# Reducing stellar noise in exoplanet observables using machine learning

Manuel Perger

Institut de Ciéncies de l'Espai (ICE-CSIC)
Institut d'Estudis Espacials de Catalunya (IEEC)
Barcelona, Spain

Institute of Space Sciences

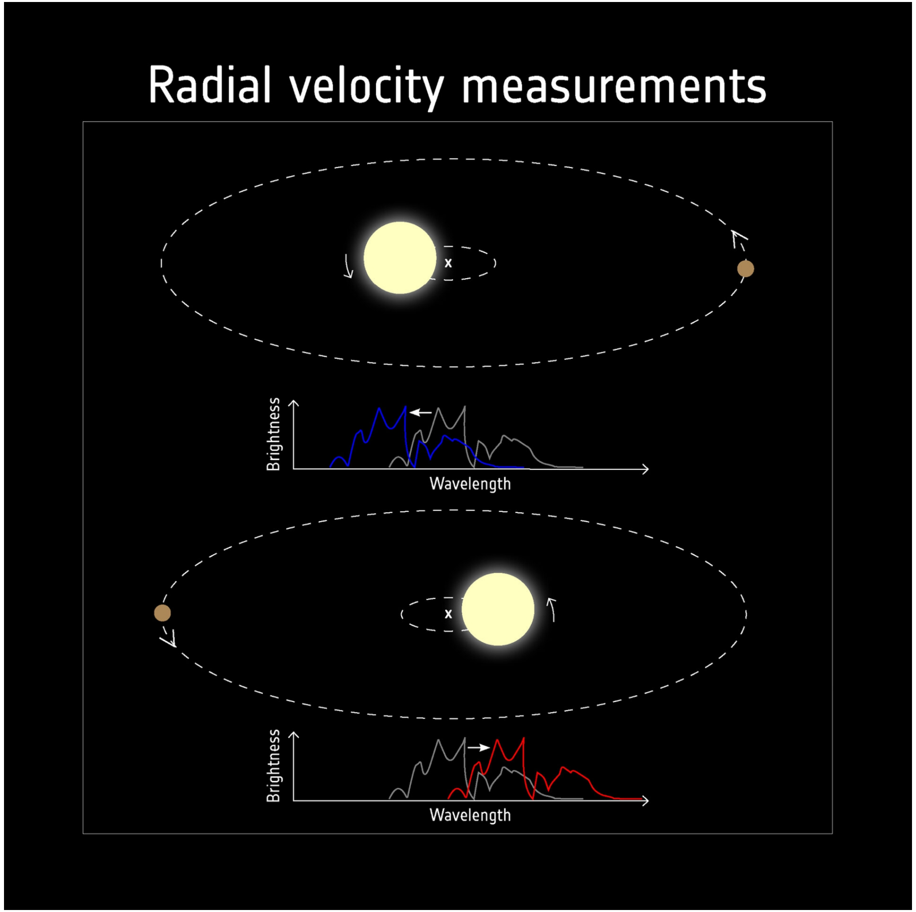
EXCELENCIA MARÍA DE MAEZTU

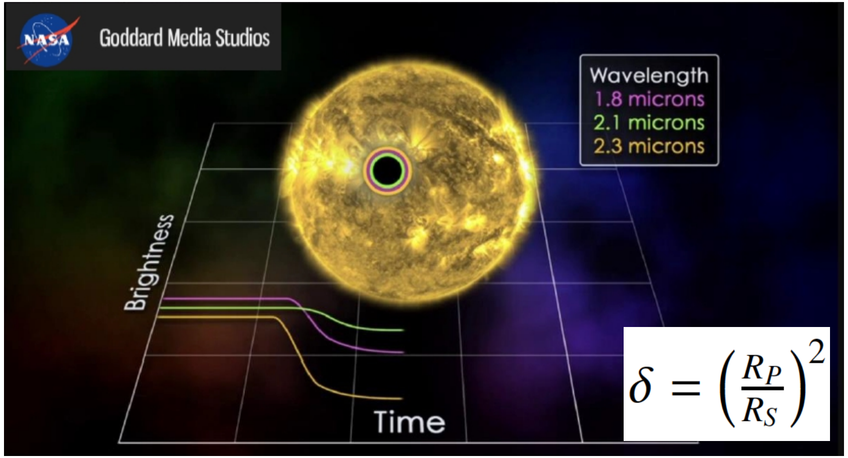
CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

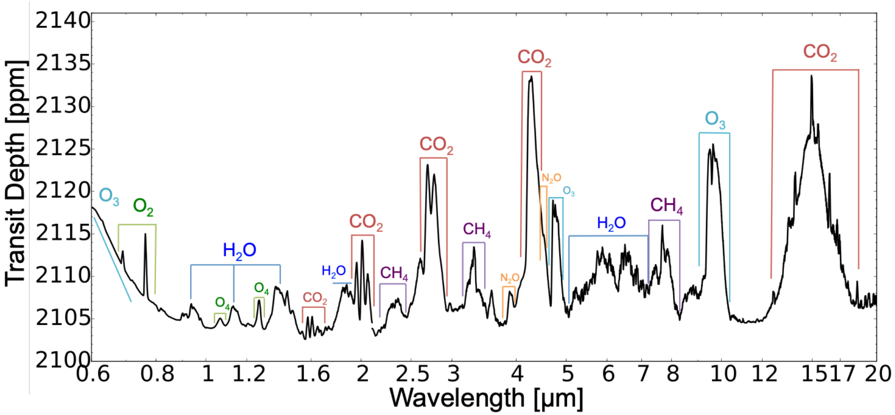
IEEC
Institut d'Estudis Espacials de Catalunya

Institute of
Space Sciences

# Radial velocities, planetary transits, transmission spectra



Institute of
Space Sciences

# Stellar phenomena affecting observations

**oscillations**: min; cm/s

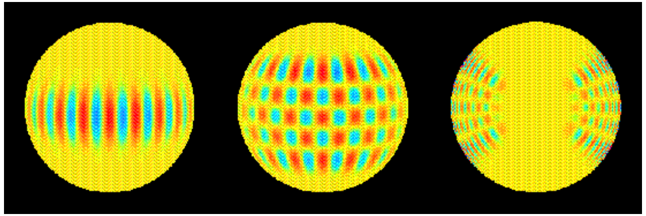
**convection/granulation/flares**
min to hrs; m/s

