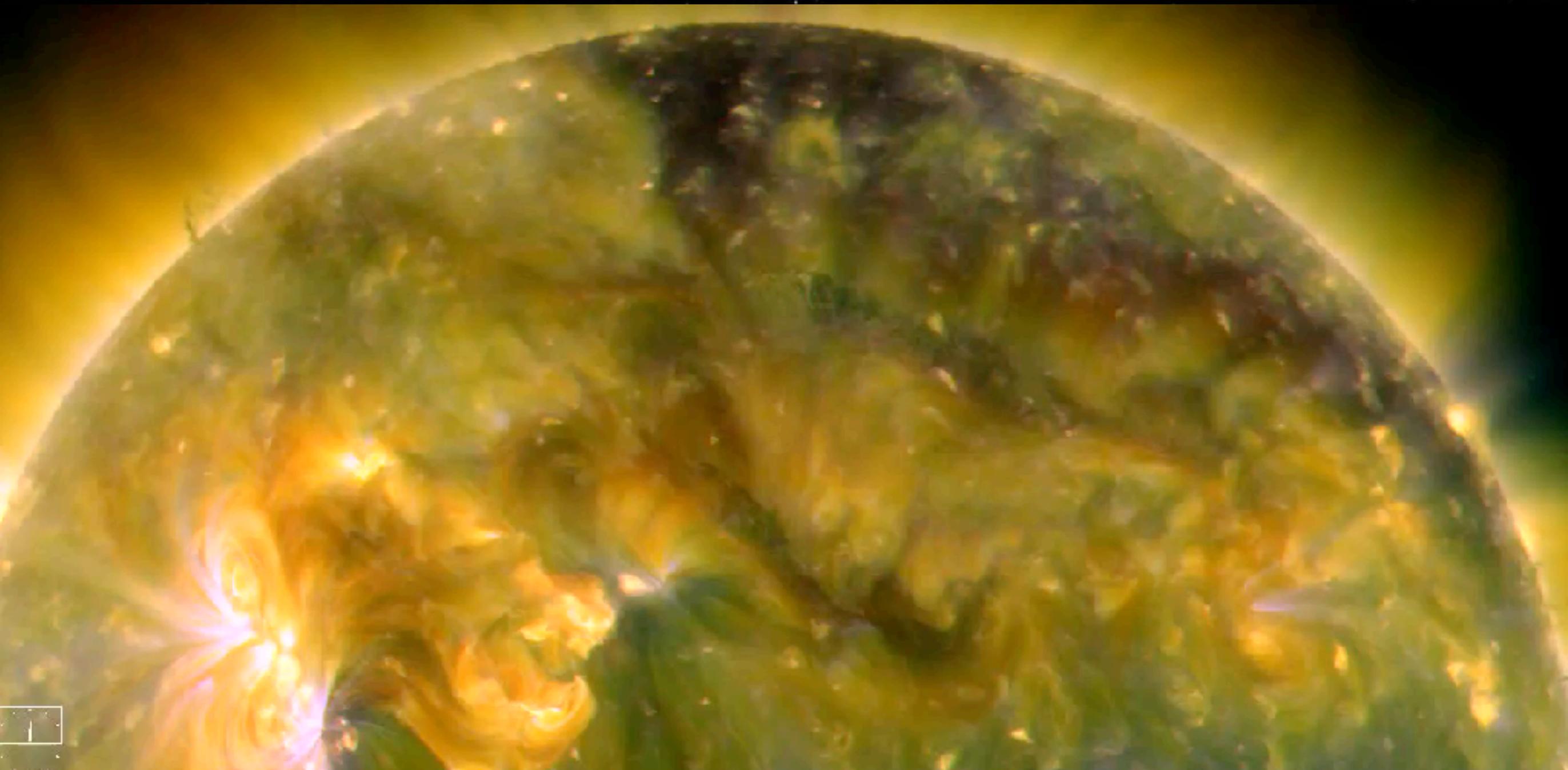


Thierry Dudok de Wit
(LPC2E, CNRS et Université d'Orléans)

+ Greg Kopp (Boulder), Claus Fröhlich (Davos),
Micha Schöll (Hamburg), Laure Lefèvre (Bruxelles)

Mesure du rayonnement solaire par satellite: de la fusion de données en jonglant avec les incertitudes



- Mesure de l'irradiance solaire : enjeux
- Pourquoi fusionner plusieurs mesures en une seule ?
- Stratégie employée
- Perspectives



surface : 1 m²
TSI : 1361 W

TSI : Irradiance solaire totale
= ensemble de la puissance rayonnée par le Soleil

Enjeux politiques

Le réchauffement de la planète

Apprenez **FRIENDS OF SCIENCE**.org
en dévotion



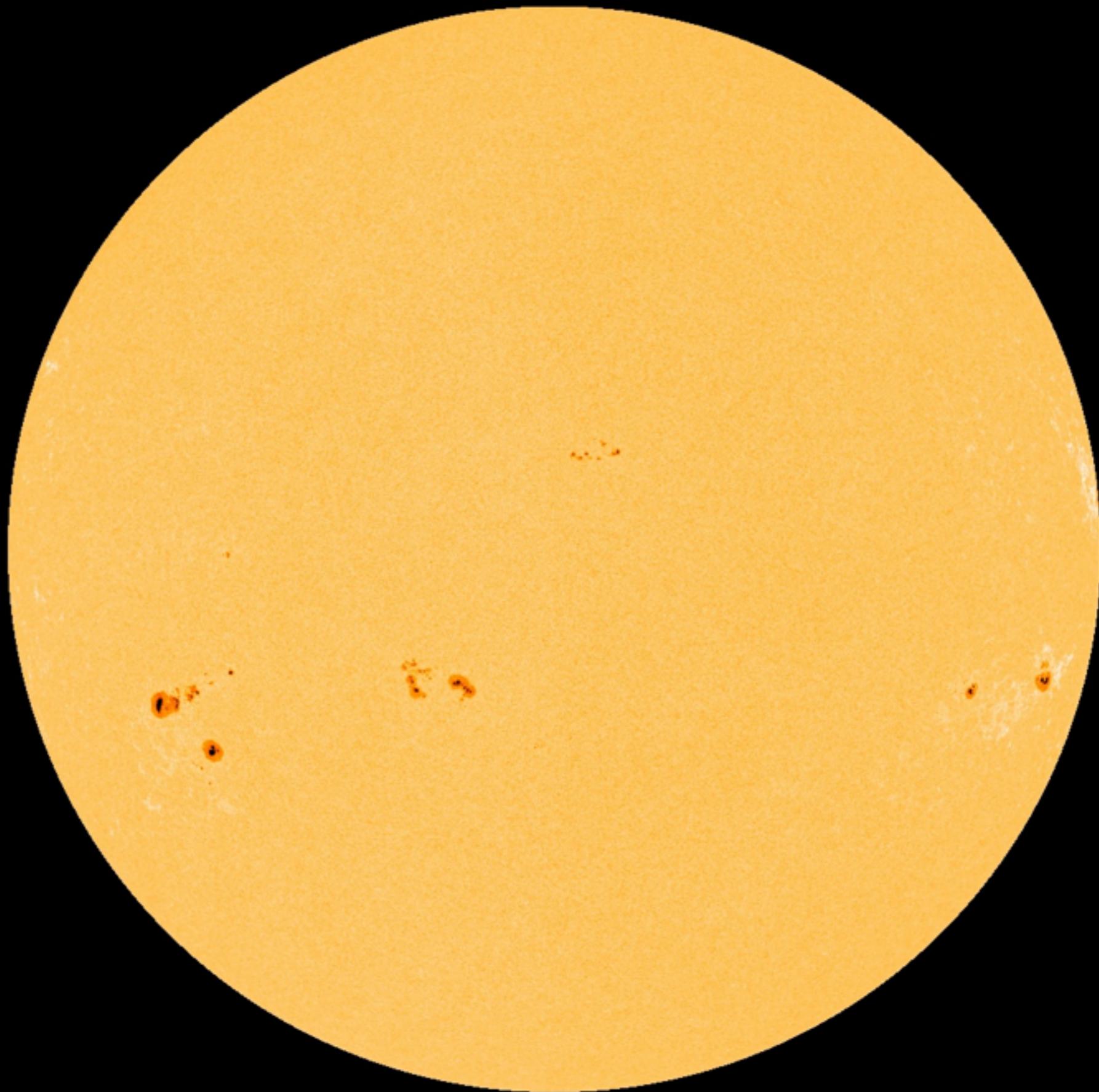
°C s'est arrêté naturellement il y a de cela plus de 16 ans.

