LOCKHEED MARTIN

Constraining the Frequency of Energy Deposition through Quantitative Comparisons of Models and Observations

9th Coronal Loops Workshop — University of St Andrews
12 June 2019

Will Barnes[1,2]

[1] Lockheed Martin Solar and Astrophysics Laboratory, [2] Bay Area Environmental Research Institute

Stephen Bradshaw[3], Nicholeen Viall[4]

[3] Department of Physics and Astronomy, Rice University
[4] NASA Goddard Space Flight Center

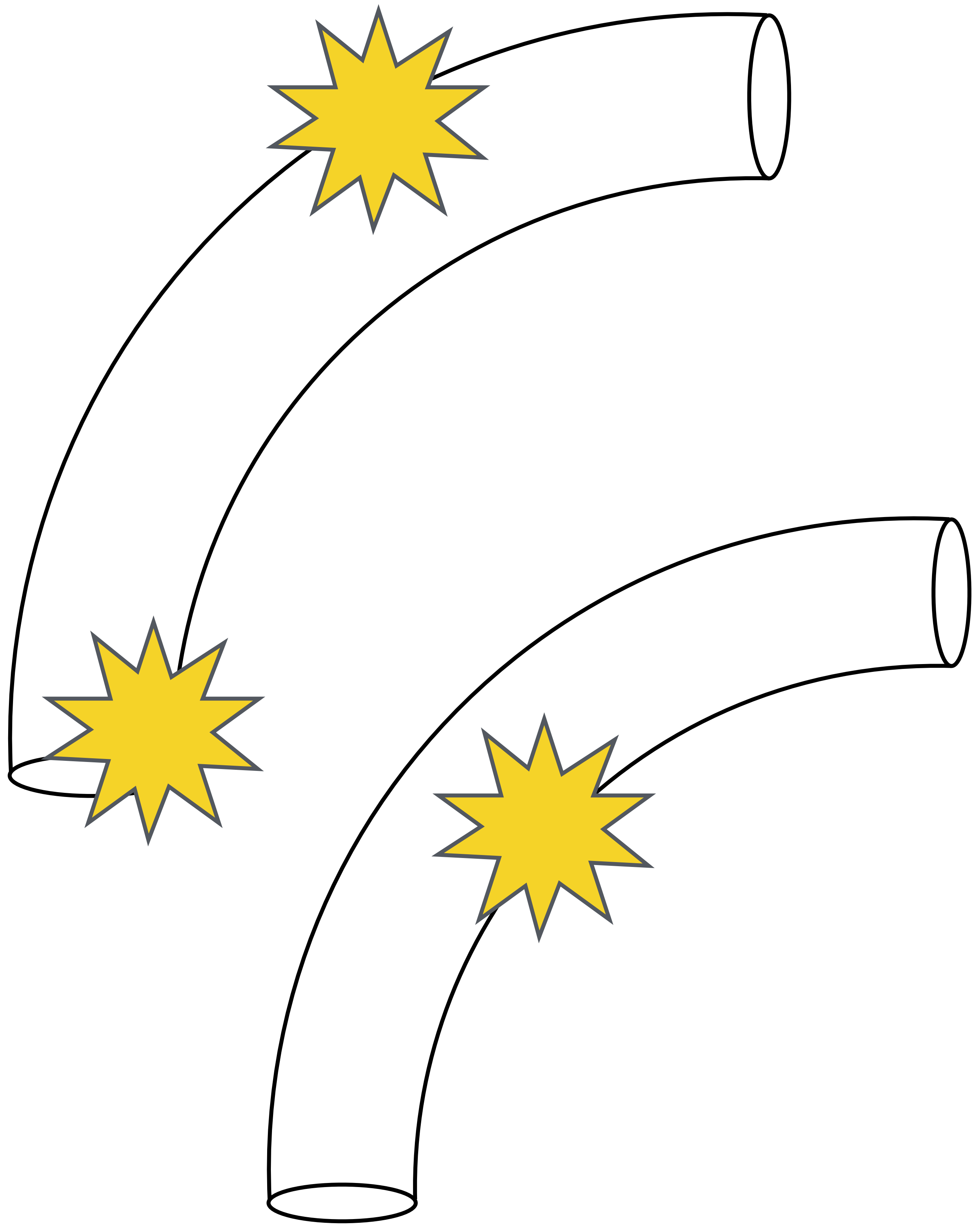
SOLAR AND ASTROPHYSICS LAB

sunpy powered

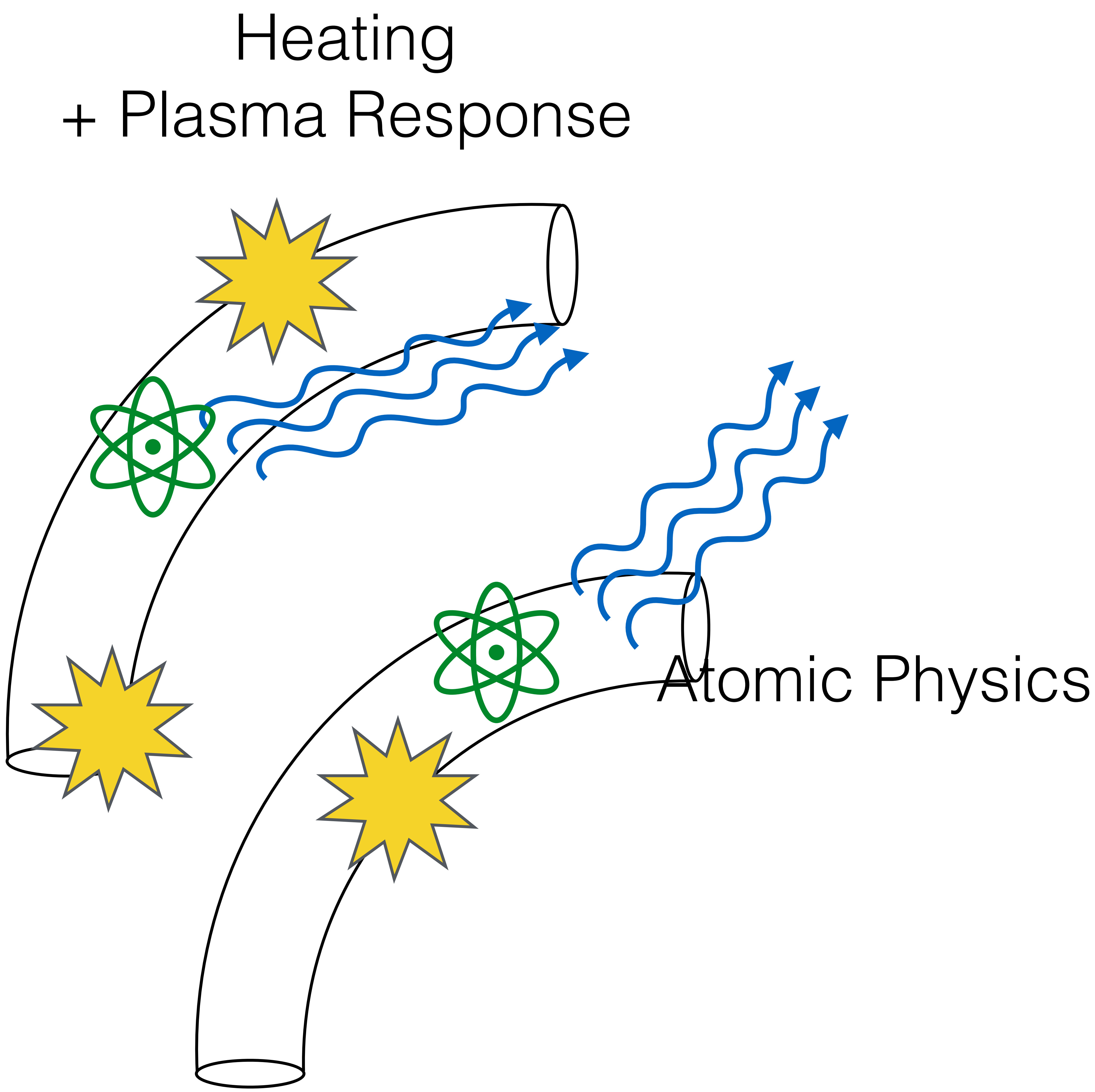
The Problem

The Problem

Heating
+ Plasma Response



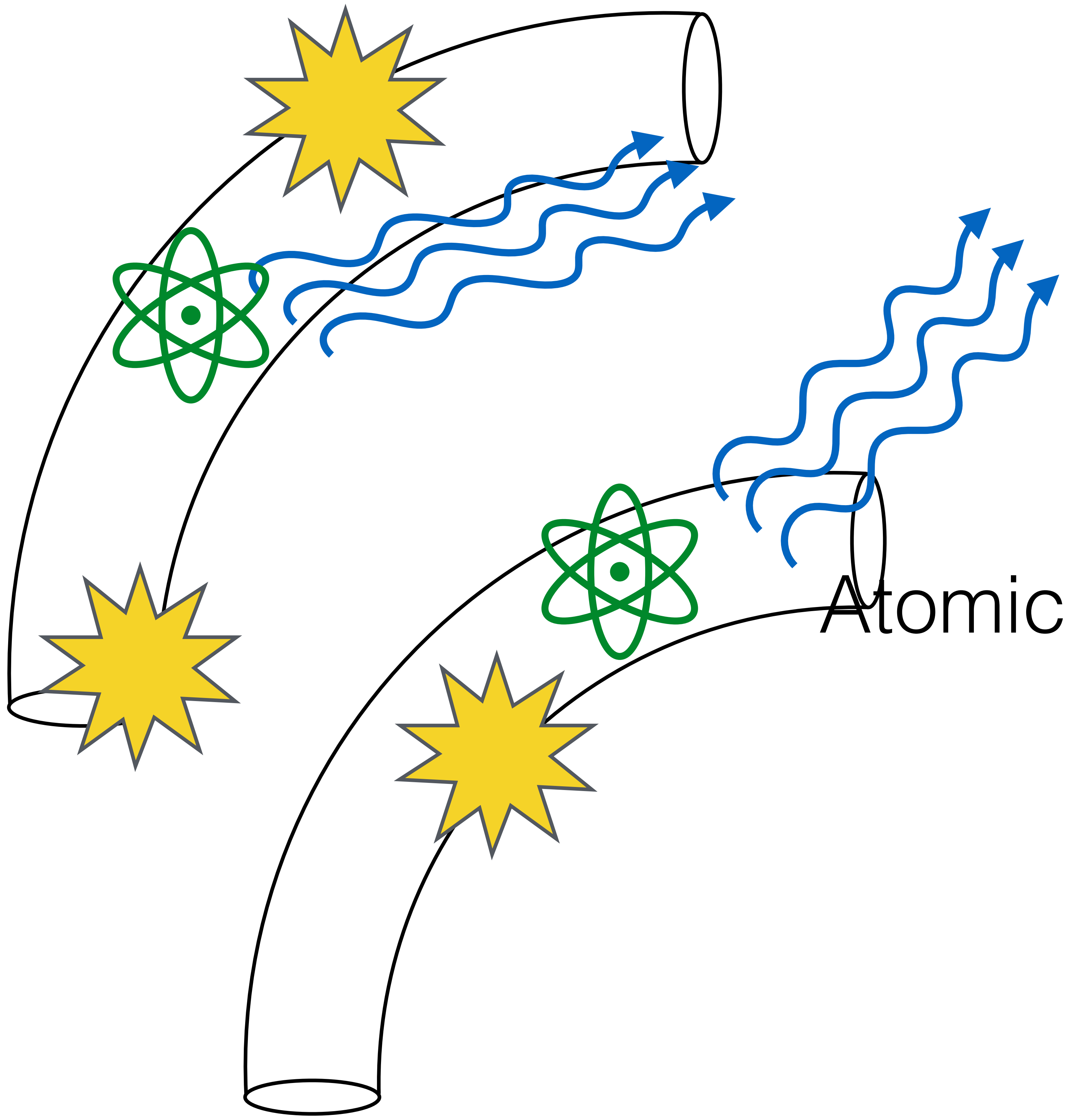
The Problem



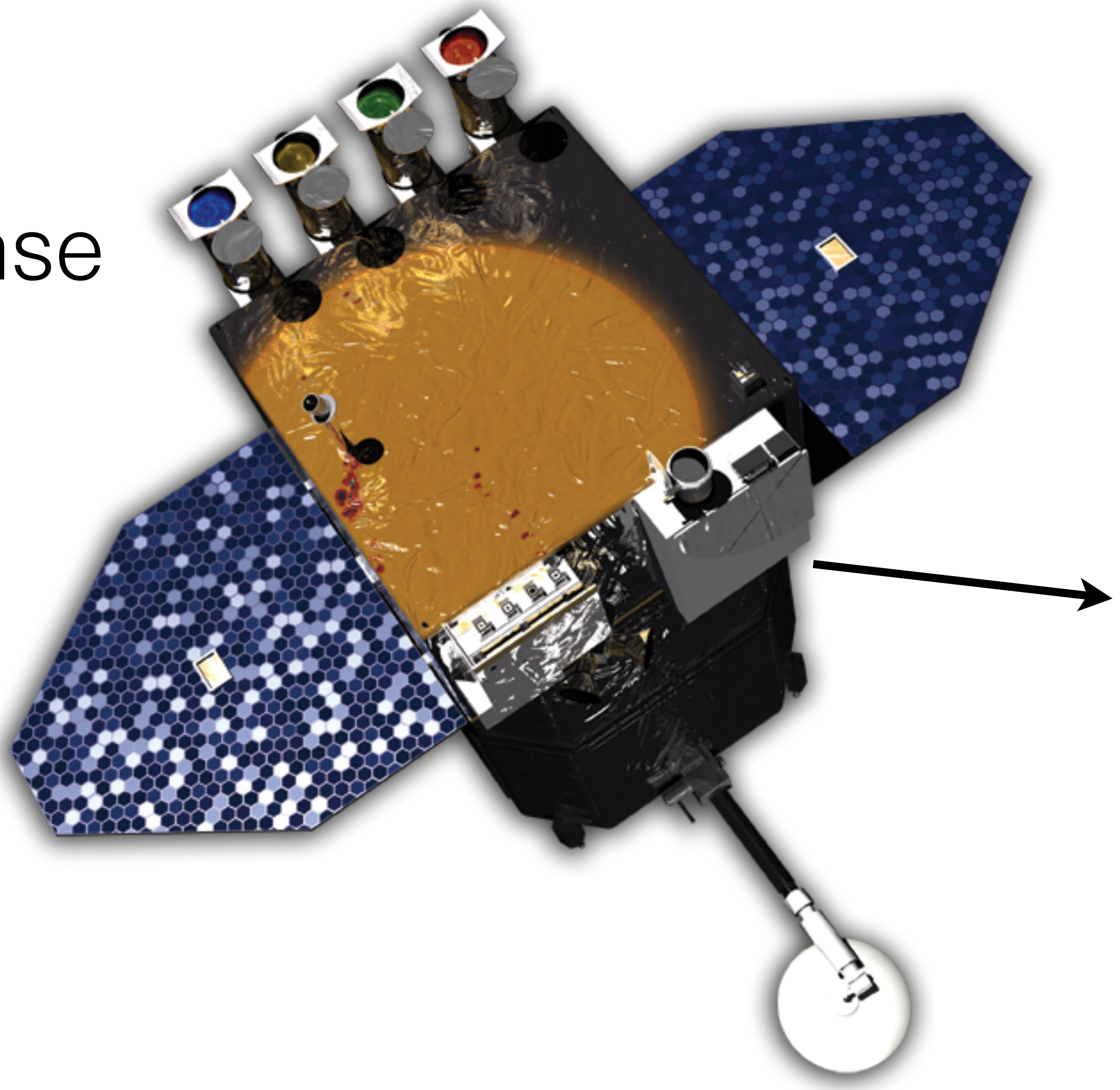
The Problem

Instrument Response

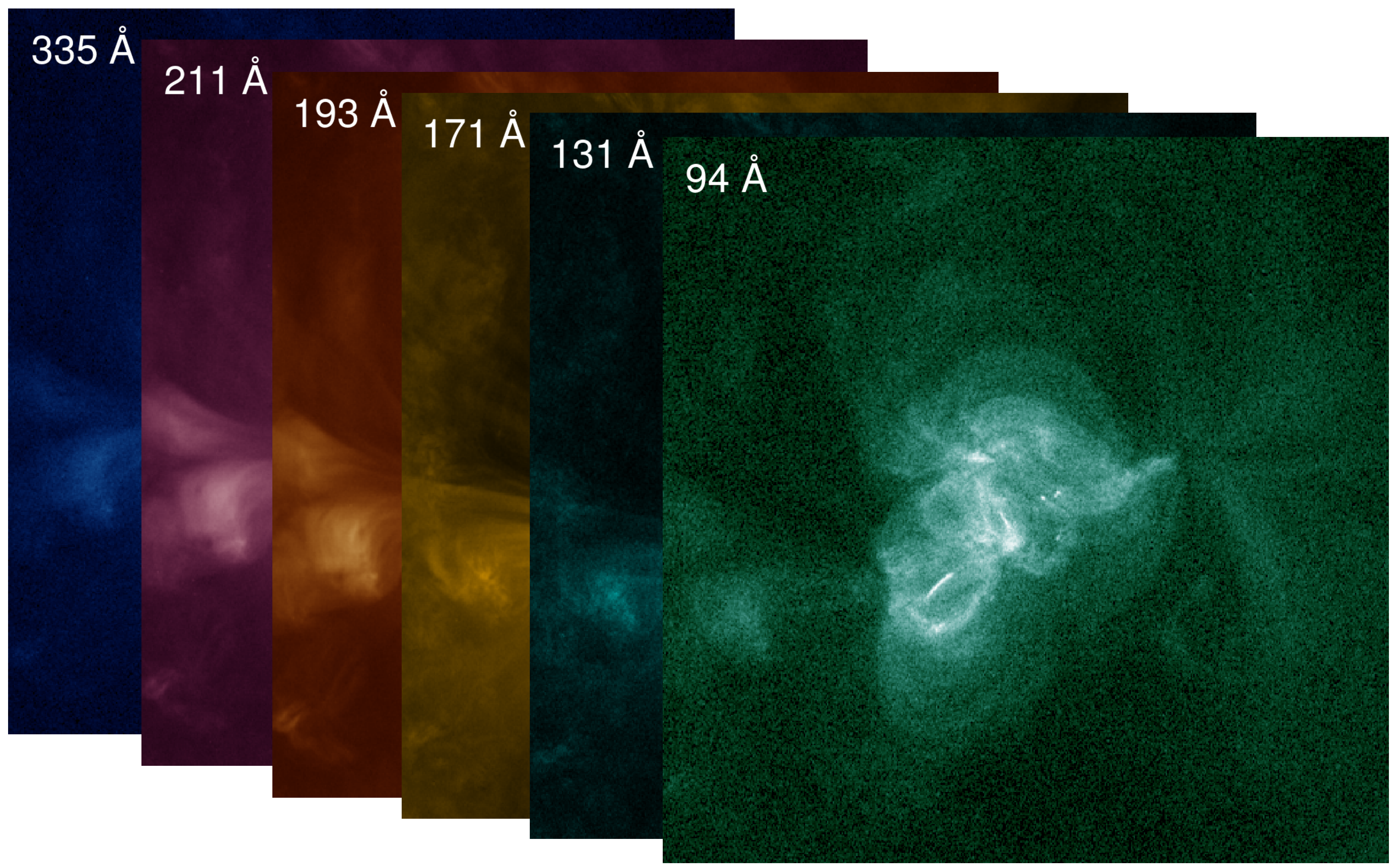
Heating
+ Plasma Response



Atomic Physics