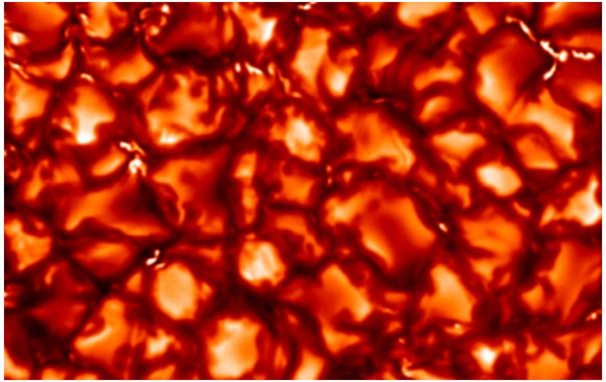
**spots/faculae/plagues**
several rotation periods
m/s to km/s

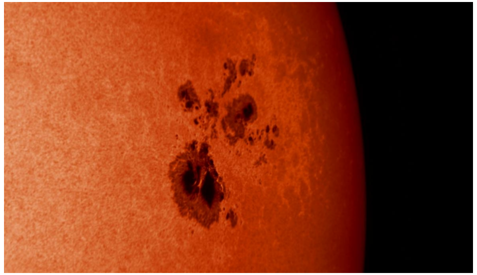
**magnetic cycle**
spot evolution, number and
position, years

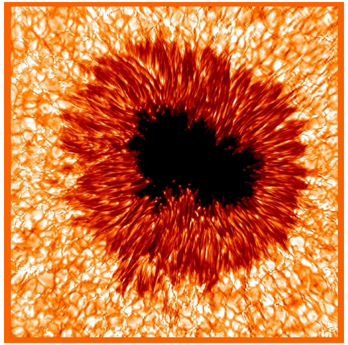
**instrumental precisions ~1 m/s**
**stellar activity >2-3 m/s**
**Earth twin RV 9 cm/s**

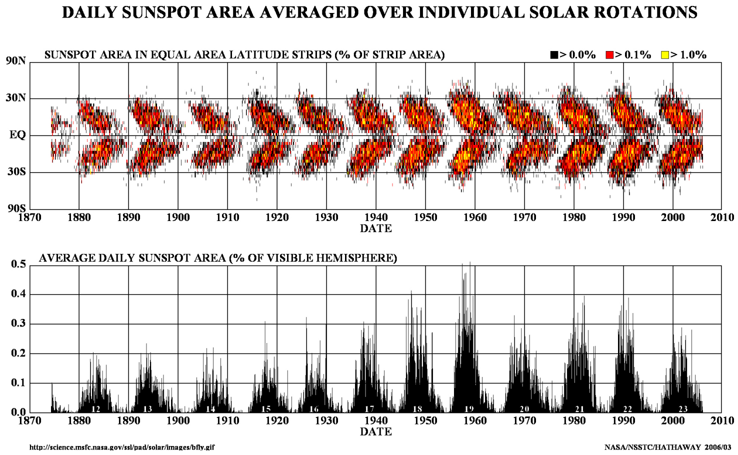
**JWST precision ~20 ppm**
**stellar activity >100 ppm**
**Earth twin ~50 ppm**











Institute of
Space Sciences

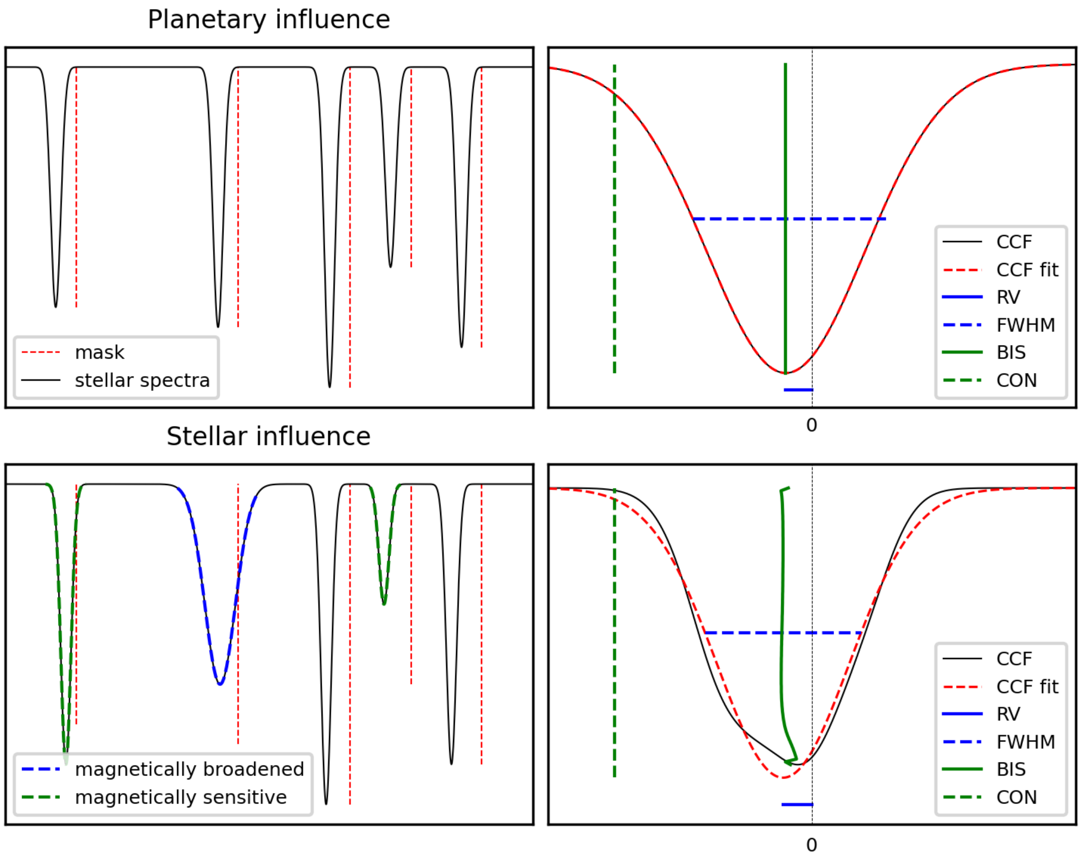
# Magnetic activity affecting stellar spectra

Parameterization with Cross - Correlation Function (**CCF**) and its *moments* (**activity indices**)

Doppler shift induced by a **planetary companion** is achromatic and is an actual translational shift.