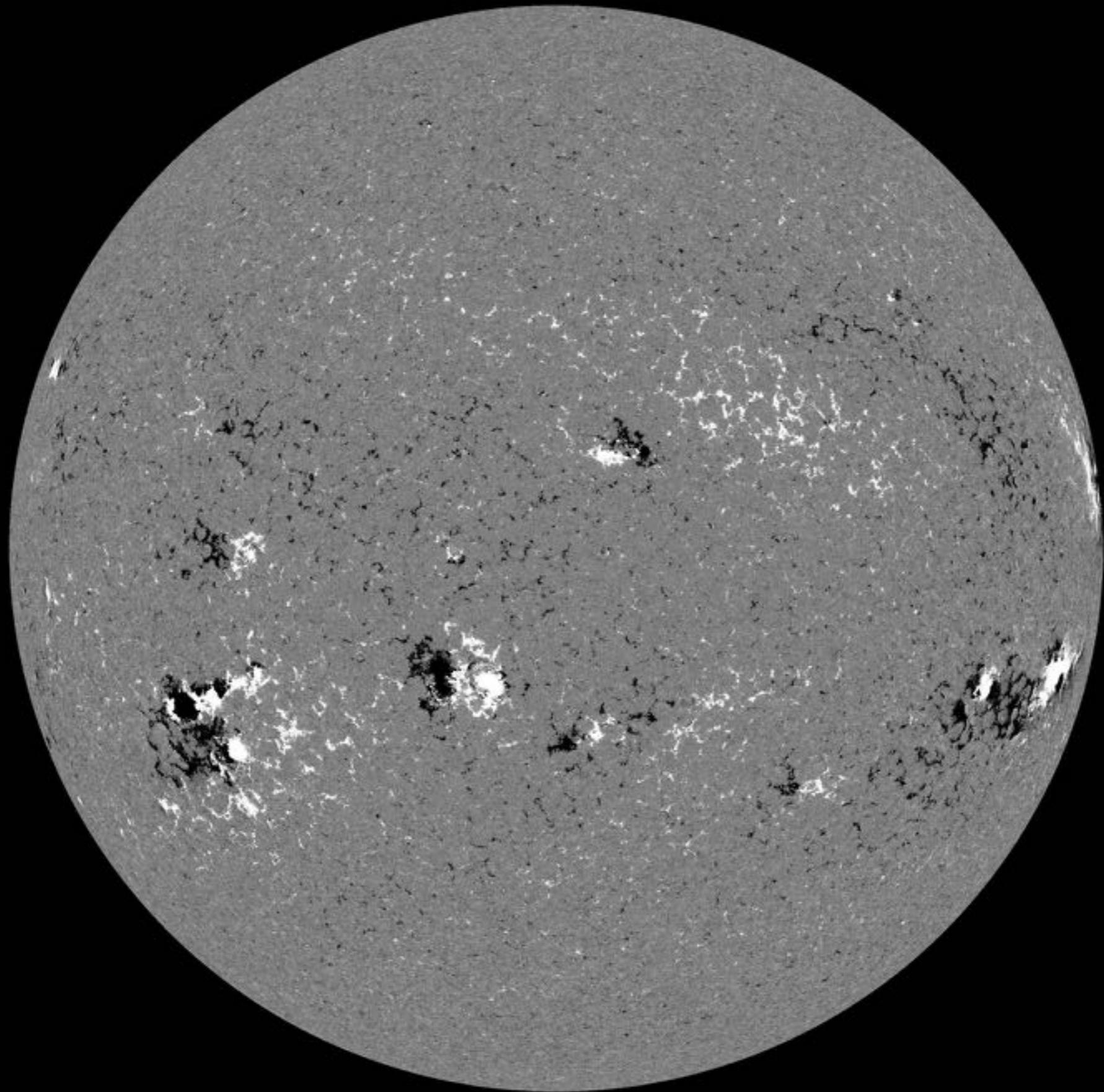


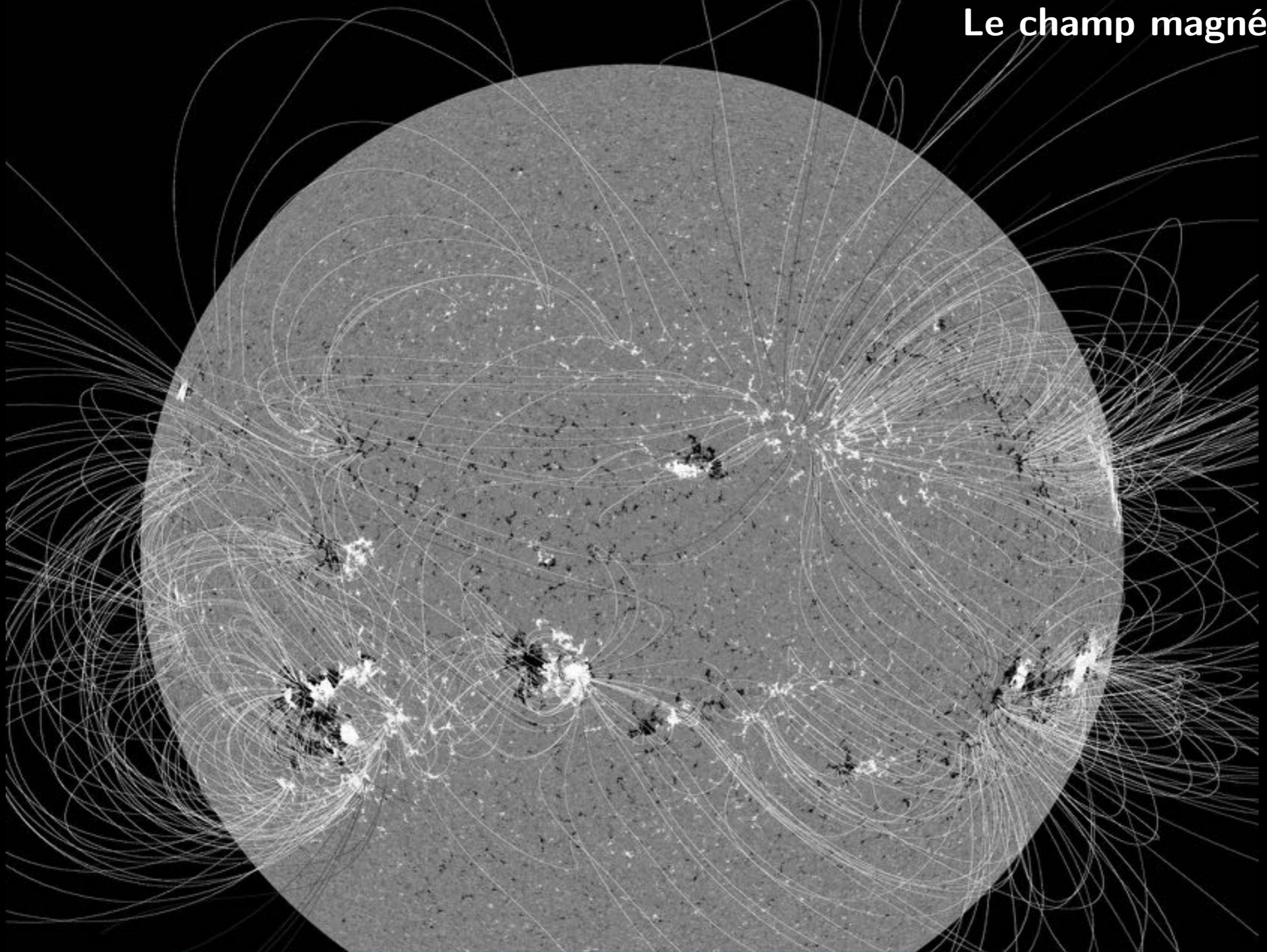
PATTISON

Enjeux scientifiques





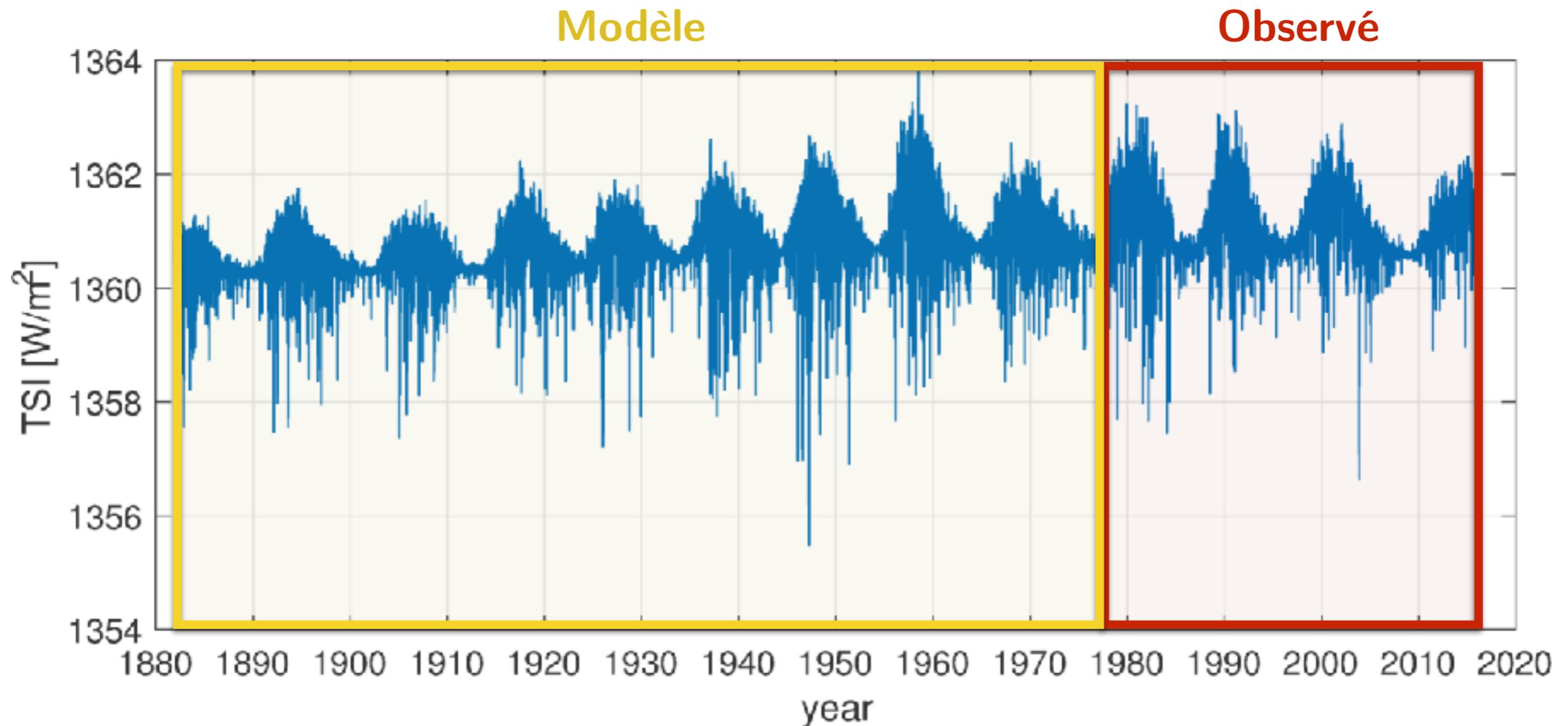




Comprendre comme varie la TSI équivaut à
comprendre le magnétisme solaire

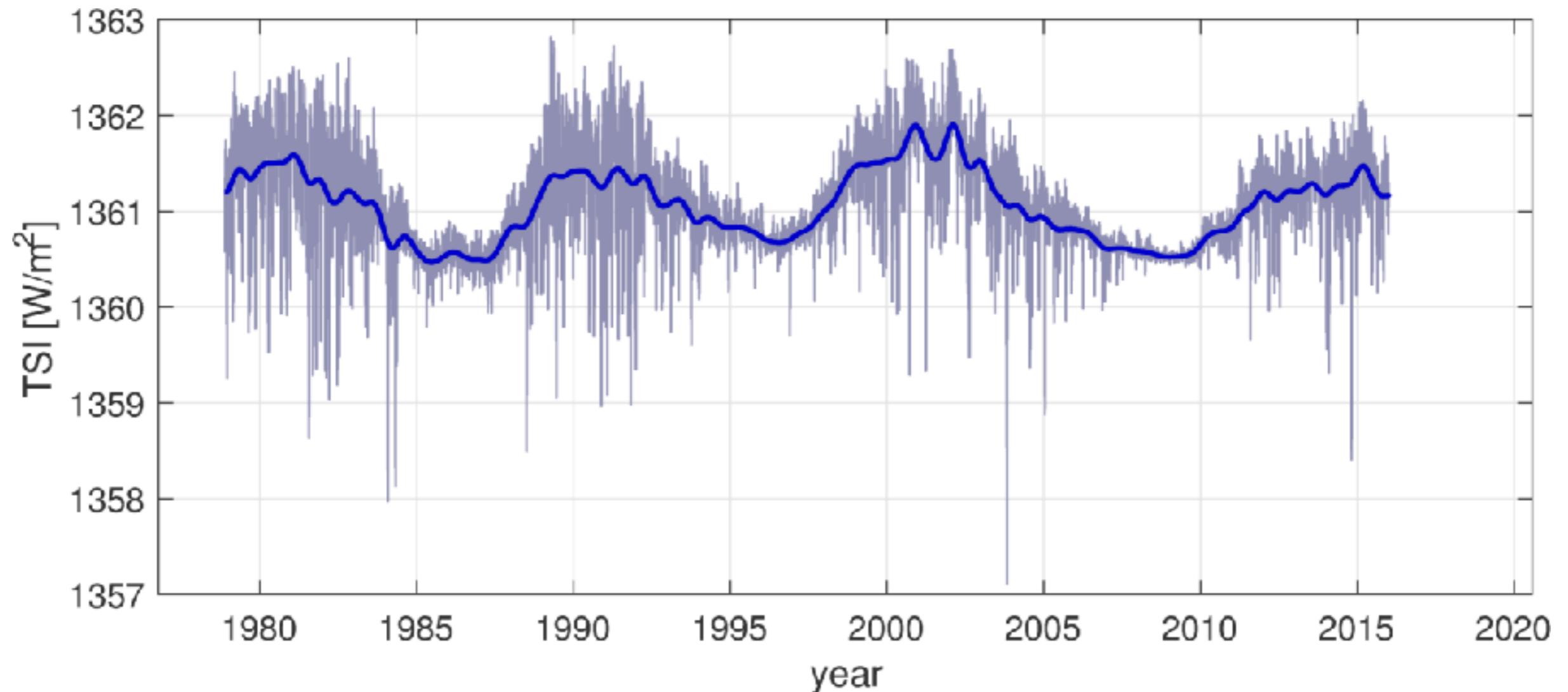