Multi-wavelength Observations



The Problem


Multi-wavelength Observations
335 Å
211 Å
193 Å
171 Å
131 Å
94 Å
Instrument Response
Heating
+ Plasma Response
Atomic Physics
Viall and Klimchuk (2012)
$\alpha = 2.7 \pm 0.32$
$\beta = 7.9 \pm 2.97$
Warren et al. (2012)
Diagnostics

The Problem

Multi-wavelength Observations
335 Å
211 Å
193 Å
171 Å
131 Å
94 Å
Instrument Response
Heating
+ Plasma Response
Atomic Physics
β = 7.9 ± 2.97
Warren et al. (2012)
Viall and Klimchuk (2012)
Diagnostics

The Problem

How do we compare models and observations in order to constrain the frequency of energy deposition in active region loops?

The Problem

How do we compare models and observations in order to constrain the frequency of energy deposition in active region loops?

To answer this question, we need:

The Problem

How do we compare models and observations in order to constrain the frequency of energy deposition in active region loops?

To answer this question, we need:

1. **Model of Coronal Plasma and Resulting Emission**

The Problem

How do we compare models and observations in order to constrain the frequency of energy deposition in active region loops?

To answer this question, we need:

1. **Model of Coronal Plasma and Resulting Emission**

2. **Multiple Diagnostics of the Heating Frequency**

The Problem

How do we compare models and observations in order to constrain the frequency of energy deposition in active region loops?

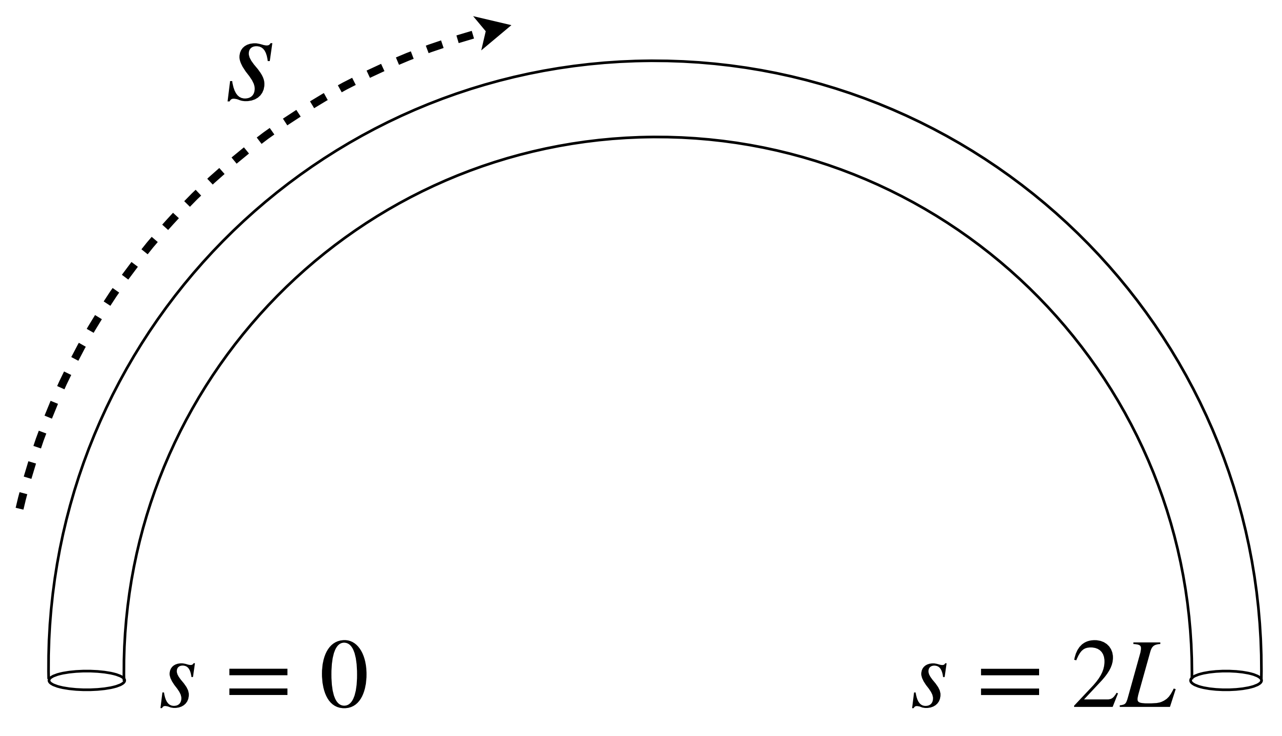
To answer this question, we need:

1. **Model of Coronal Plasma and Resulting Emission**
2. **Multiple Diagnostics of the Heating Frequency**
3. **A Quantitative Measure of the "distance" Between Modeled and Observed Diagnostics**