Shifts induced by **magnetic effects** (e.g., spots & faculae) depend on wavelength and are only due to asymetries and shape changes of the lines/CCF.

line features can be (more or less) **sensitive to magnetism** – Zeeman effect, line broadening, varying absorption level, etc.



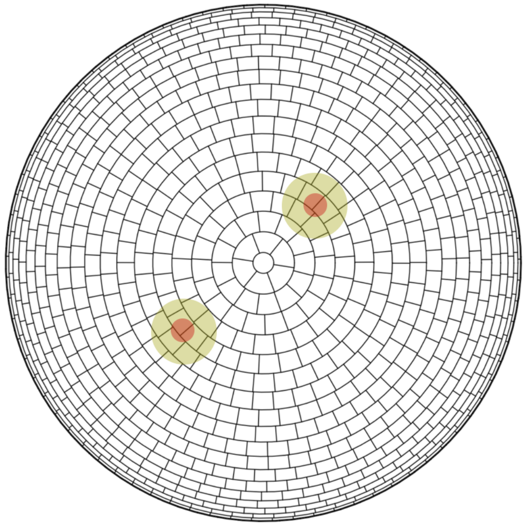
# **StarSim** forward and inverse modelling

## Synthetic observables of a spotted rotating star

Forward modellling of radial velocities and light curves

### Modes

RV: CCF of high resolution spectra
Photometry: low-resolution spectra
Phoenix spetra models (Husser et al. 2013)
500 to 50000 nm



### Stellar parameters

Teff = Tphot = 2600 – 12000 K
logg 3.5 – 5.0
Metallicity -4.0 – 1.0
Radius
Mass
Rotation period
Inclination
Differential rotation
Limb darkening laws
Convective shift & Center-to-limb
Bisector (CIFIST models,
Ludwig et al. 2009)

### Evolving spot map

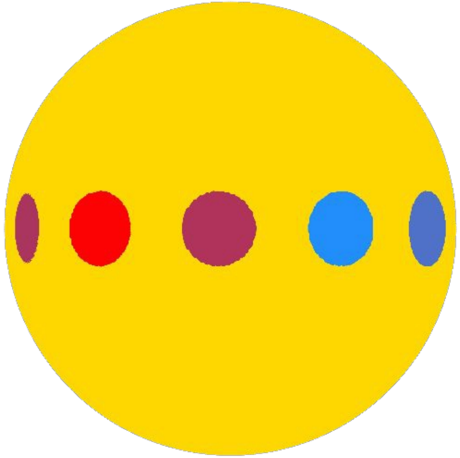
date of appearance
lifetime of spot
latitude
longitude
Radius
Evolution law
N spots
Spot ΔT = Tphot - Tspot
Facula ΔT = Tfac - Tphot

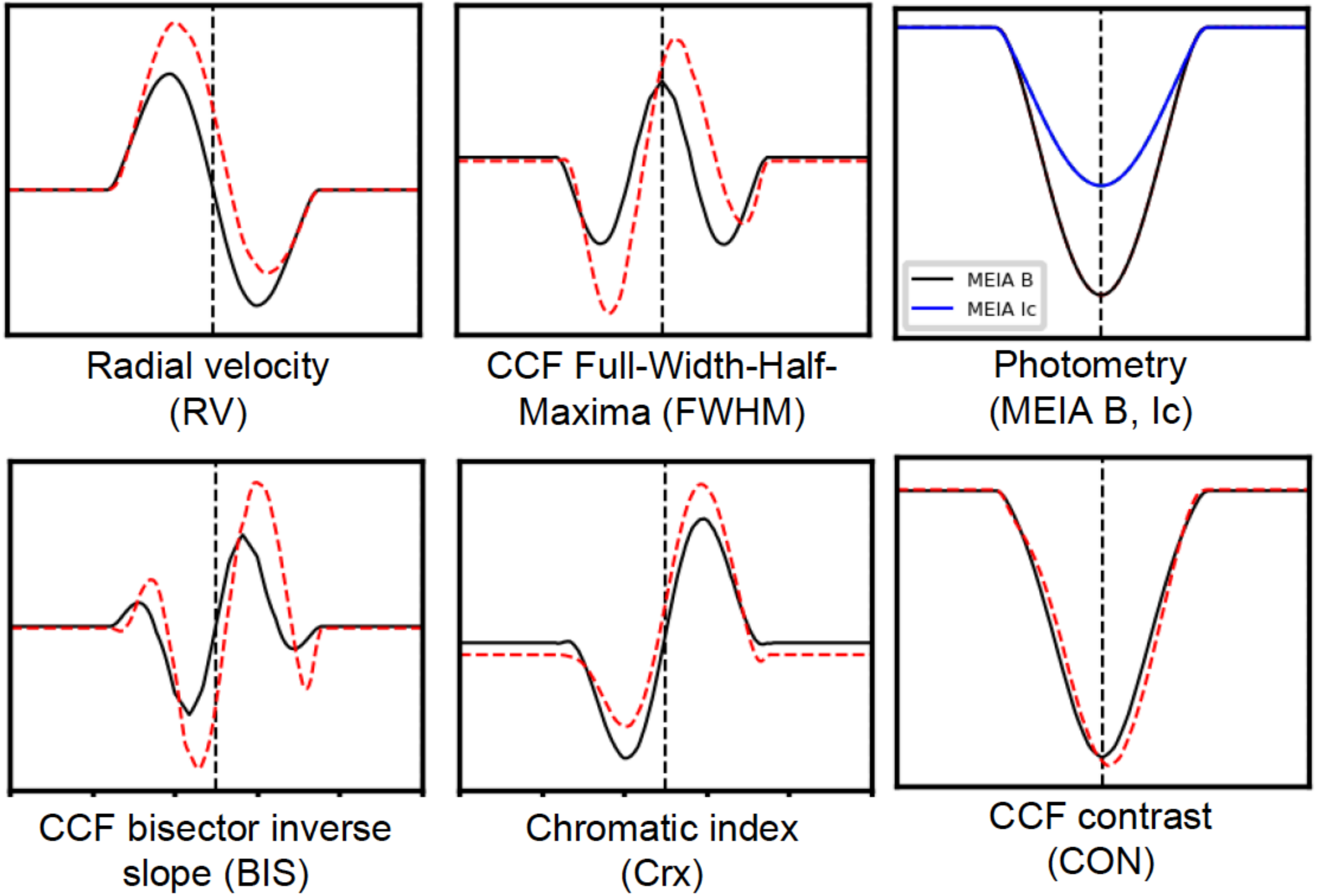
**Herrero et al. (2016)**
**Rosich et al. (2020)**
**ICE team (in prep.)**