La TSI varie : de combien ?

Le Soleil (et la TSI) ont un cycle d'activité d'environ 11 ans

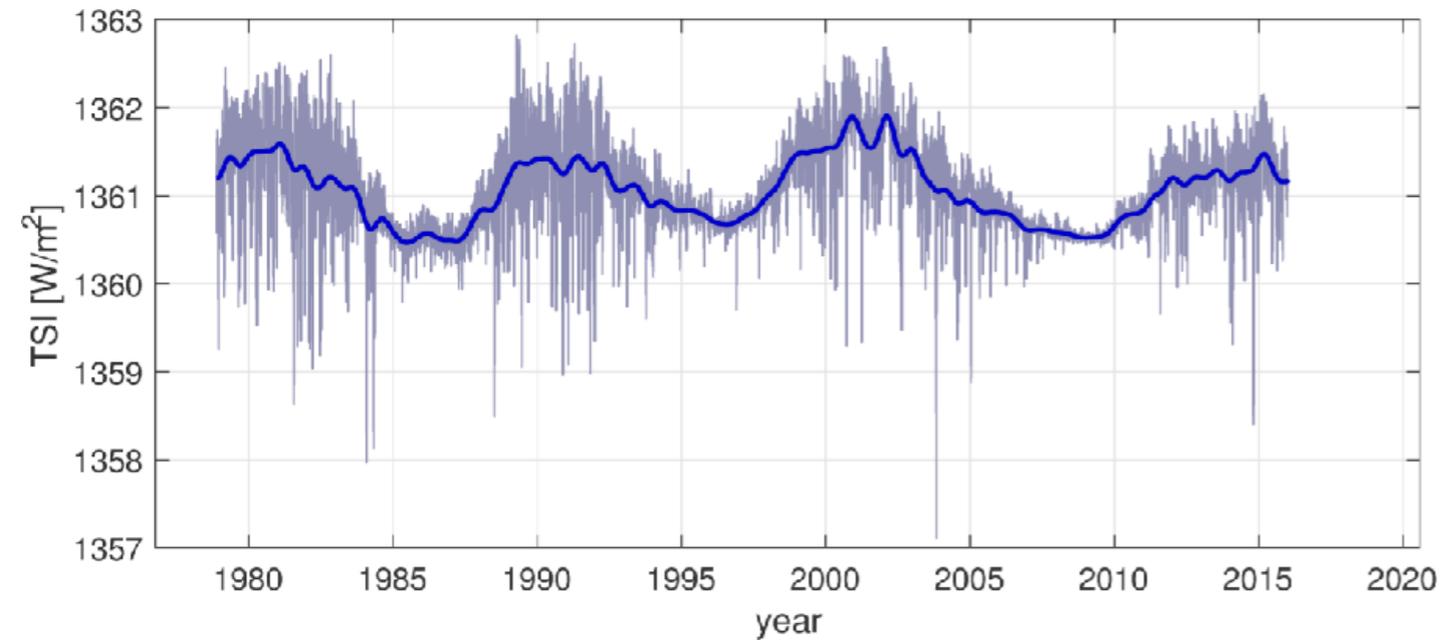


Les observations actuelles

- La TSI n'est mesurée que depuis 1978
- Elle varie peu au cours d'un cycle solaire ($< 0.1\%$)