1. Modeling Plasma Dynamics and Emission

1. Modeling Plasma Dynamics and Emission

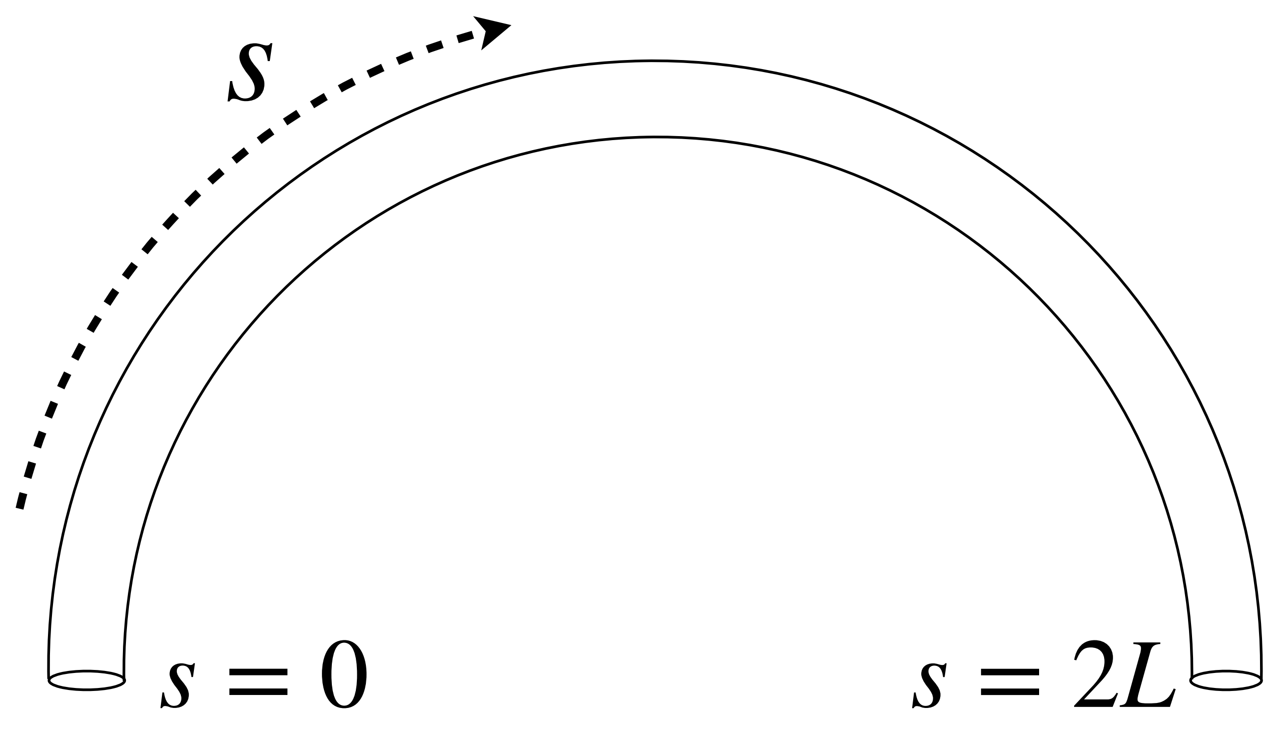
Single Loop Models



- HYDRAD
  (Bradshaw and Cargill, 2013)
- RADYN
  (Carlsson and Stein, 1992;
  Allred et al., 2015)
- Predictive Sciences
  (Mikić et al., 2013)
- EBTEL (0D)
  (Klimchuk et al. 2008;
  Cargill et al. 2012a,b)

1. Modeling Plasma Dynamics and Emission

Single Loop Models



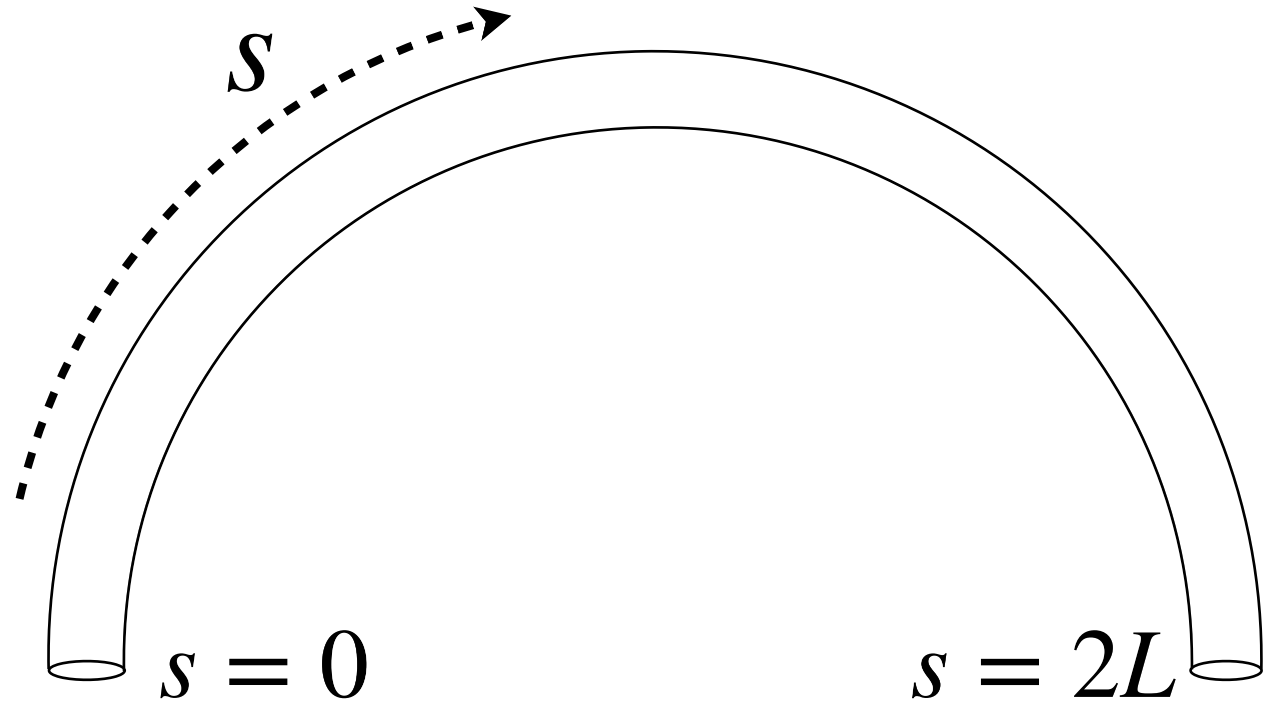
- HYDRAD
  (Bradshaw and Cargill, 2013)
- RADYN
  (Carlsson and Stein, 1992;
  Allred et al., 2015)
- Predictive Sciences
  (Mikić et al., 2013)
- EBTEL (0D)
  (Klimchuk et al. 2008;
  Cargill et al. 2012a,b)

Multi-Loop Models

- Allred et al. (2018)
- Bradshaw and Viall (2016)
- Lundquist et al. (2008a,b)
- Warren and Winebarger (2007)
- Schrijver et al. (2004)

1. Modeling Plasma Dynamics and Emission

Single Loop Models

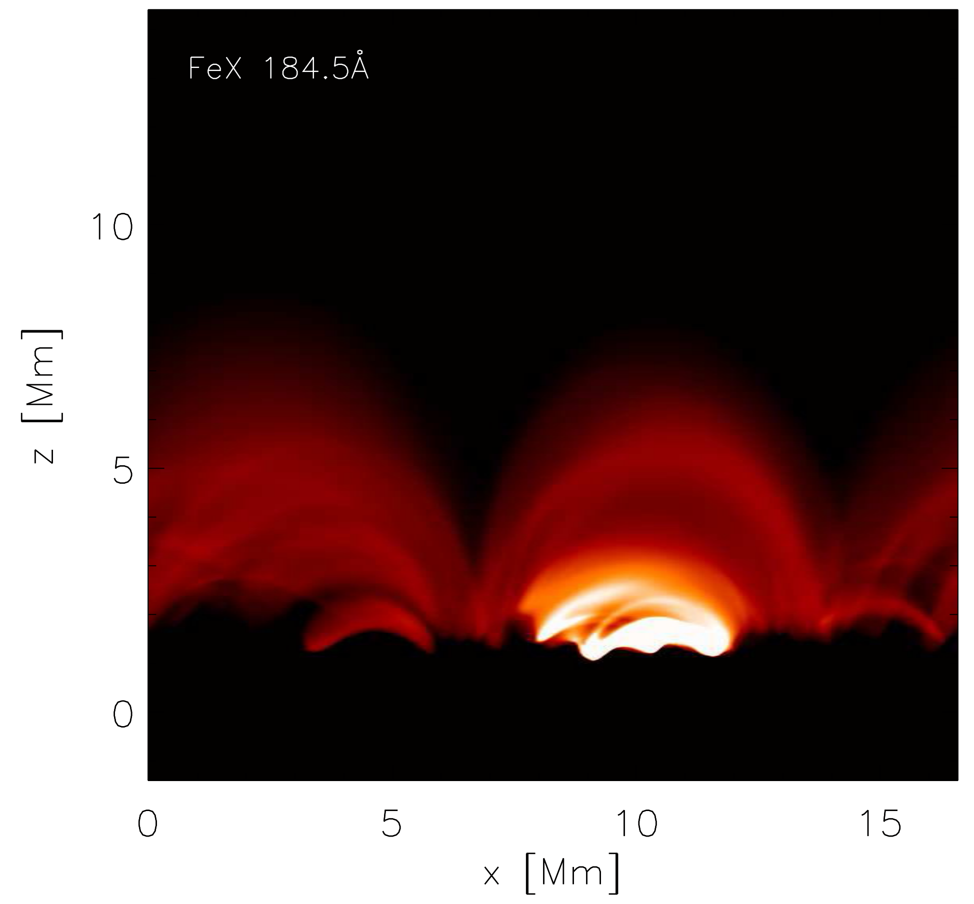


- HYDRAD (Bradshaw and Cargill, 2013)
- RADYN (Carlsson and Stein, 1992; Allred et al., 2015)
- Predictive Sciences (Mikić et al., 2013)
- EBTEL (0D) (Klimchuk et al. 2008; Cargill et al. 2012a,b)