**https://github.com/dbarochlopez/starsim**

Institute of
Space Sciences

# **StarSim** time-series data products






## **Other modelling code**

SOAP (Boisse et al. 2012)
SOAP-T (Oshagh et al. 2013)
SOAP2.0 (Dumusque et al. 2014)
SOAP-GPU (Zhao & Dumusque 2023)

# Sun-as-a-star time-series data

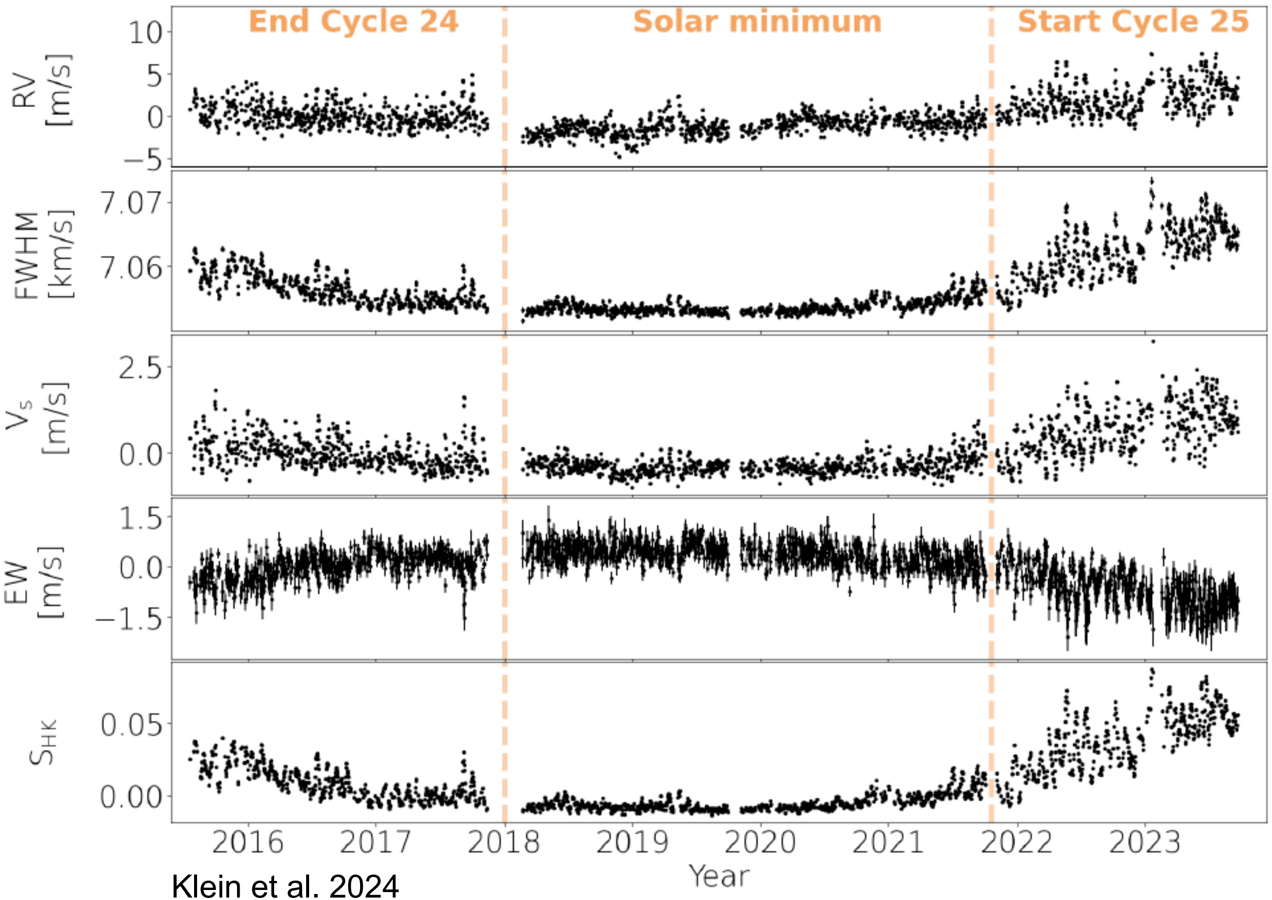
## HARPS-N spectrograph

- **H**igh **A**ccuracy **R**adial velocity **P**lanetary **S**earch project
- stabilised cross-dispersed échelle spectrograph
- Roque de Los Muchachos observatory (La Palma, Spain)
- 3.58-m Telescopio Nazionale Galileo (TNG)
- High resolution R=115 000
- Optical wavelength: 383 to 690 nm

## Solar observations

- Since 2015
- 5-min cadence (averaging out oscillations)
- 150 000 disk integrated spectra (8000 nightly-binned)
- Median SNR =350