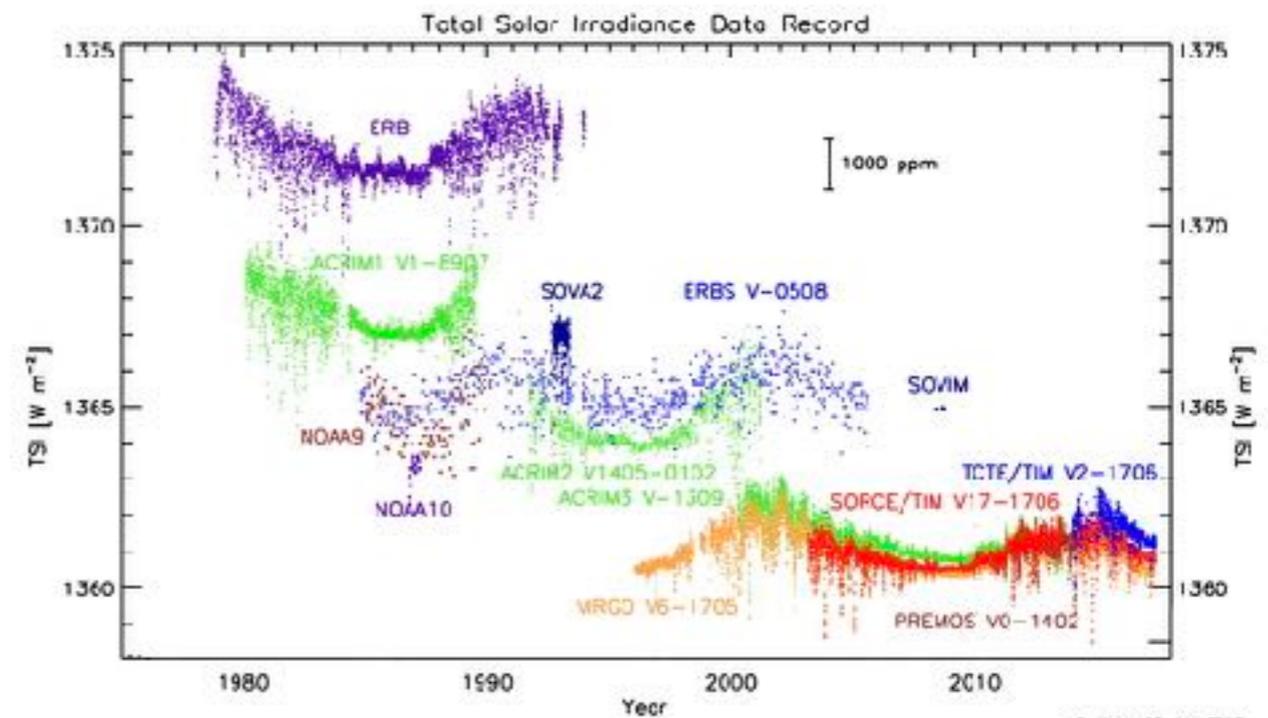


Les observations actuelles



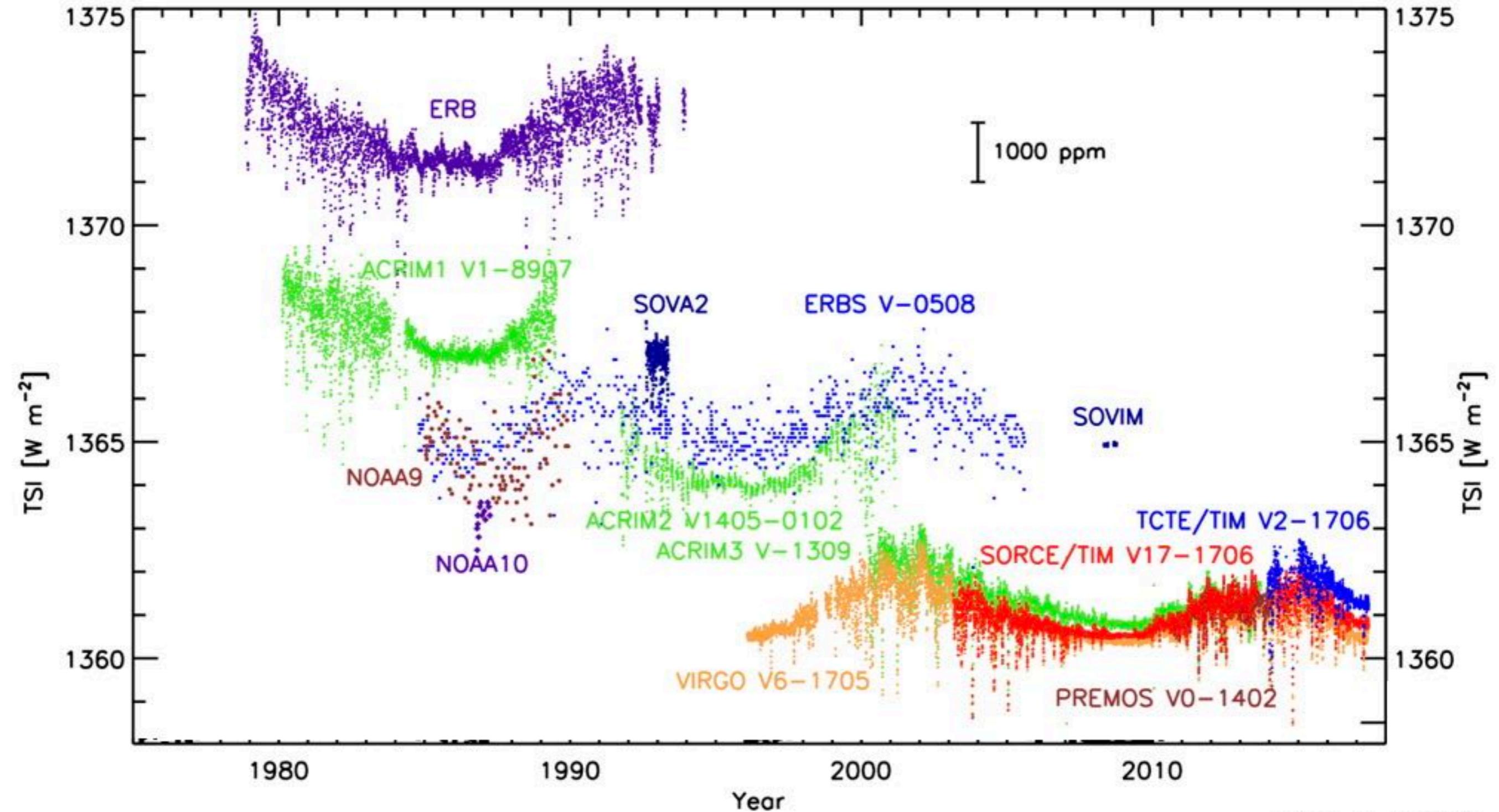
La mesure telle qu'on
la souhaiterait

La mesure telle
qu'elle est réellement



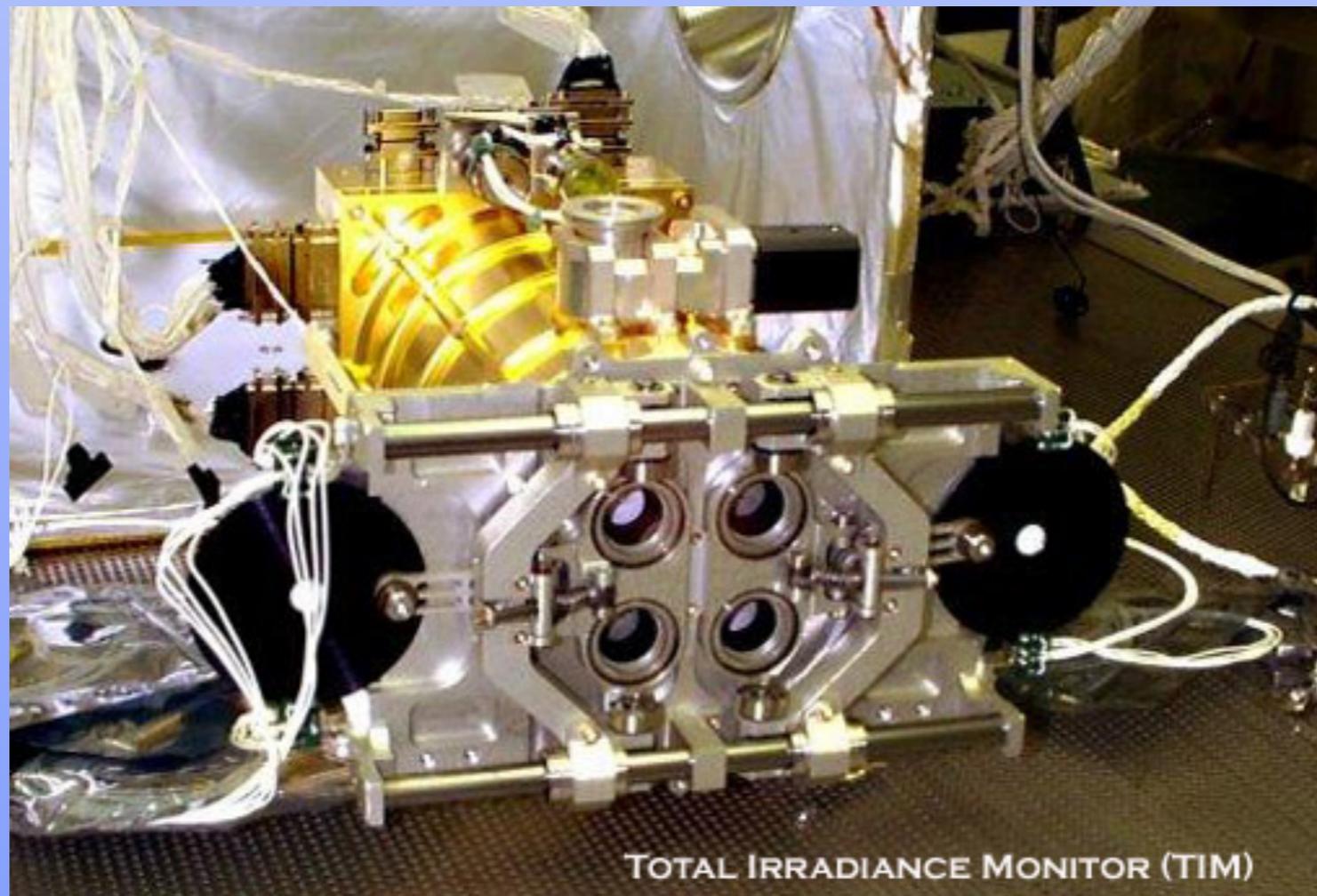
La réalité est plus complexe

Total Solar Irradiance Data Record



Interlude

- Ces mesures de TSI sont particulièrement difficiles à effectuer
 - étalonnage précis des instruments
 - dégradation des instruments → durée de vie < 10 ans
 - manque de coordination au niveau international



TIM (NASA)

- **Objectif:** **Fusionner** toutes les observations en une seule série temporelle composite avec des valeurs journalières et des incertitudes réalistes
- Requis : fiabilité, reproductibilité, traçabilité



- La **fusion de données** se rencontre dans de nombreux contextes.
- Exemple: *pansharpening*



Berne et al., IEEE IP (2012)

De nombreux obstacles

- les instruments sont en désaccord
 - comment trouver la meilleure “moyenne” ?

- les cadences d’observation sont différentes : de la minute au mois
 - comment tout ramener à une valeur journalière ?

- le nombre d’instruments qui observent par jour varie de 0 à 3
