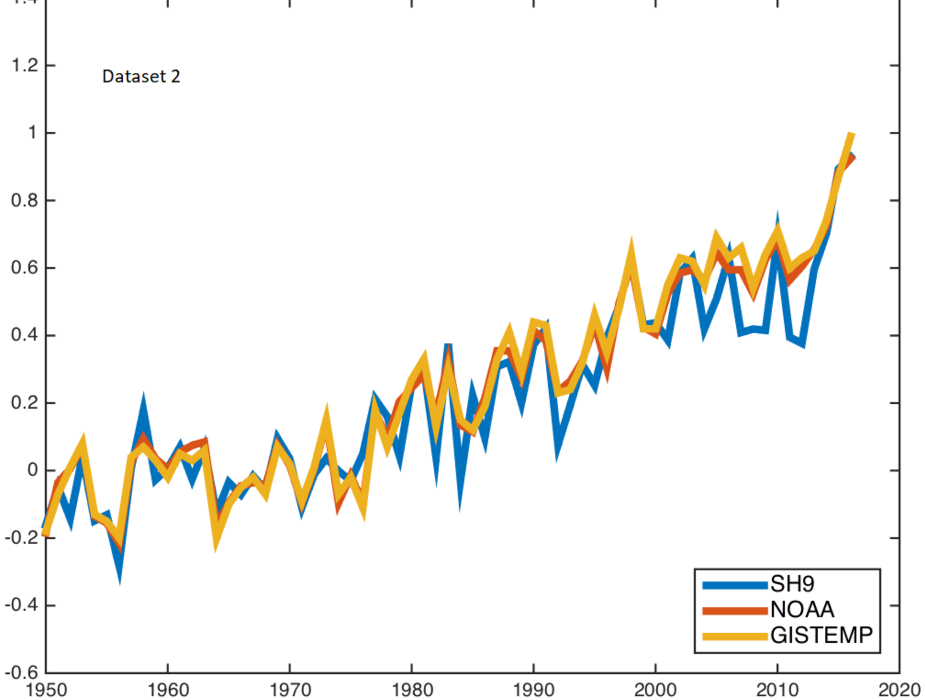
9187 locations in dataset 2




Dataset1
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
1950
1960
1970
1980
1990
2000
2010
2020
PCF
NOAA
GISTEMP


Dataset1
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1950
1960
1970
1980
1990
2000
2010
2020
SH9
NOAA
GISTEMP

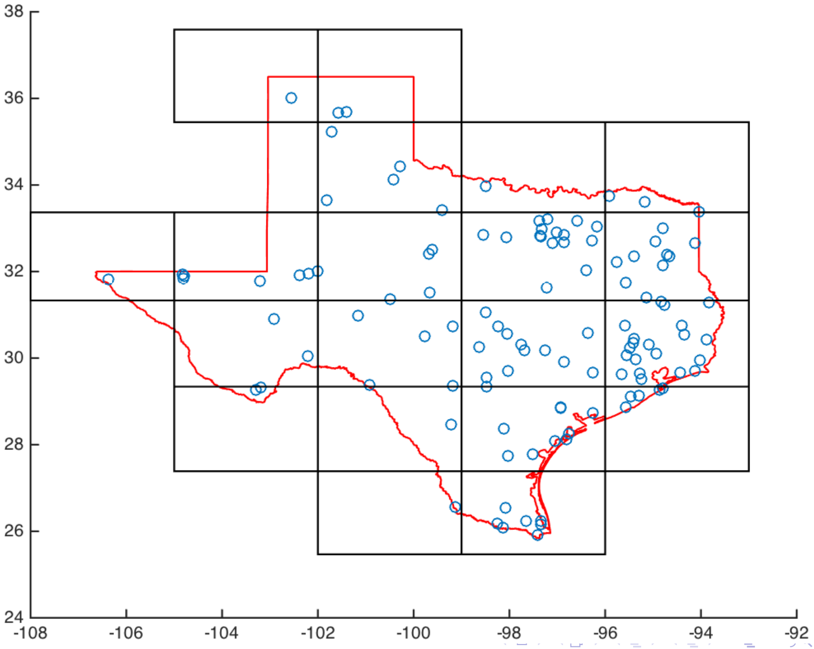

Dataset2
0.8
0.6
0.4
0.2
0
-0.2
-0.4
1950
1960
1970
1980
1990
2000
2010
2020
PCF
NOAA
GISTEMP