**Dumusque et al. (2020)**
https://dace.unige.ch/sun/?




Klein et al. 2024

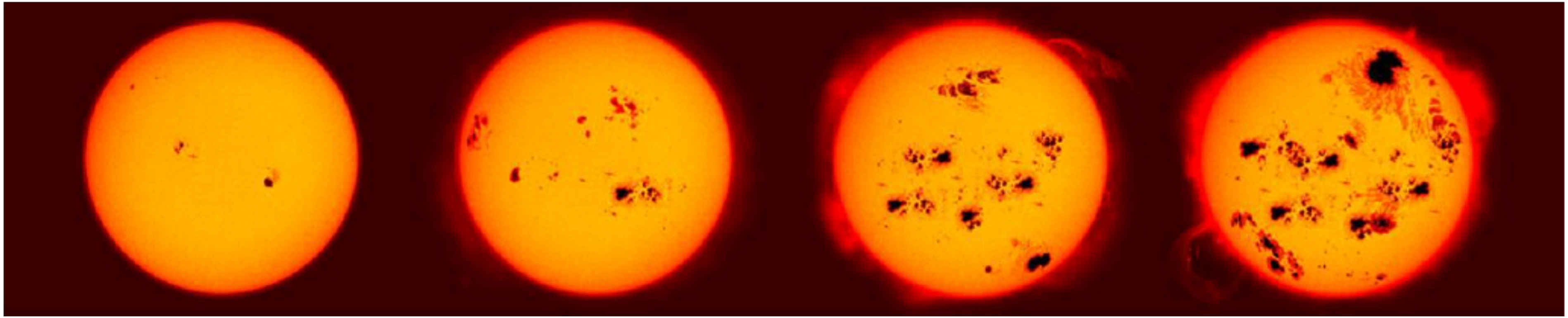
# Problem set

## Problem
- Stellar phenomena have to be mitigated in all exoplanet observations

## Input data
- Test star data (e.g. Sun), model data (StarSim)
- High-resolution spectra, CCF, activity indices time series, temporal correlation
- (*Data contemporaneous to transit event*)

## Output data
- Radial velocity data
- (*Transmission spectrum*)

## Results
- Model data shows good capabilities of the method (down to 2% rms reduction)
- Test star data can be reduced from 10% (AU Mic, active star) to 50% (Sun)
- Better modelling (StarSim3)
- Different (better) spectral parameterization



# Papers on radial velocities

## A machine learning approach for correcting radial velocities using physical observables

M. Perger[1,2], G. Anglada-Escudé[1,2], D. Baroch[1,2], M. Lafarga[3], I. Ribas[1,2], J. C. Morales[1,2], E. Herrero[1,2], P. J. Amado[4], J. R. Barnes[5], J. A. Caballero[6], S. V. Jeffers[7], A. Quirrenbach[8], and A. Reiners[9]

1 Institut de Ciències de l'Espai (ICE, CSIC), Campus UAB, Carrer de Can Magrans s/n, 08193 Bellaterra, Spain
e-mail: perger@ice.cat
2 Institut d'Estudis Espacials de Catalunya (IEEC), c/ Gran Capità 2-4, 08034 Barcelona, Spain
3 Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK
4 Instituto de Astrofísica de Andalucía (IAA-CSIC), Glorieta de la Astronomía s/n, 18008 Granada, Spain
5 School of Physical Sciences, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK
6 Centro de Astrobiología (CAB), CSIC-INTA, ESAC, Camino bajo del castillo s/n, 28692 Villanueva de la Cañada, Madrid, Spain
7 Max Planck Institut für Sonnensystemforschung (MPS), Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany
8
9 Institut für Astrophysik, Georg-August-Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

| | |
|---|---|
| De Beurs et al. (2022) | https://iopscience.iop.org/article/10.3847/1538-3881/ac738e/pdf |
| **Perger et al. (2023)** | https://www.aanda.org/articles/aa/pdf/2023/04/aa45092-22.pdf |
| Liang et al. (2023) | https://iopscience.iop.org/article/10.3847/1538-3881/ad0e01/pdf |
| Colwell et al. (subm.) | https://arxiv.org/pdf/2304.04807 |
| Zhao et al. (subm.) | https://arxiv.org/pdf/2405.13247 |

# StarSim input data modelling

## Different spot maps

600000 EpsEri
1000000 AUMic

| | |
|---|---|
| Number of spots | 25 to 40 |
| Appearance | -lifetime to 100 |
| Lifetime | 5 to 100 days |
| Colatitude | 0 to 180 deg |
| Longitude | 0 to 360 deg |
| Radius | 2.5 to 4 deg |

## Different time samplings

100 time stamps
0<t<100 days
all sets simultaneous
1 x uniformely distributed
2 x randomly distributed
2 x seasonal gap
observed sampling

## Input time-series data

photometric light curves in *VBI* filters
CCF activity indicators: FWHM, BIS and CON
uncertainties extracted from observed data
- AU Mic 30 to 60%
- Eps Eri 17 to 25%

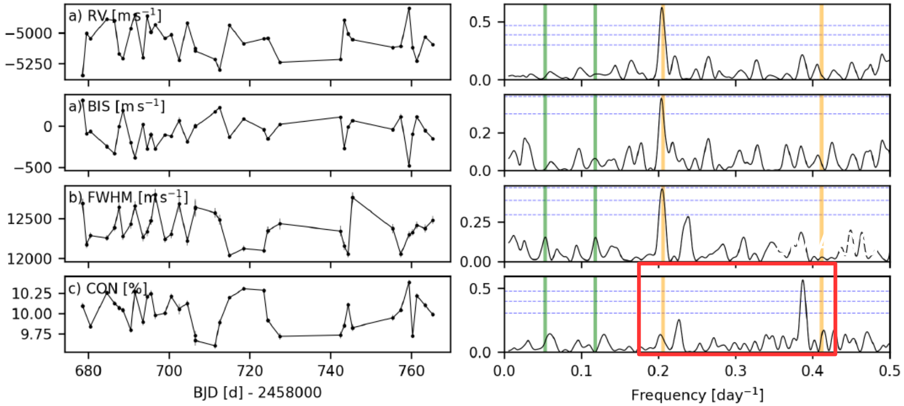
## Output data StarSim models (labels)