 - comment gérer les jours sans observations ?

- les incertitudes fournies avec les instruments ne sont pas comparables
 - quelles données favoriser ?

- et ce n’est pas tout...

Quelle stratégie ?



1. Estimer la précision de chaque instrument

- modèle autorégressif

2. Estimer les valeurs manquantes

- méthode d'expectation-maximization

3. Décomposer les données en différentes échelles des temps

- décomposition en ondelettes

4. Faire une moyenne pondérée des observations (échelle par échelle)

- Max de vraisemblance (puis Bayésien)

5. Estimer l'incertitude

- méthode de Monte-Carlo

Etape 1

Estimer la précision

1. Estimer la précision

■ **Objectif** : obtenir une mesure indépendante et fiable de la précision de chaque instrument

■ Modèle autorégressif

$$\hat{y}(t_k) = a_1 y(t_{k-1}) + a_2 y(t_{k-2}) + \dots + a_p y(t_{k-p})$$

$$\epsilon(t_k) = \hat{y}(t_k) - y(t_k)$$

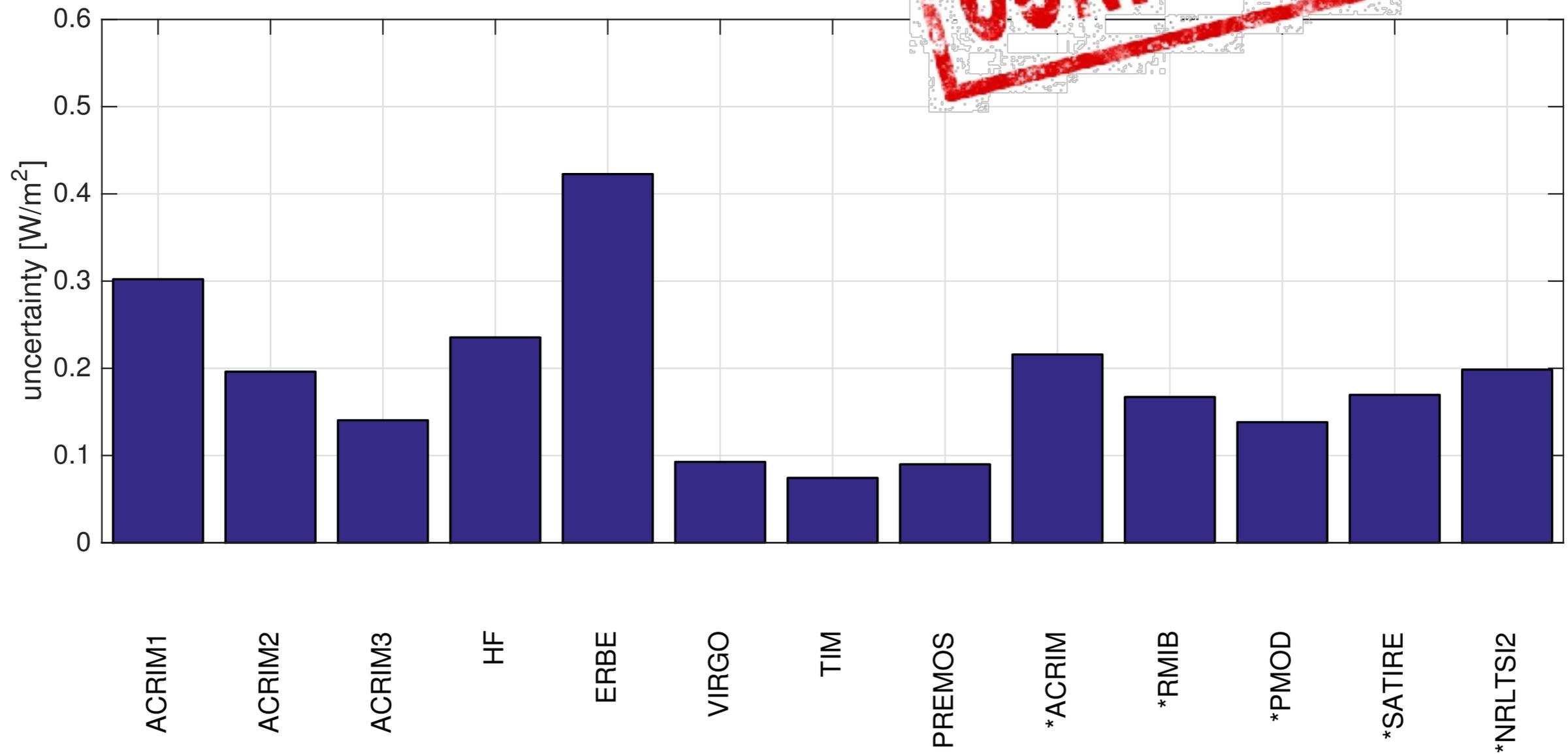
The logo for TSI (Technical Systems International) is a dark red, rounded rectangular shape with a white border. It is positioned to the right of the equations, with a small white triangle pointing towards the equation for $\hat{y}(t_k)$.