Multi-Loop Models

- Allred et al. (2018)
- Bradshaw and Viall (2016)
- Lundquist et al. (2008a,b)
- Warren and Winebarger (2007)
- Schrijver et al. (2004)

3D MHD

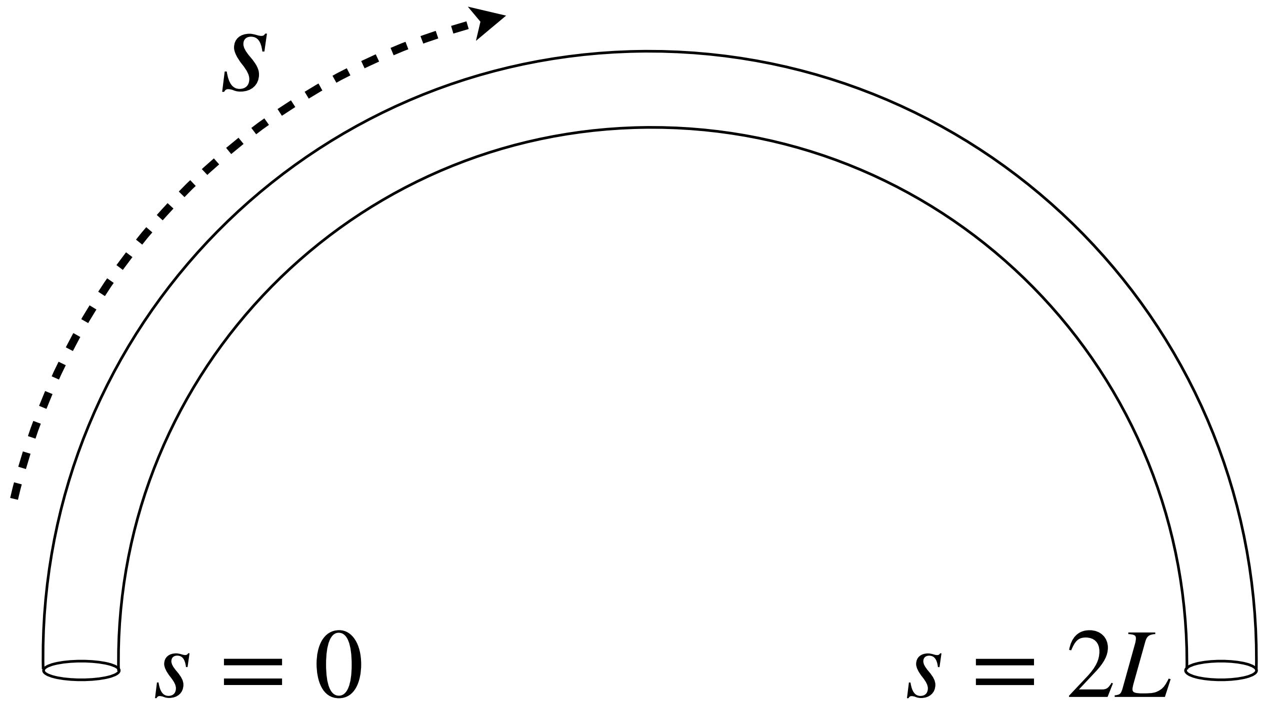


Testa et al. (2012)

- BIFROST (Gudiksen et al., 2011)
- MAS code (e.g. Mikić et al., 2017)
- MURaM (Rempel 2017)

1. Modeling Plasma Dynamics and Emission

Single Loop Models



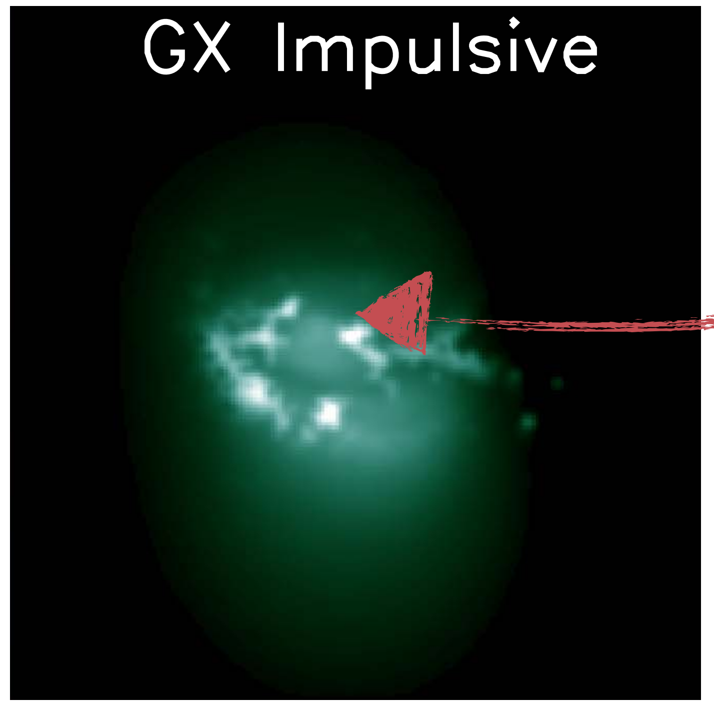
- HYDRAD (Bradshaw and Cargill, 2013)
- RADYN (Carlsson and Stein, 1992; Allred et al., 2015)
- Predictive Sciences (Mikić et al., 2013)
- EBTEL (0D) (Klimchuk et al. 2008; Cargill et al. 2012a,b)

Multi-Loop Models

- Allred et al. (2018)
- Bradshaw and Viall (2016)
- Lundquist et al. (2008a,b)
- Warren and Winebarger (2007)
- Schrijver et al. (2004)

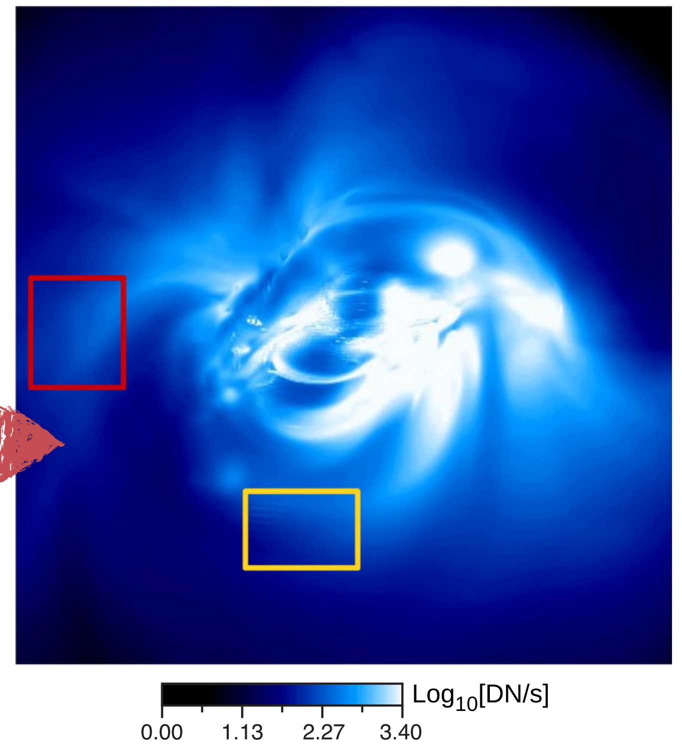
See talk by Schonfeld

GX Simulator



Nita et al. (2017)