radial velocities
90 % training data
10 % test data
observed data is test data

# Test stars and neural network

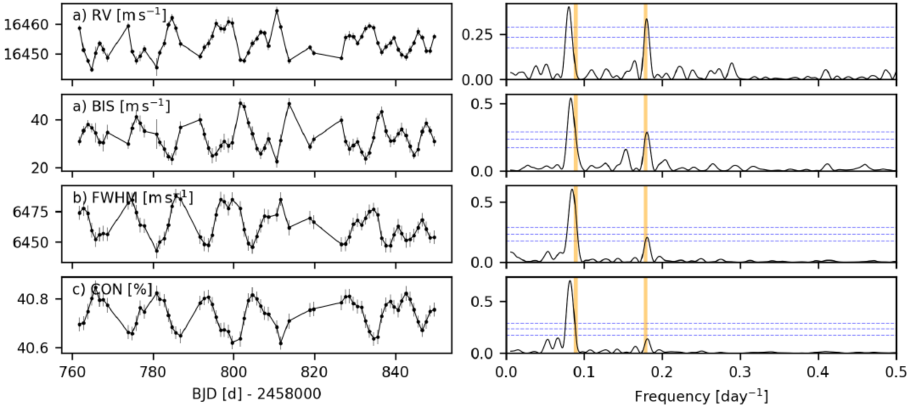
## AU Microscopii

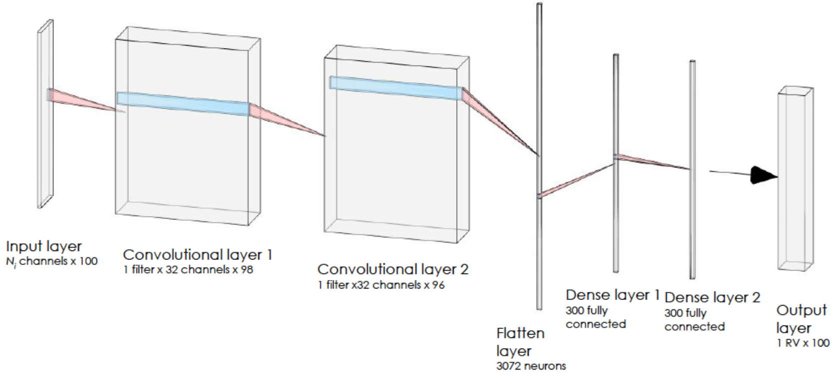
Active M1-type star with 2 planets




## Epsilon Eridani

Young K2-type star with one long-period planet

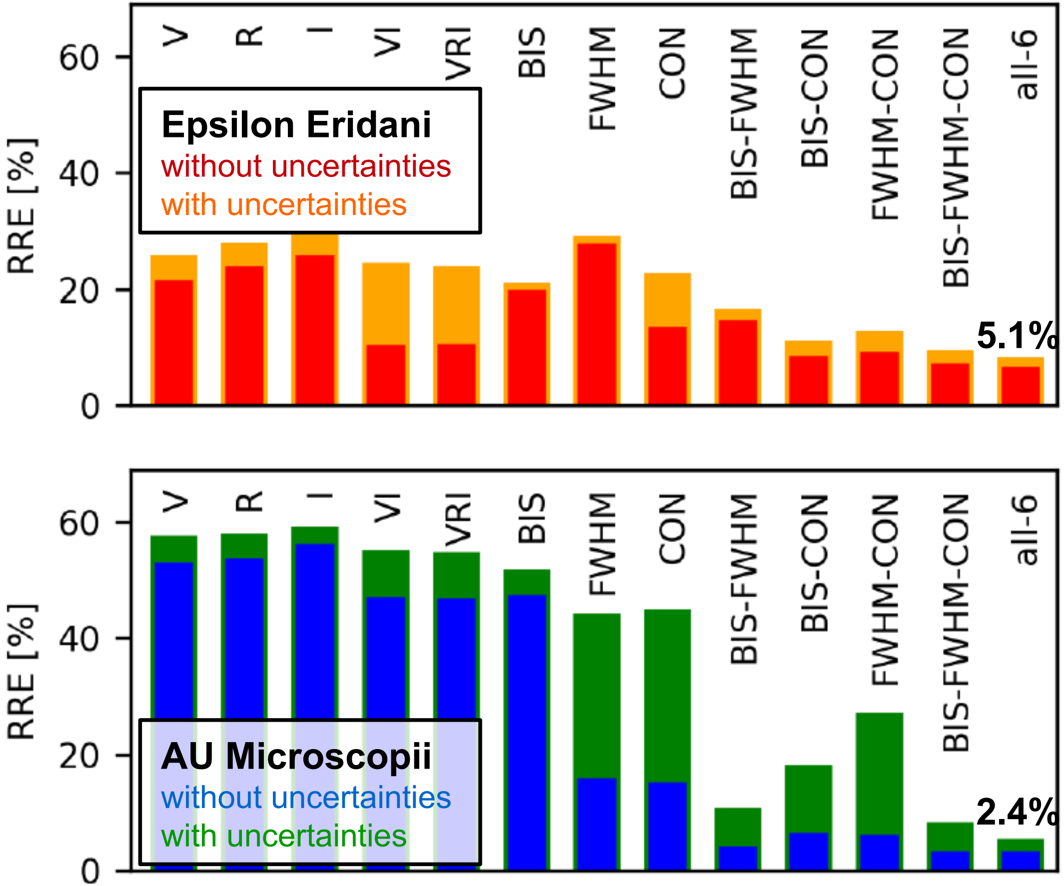


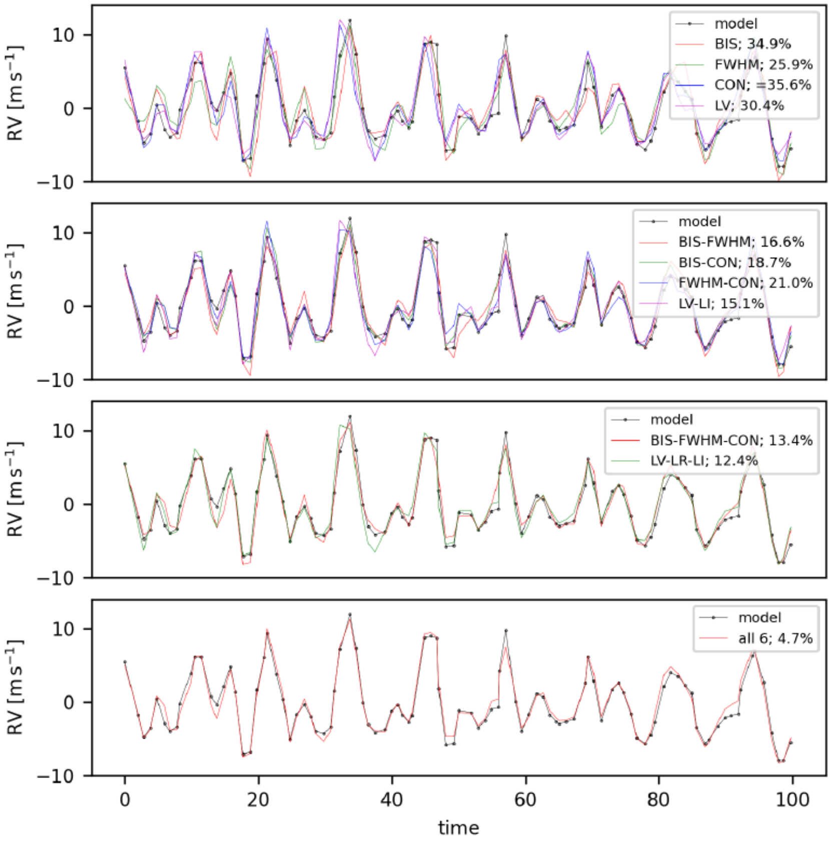




- $N_i$ channels: number of input activity indices
- kernelsize 3
- default 50 epochs