■ L'erreur résiduelle $\epsilon(t)$ capte les fluctuations aléatoires d'origine solaire et instrumentale → estimateur de la précision σ_y

■ (A partir de la précision nous avons estimé l'incertitude à différentes échelles de temps.)

1. Estimer la précision

■ Précision moyenne de chaque instrument



Etape 2

Estimer les valeurs manquantes

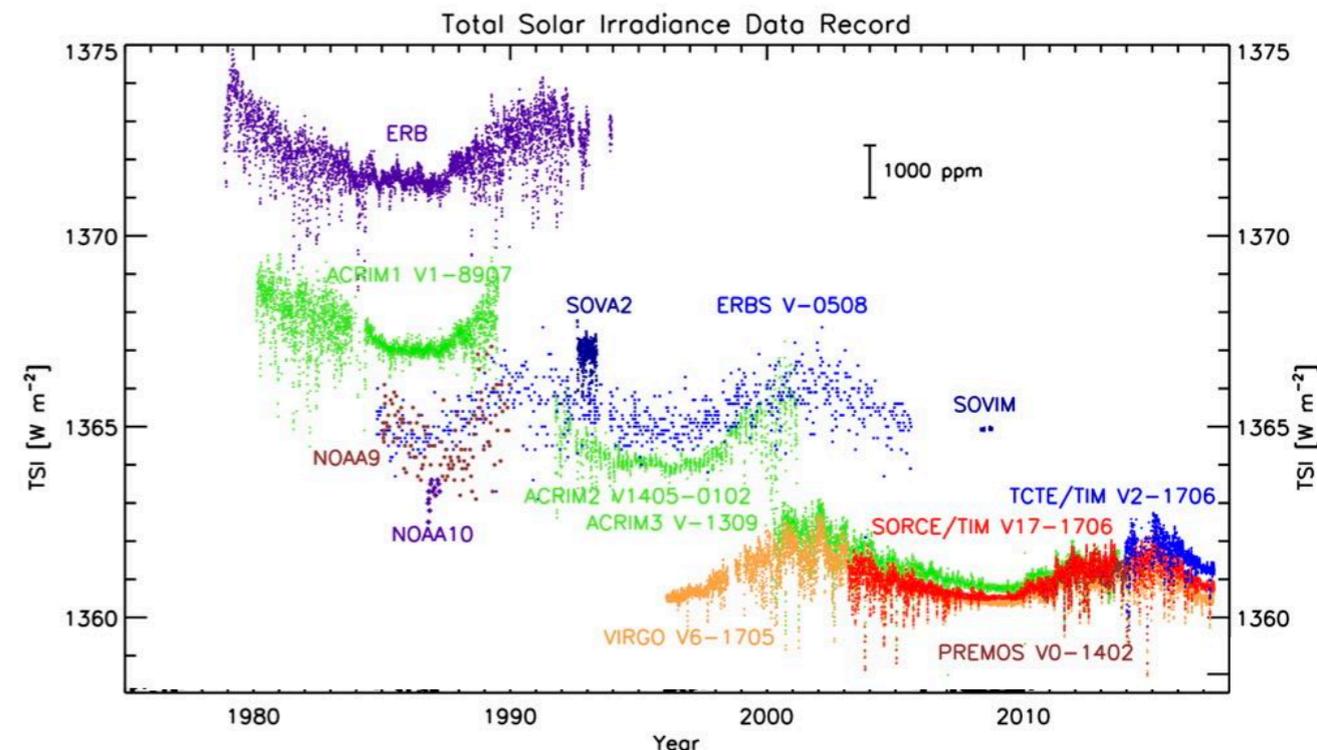
2. Estimer les valeurs manquantes

- **Objectif** : obtenir des séries temporelles (cadence = 1 jour) sans valeurs manquantes: requis pour effectuer la transformation en ondelettes.
- Comme les mesures sont redondantes, en première approximation

$$y_k(t) = \bar{y}_k + \alpha_k f(t)$$

TSI de
l'observateur k

la "vraie" TSI



2. Estimer les valeurs manquantes

- Approximations successives

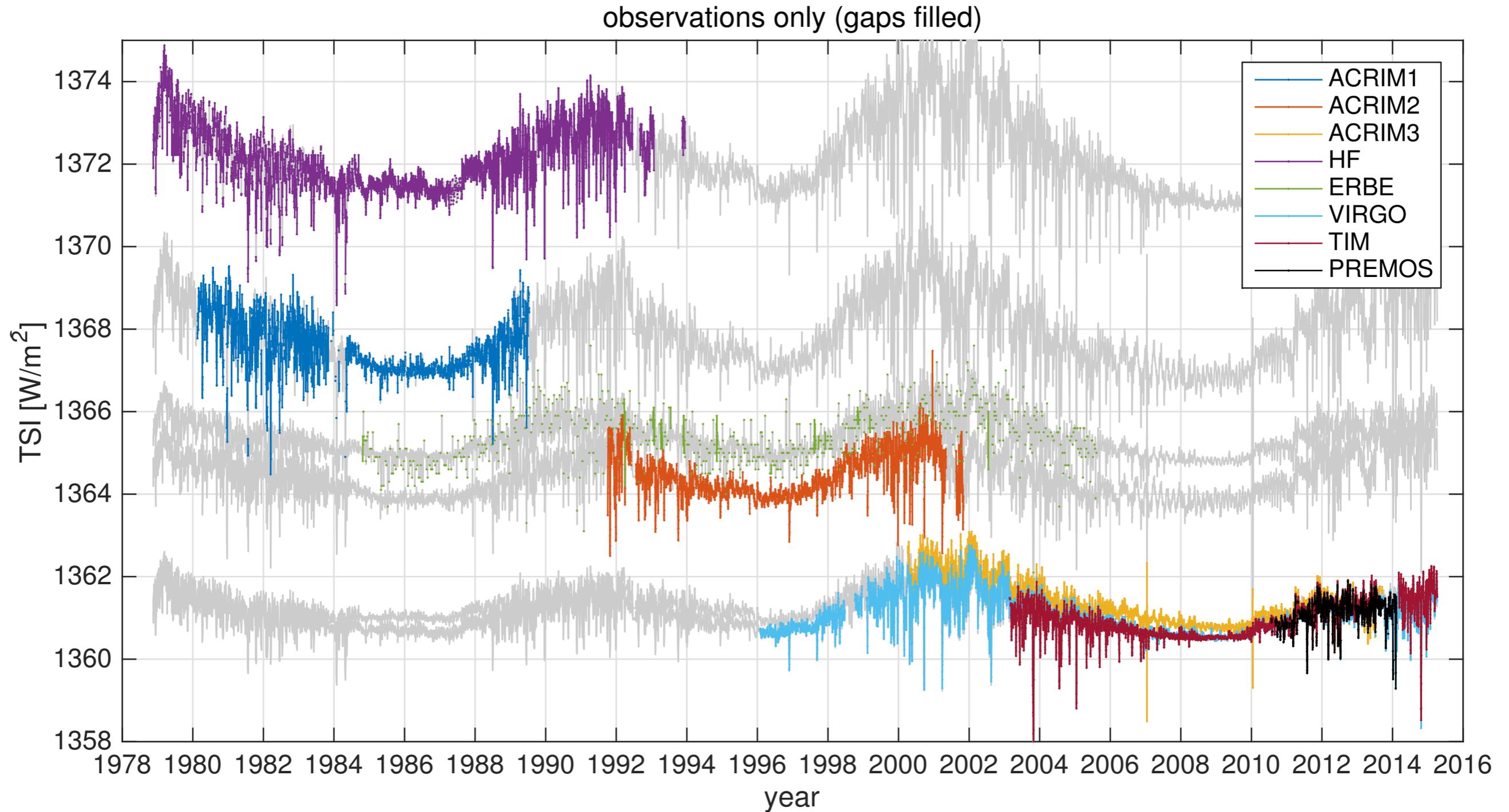
$$y_k(t) = \bar{y}_k + \alpha_k f(t)$$

$$y_k(t) = \bar{y}_k + \alpha_k f(t) + \beta_k g(t) + \dots$$

- Les approximations sont obtenues par SVD (singular value decomposition)

- Validation : en éliminant délibérément certaines valeurs et en testant la reconstruction