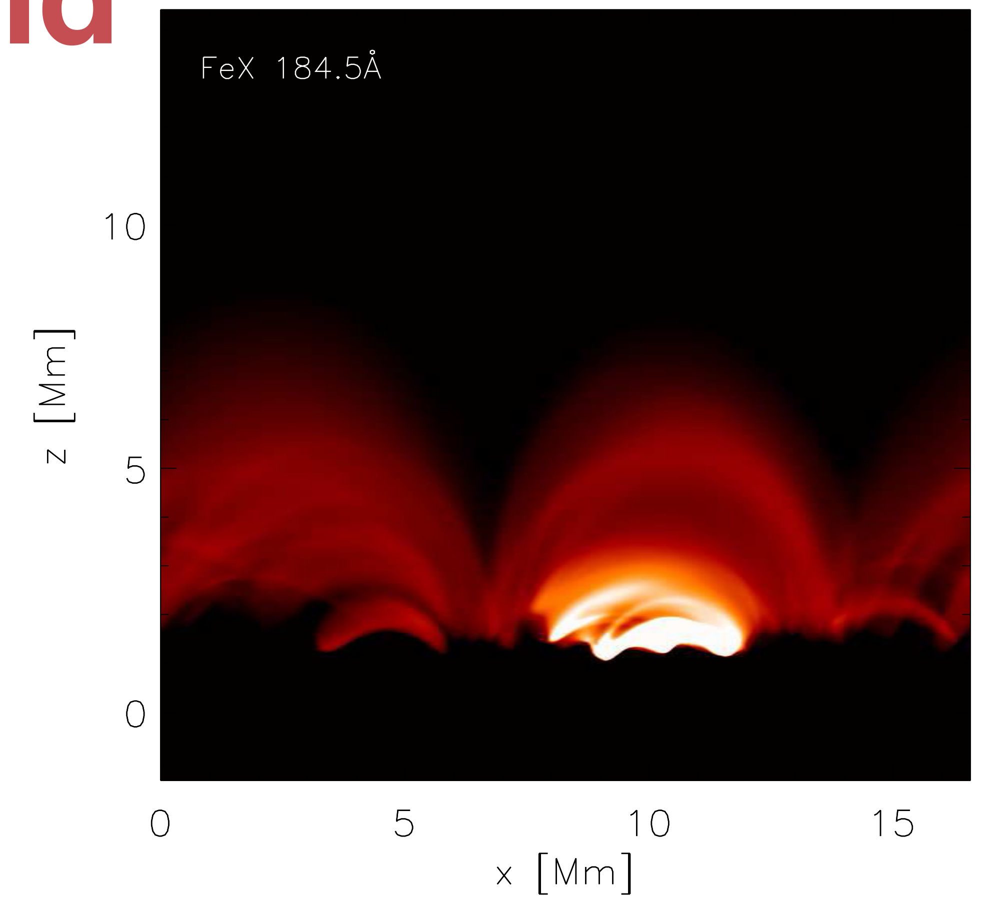
Frozen field



Mok et al. (2016)

See talk by Downs

3D MHD

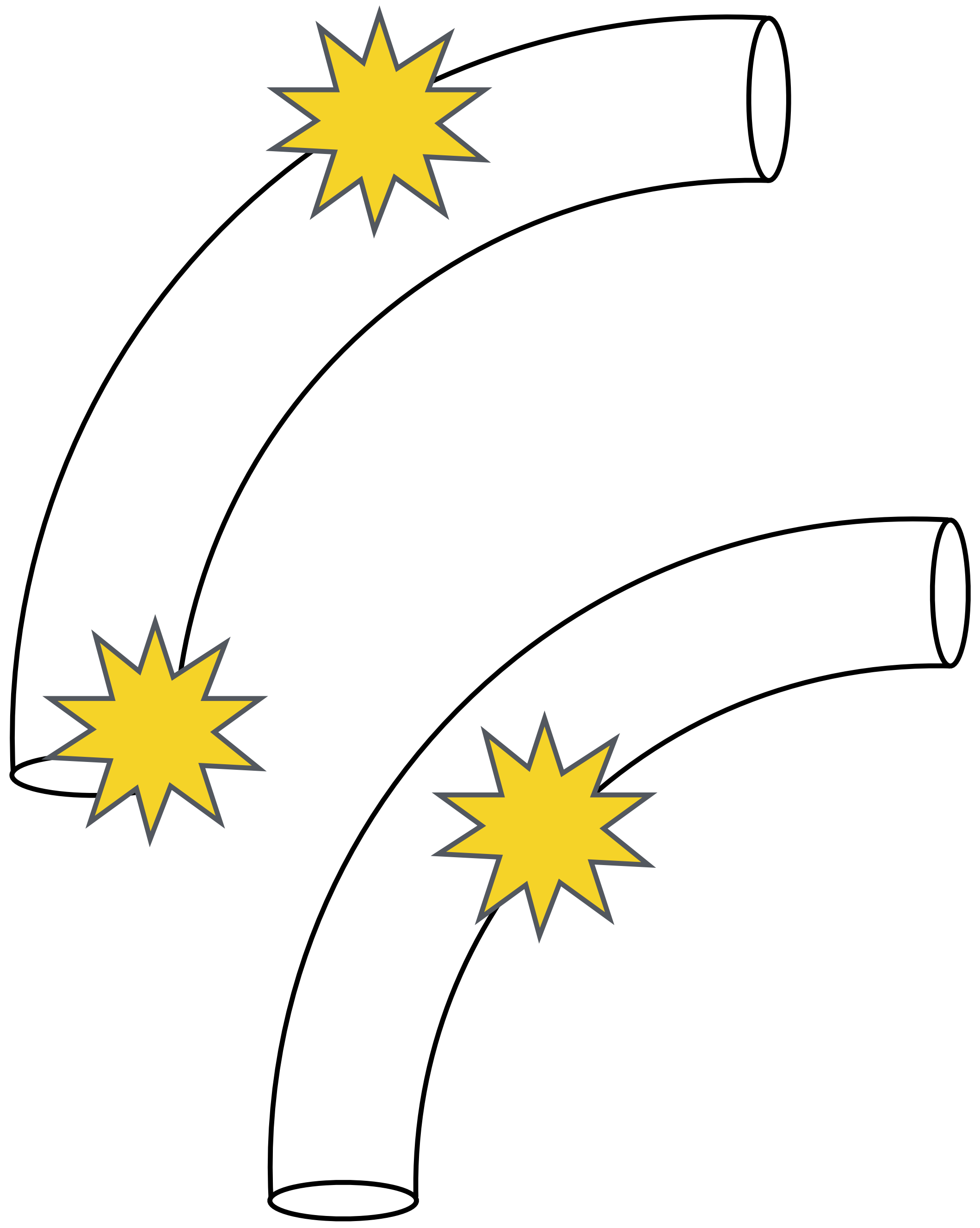


Testa et al. (2012)

- BIFROST (Gudiksen et al., 2011)
- MAS code (e.g. Mikić et al., 2017)
- MURaM (Rempel 2017)