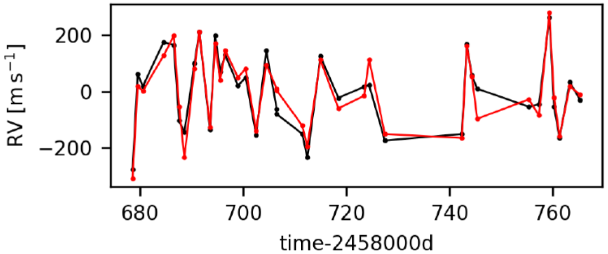
Institute of
Space Sciences

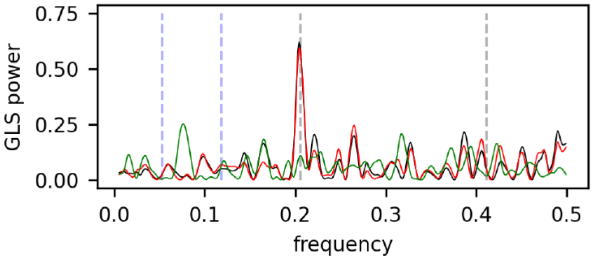
# Results on StarSim models

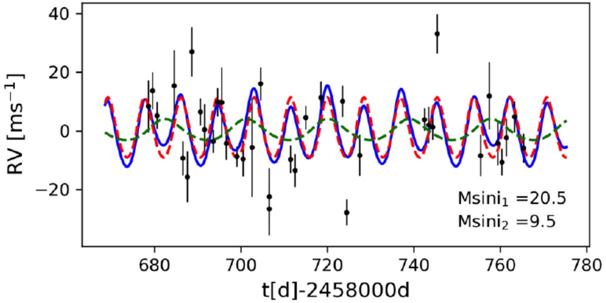


# Results on test stars

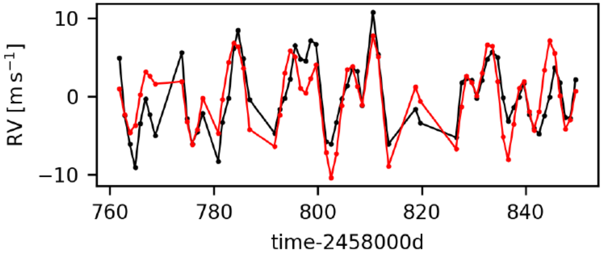


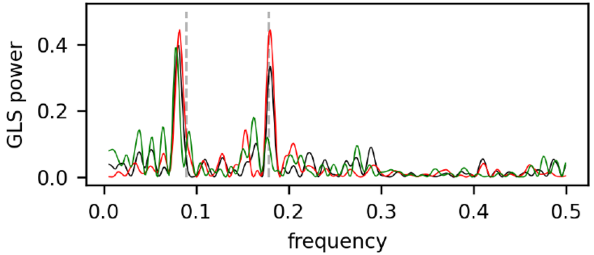




## AU Microscopii

- observed time stamps
- 400000 simulation
- Input: FWHM, and BIS
- Output: RV
- Reduced rms from 132.2 to 13.0 ms-1 (10%)




## Epsilon Eridani

- observed time stamps
- 100000 simulation
- Input: FWHM, BIS, and CON
- Output: RV
- Reduced rms from 4.4 to 2.0 m/s (45%)

# Study on different input data

**CCF decomposition**:

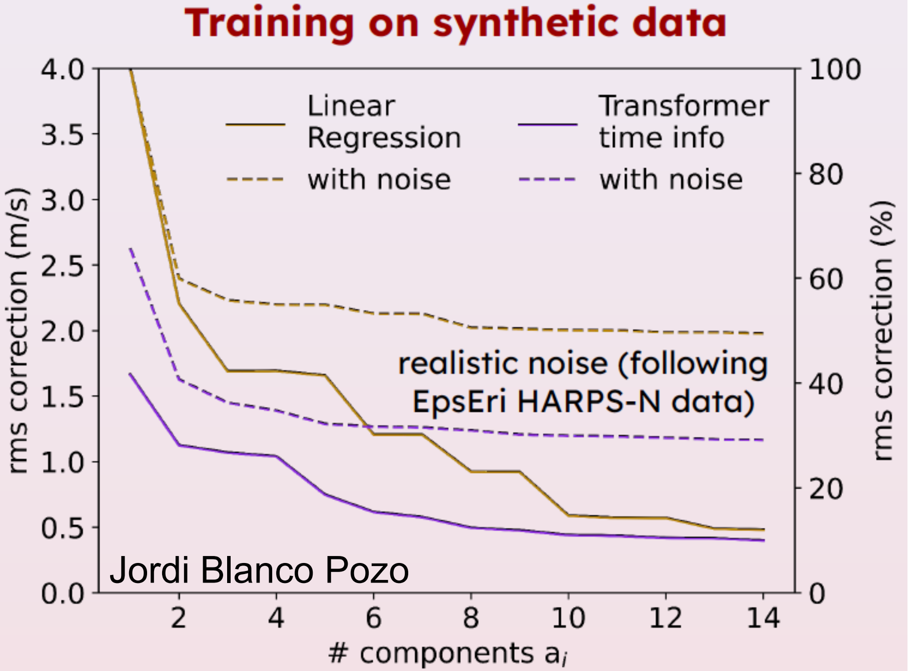
- Auto correlation function (ACF)
  shift invariant
- Moments
  similar to FWHM, BIS, contrast
- CCF principal component analysis (PCA)
  Eigenvectors
- Autoencoder latent space analysis
- Orthonormal functions
  base function G, orthonormalized to CCF'
  G from PCA or Gaussian fit
  $a_0$: output, stellar activity induced RV
  $a_i$: input, show line distortions perpendicular to CCF'