2. Estimer les valeurs manquantes

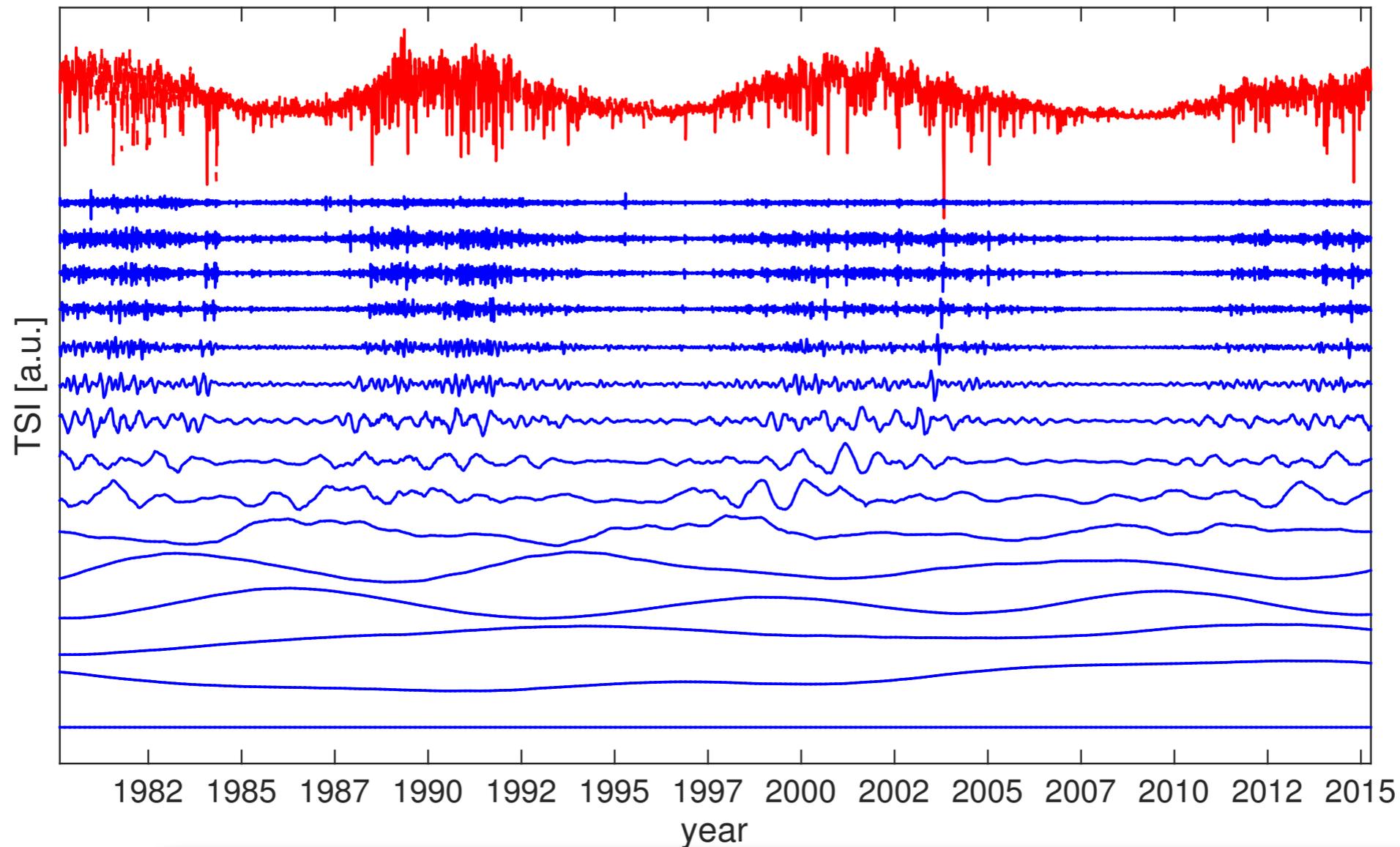


Etape 3

Traiter échelle par échelle

3. Décomposer les données en différentes échelles

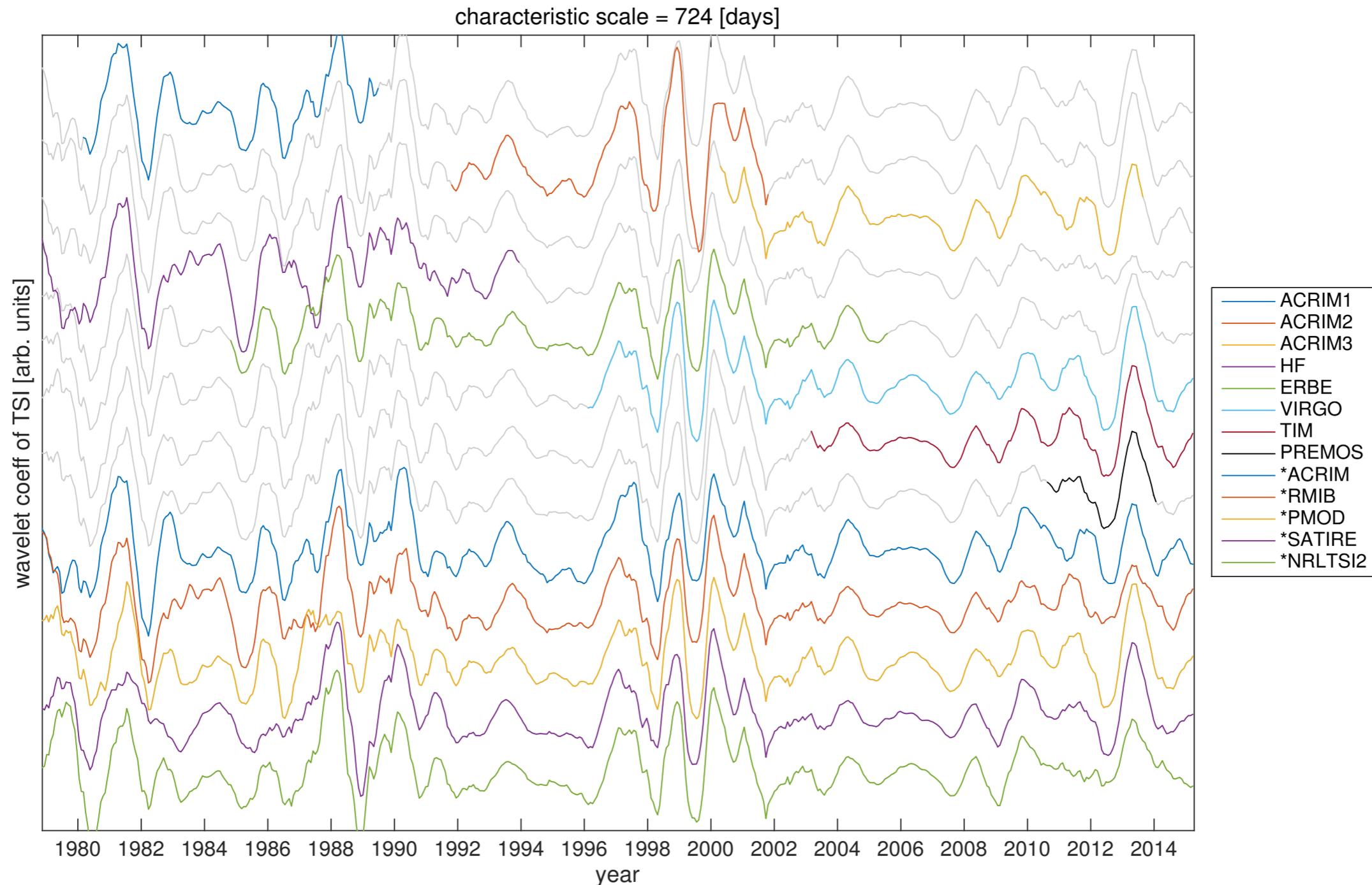
$$TSI(t) = \sum_a TSI(a, t)$$



Orthogonalité = chaque échelle peut être traitée indépendamment des autres

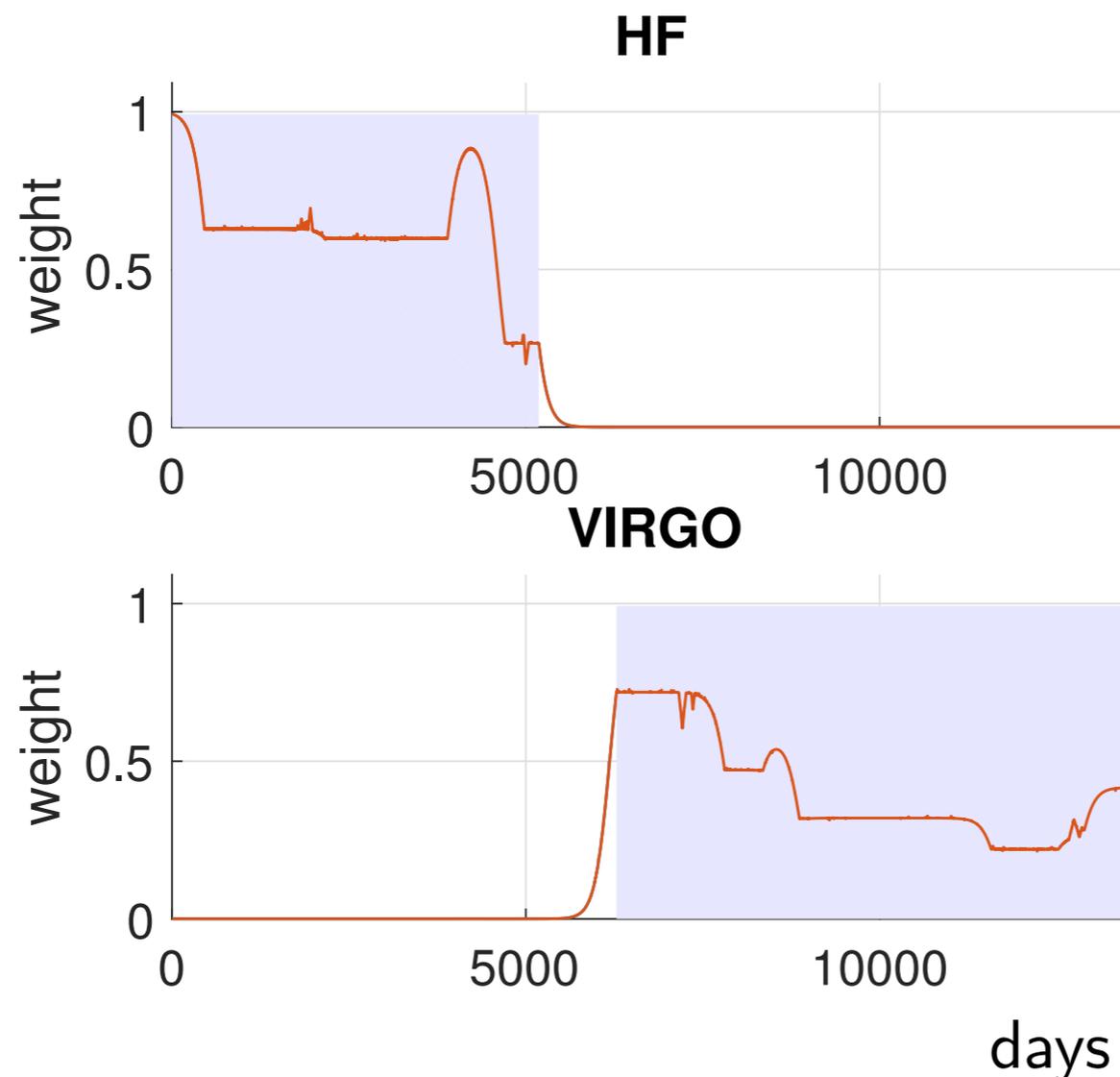
3. Décomposer les données en différentes échelles

■ Pour chaque échelle on effectue une moyenne pondérée des observations



3. Décomposer les données en différentes échelles

- La rémanence de la transformation en ondelettes permet de combler les intervalles de temps avec des valeurs manquantes



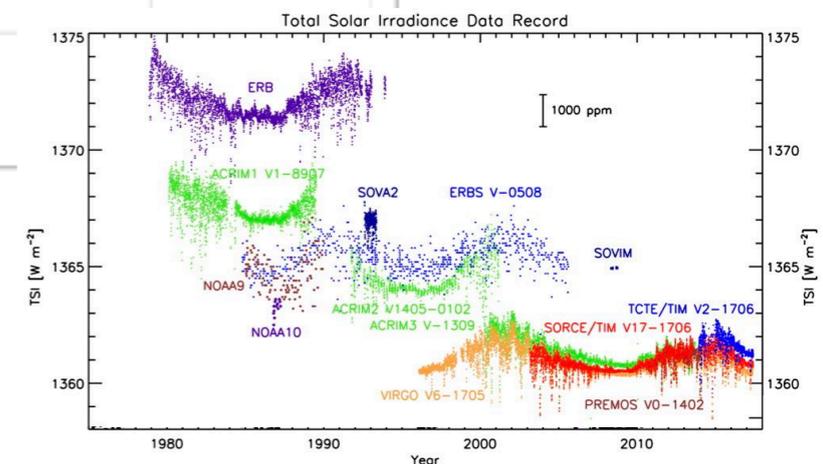
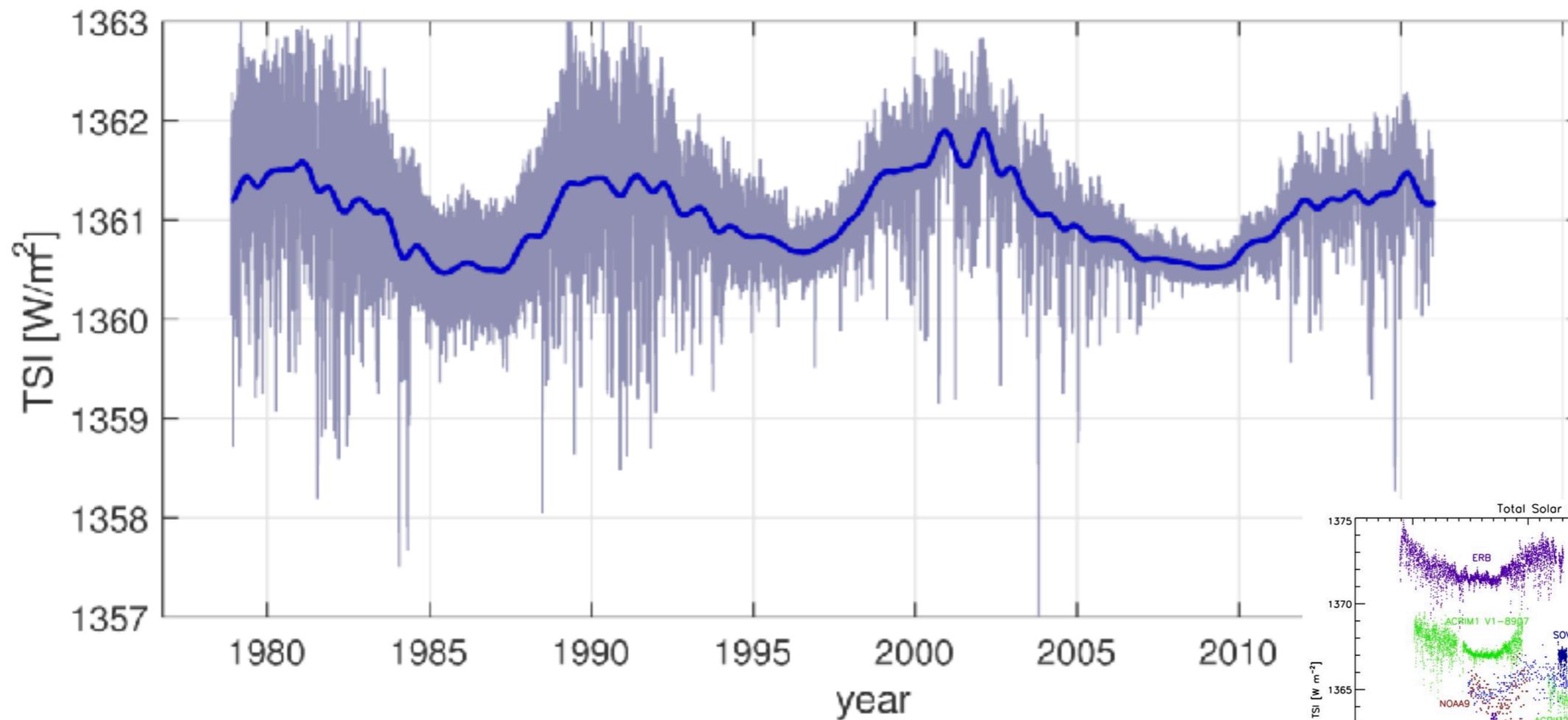
pondération de deux instruments pour une échelle de temps de 416 jours

Etape 4

Estimer les incertitudes

4. Estimer les incertitudes

- L'estimation des incertitudes se fait par Monte-Carlo (incertitudes réalistes : pas d'hypothèse de bruit blanc !)



Pour conclure



Conclusions et perspectives

- Une stratégie “sur mesure” pour un problème fréquent dans la fusion de données environnementales, avec
 - traçabilité
 - obtenir des incertitudes réalistes
- Dans la pratique
 - Matlab: 1 minute pour traiter 40 ans de données
 - portage en Python (open source, open data visé)
- Perspectives
 - intégrer des informations externes (max de vraisemblance → Bayésien)
 - appliquer au nombre de taches solaires (données opportunistes avec biais humain)
 - questions, suggestions ou autres: ddwit@cnrs-orleans.fr