1. Modeling Plasma Dynamics and Emission

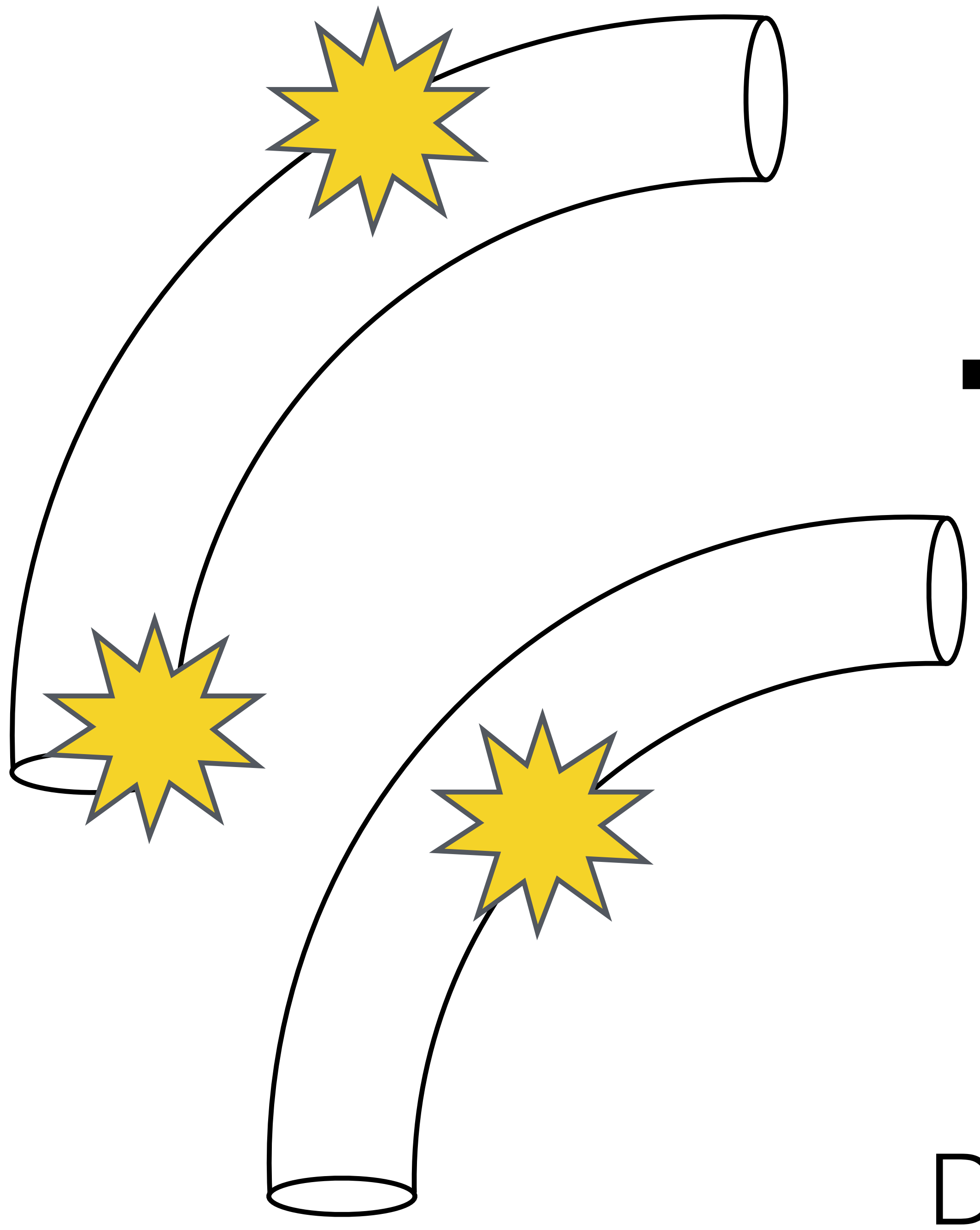
$T(s, t), n(s, t)$



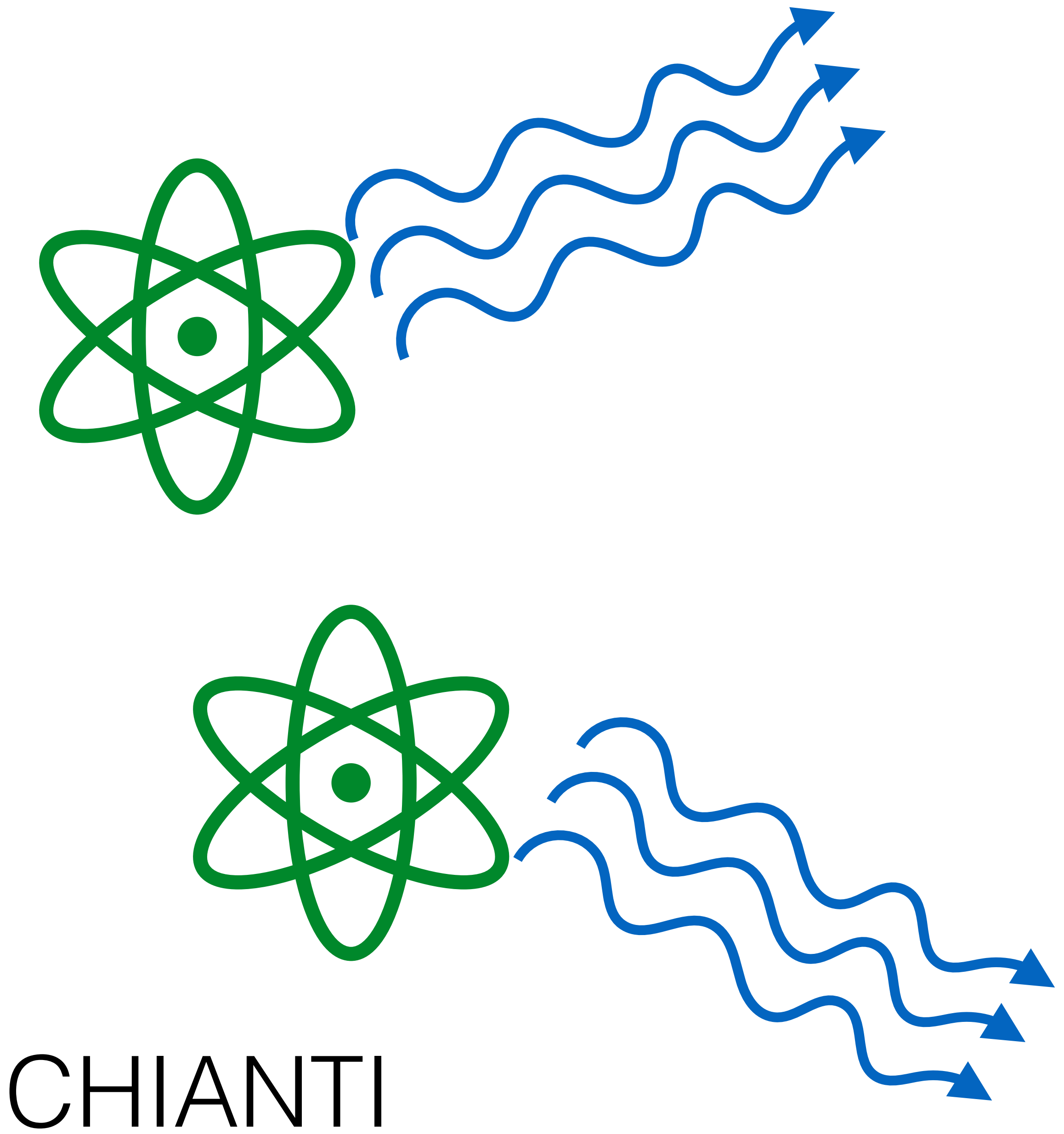
1. Modeling Plasma Dynamics and Emission

$T(s,t), n(s,t)$

$\varepsilon(\lambda, n, T)$



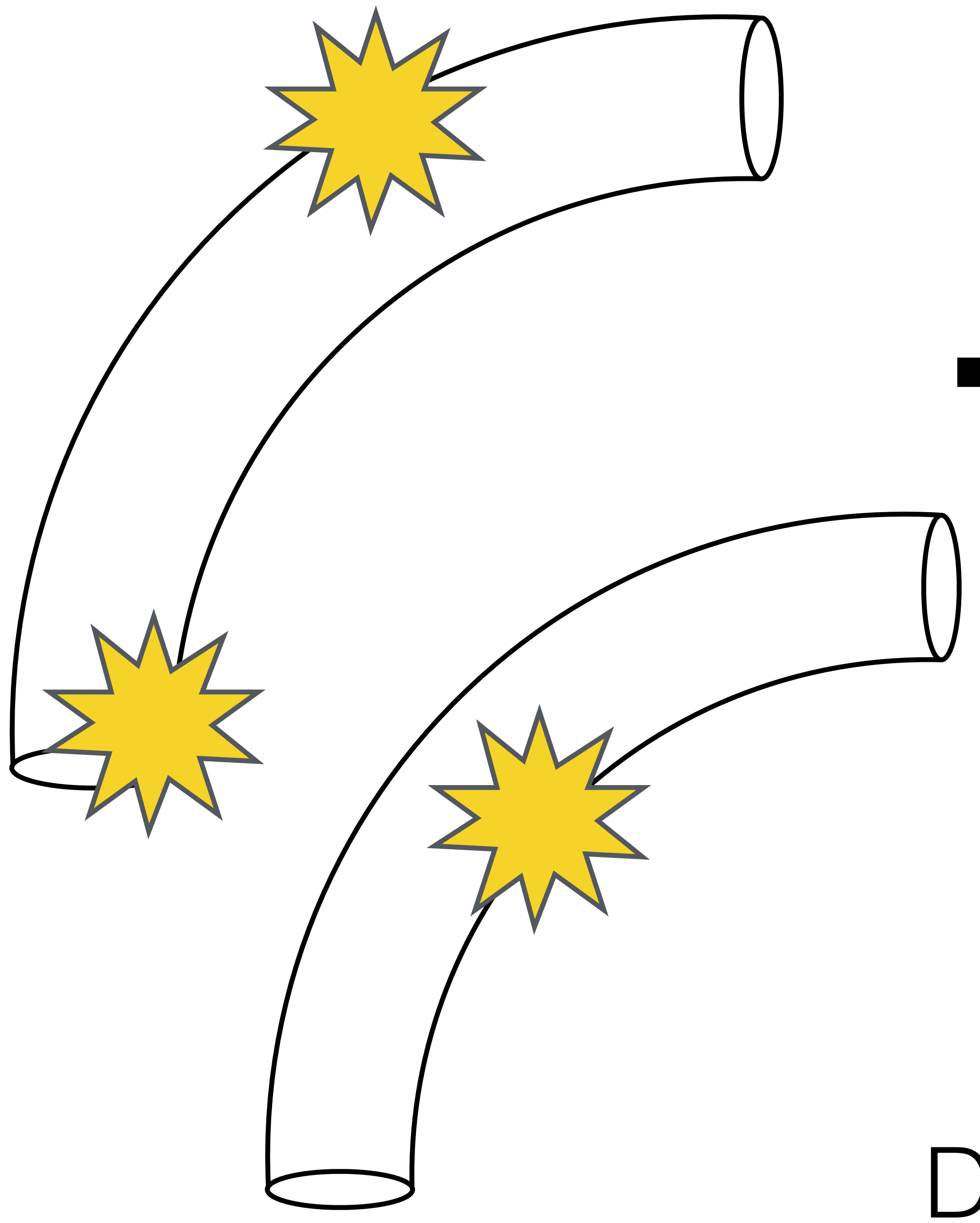
+



CHIANTI
(Dere et al., 1997,
Del Zanna et al., 2015)