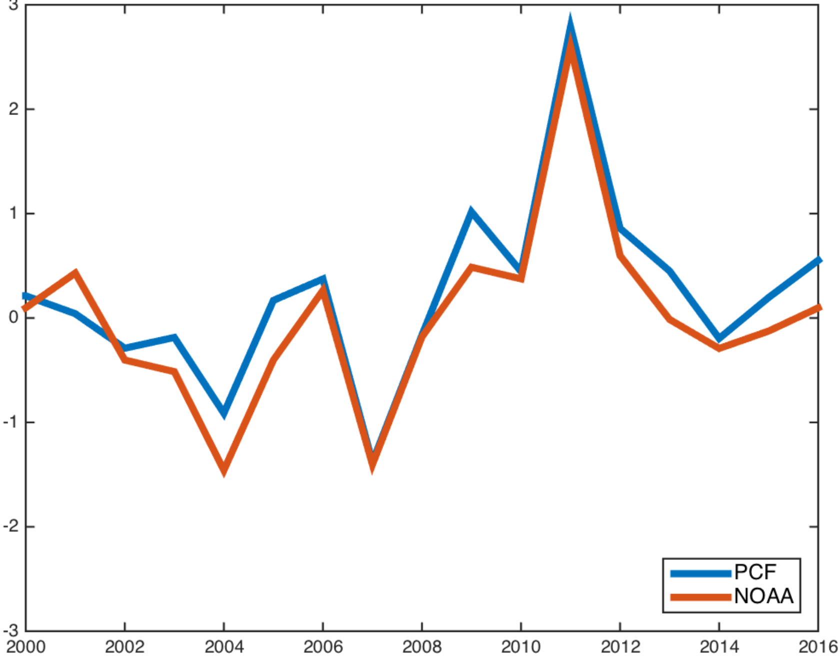


Dataset 2
SH9
NOAA
GISTEMP
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1950
1960
1970
1980
1990
2000
2010
2020

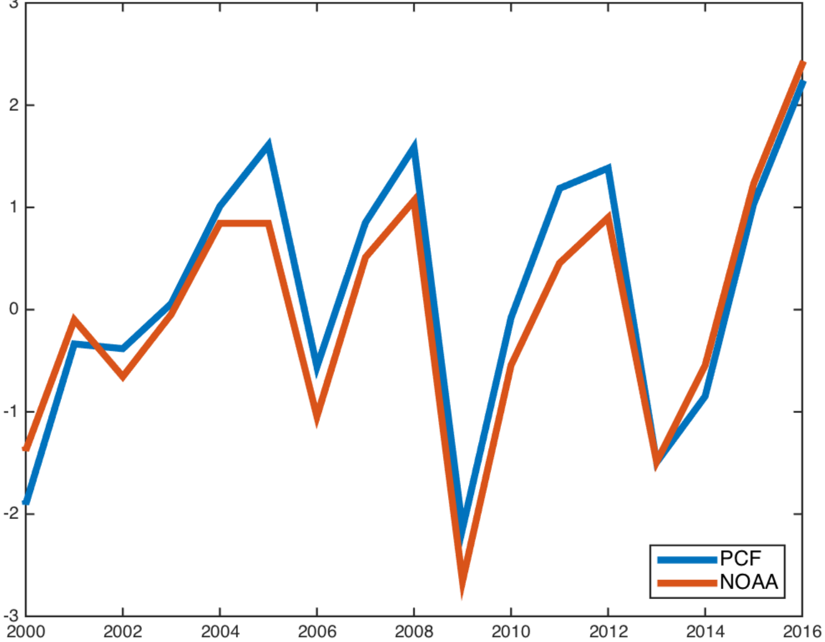
Temperatura w Texasie

Liczymy zimowe i letnie temperatury (trzy miesiące) w Teksasie. Znamy tygodniowe średnie z każdej stacji. V to przestrzeń funkcji $f(x,t)$ stałych na kwadratach "dokładnej siatki" PCF i kawałkami liniowych w czasie z węzłami co tydzień.
Mamy 115 stacji.


38
36
34
32
30
28
26
24
-108
-106
-104
-102
-100
-98
-96
-94
-92



Texas summer



Texas winter


Dziekuję za uwagę
Życzę miłej pogody