**Transformer models**
include time problem through positional embedding

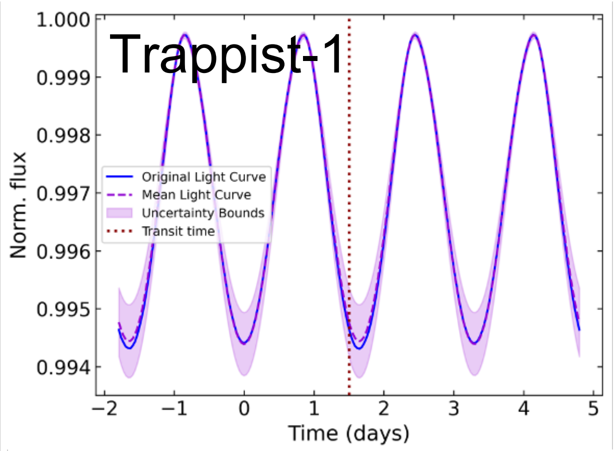
**Test stars**: AU Microscopii, Epsilon Eridani, Sun, etc.

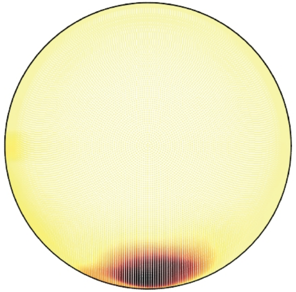
$$\mathbf{F}(v,t) = \langle \mathbf{F}(v) \rangle + \underbrace{(\epsilon(t) + a_0(t))}_{RV_{obs}} \langle \mathbf{F}'(v) \rangle + \sum_{i=1}^{N} a_i(t)\mathbf{G_i}(v)$$

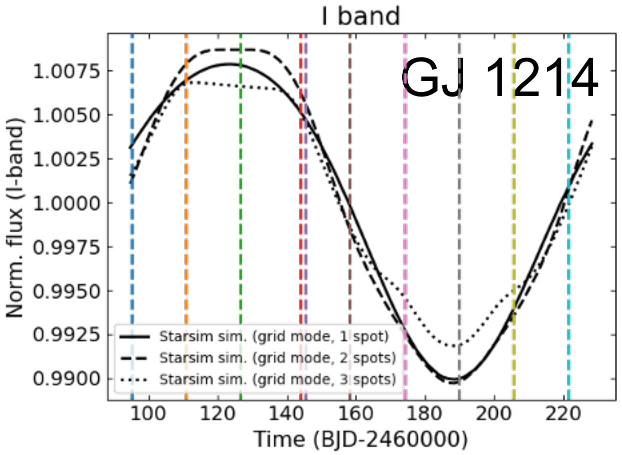


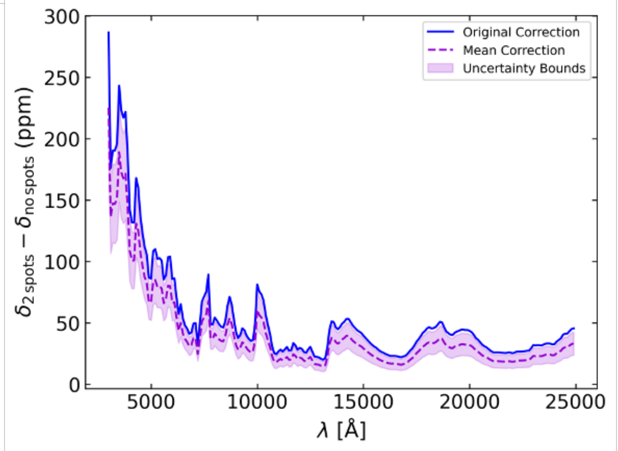


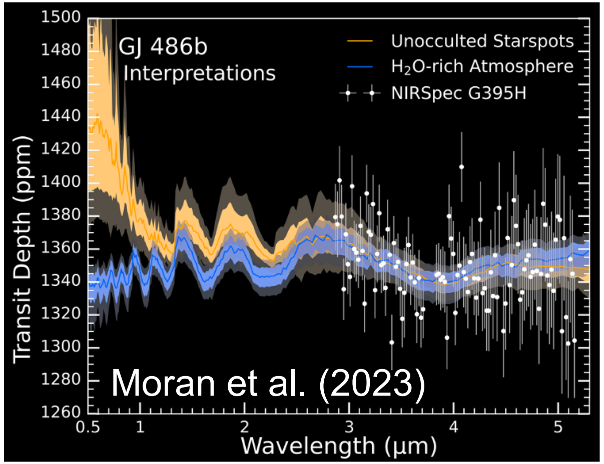
# First results on transmission spectra

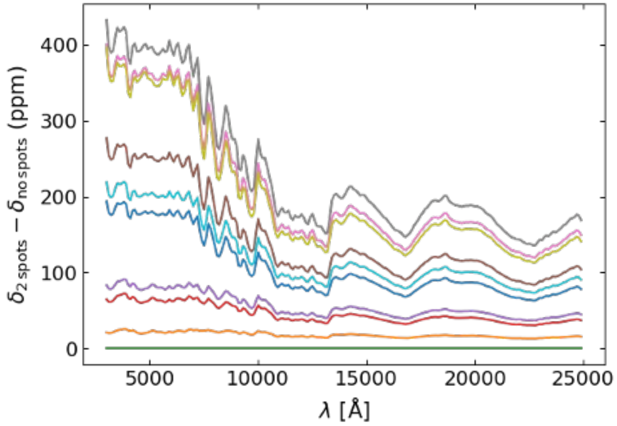




# ERC (European Research Council) Advanced Grant
# Ignasi Ribas

- Detecting ExoEarths
- Starsim3 development
- Stellar data modelling
- General Machine Learning algorithms
- RV variations down to 10 cm/s
- Transit spectroscopy down to 10 ppm

https://www.ieec.cat/en/ieec/job-offers/

https://ice.csic.es/about-us/jobs

Institute of Space Sciences
EXCELENCIA MARÍA DE MAEZTU
CSIC CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS
IEEC Institut d'Estudis Espacials de Catalunya
erc European Research Council Established by the European Commission