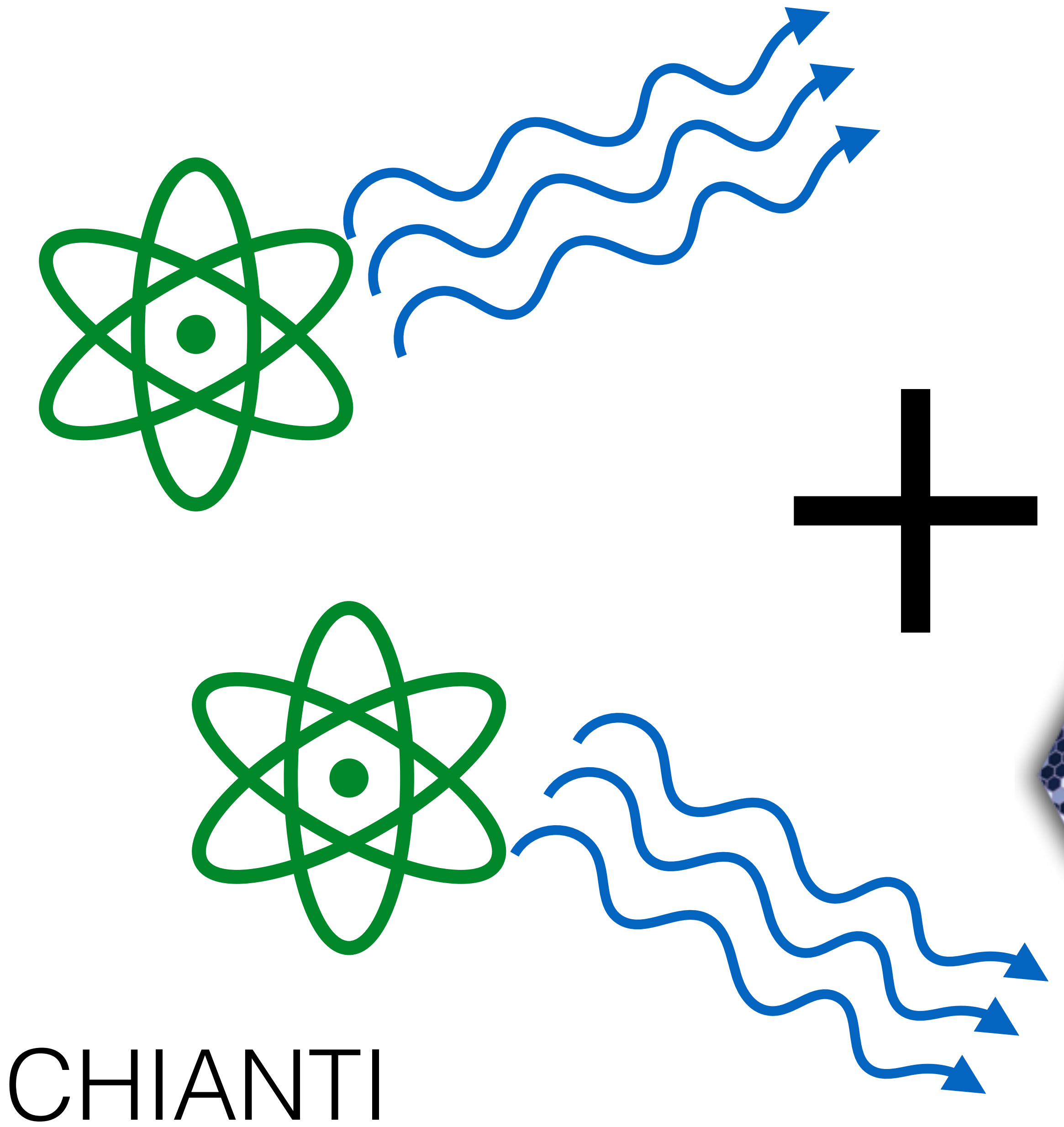
1. Modeling Plasma Dynamics and Emission

$T(s,t), n(s,t)$



+

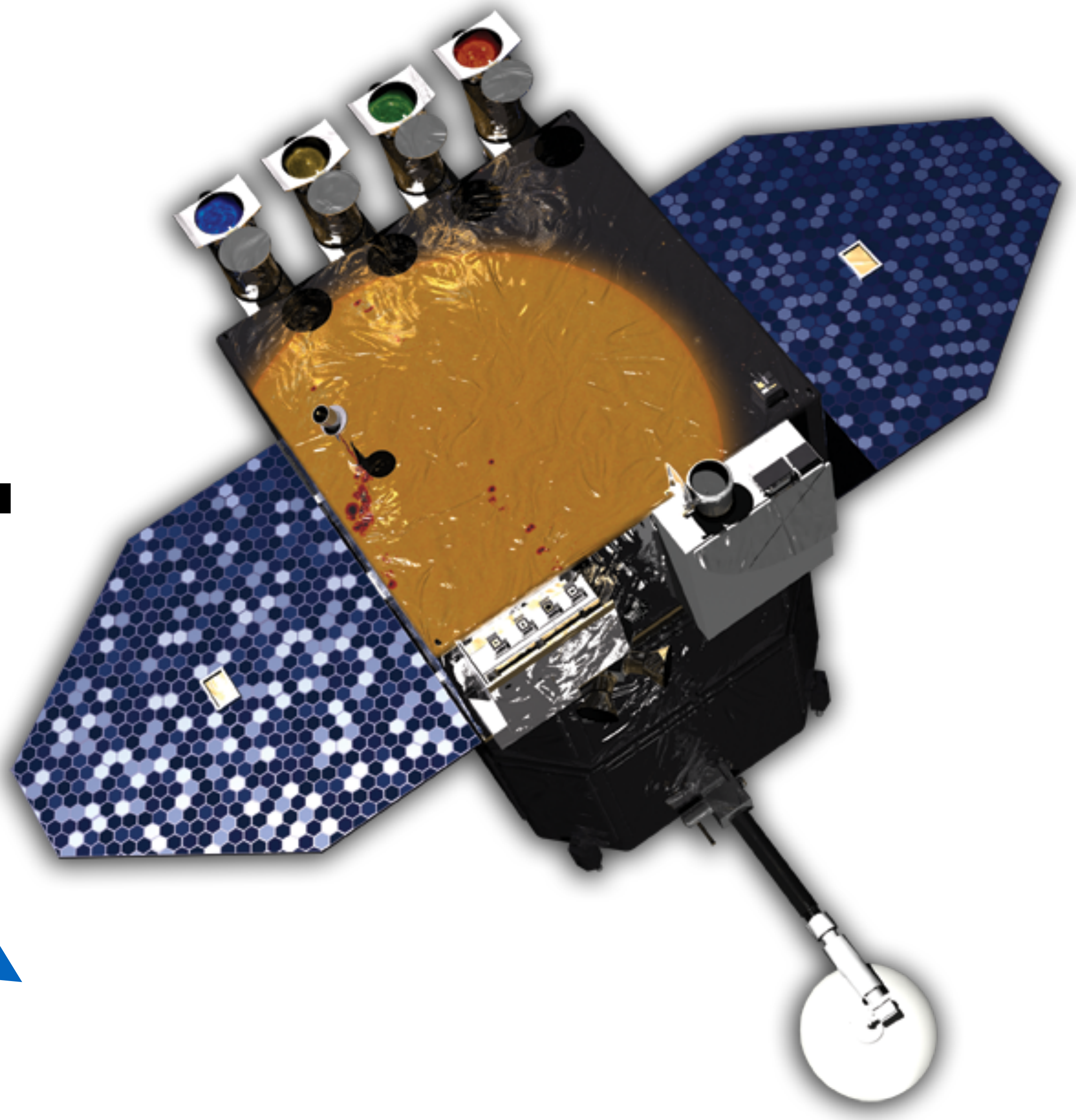
$\varepsilon(\lambda, n, T)$



CHIANTI
(Dere et al., 1997,
Del Zanna et al., 2015)

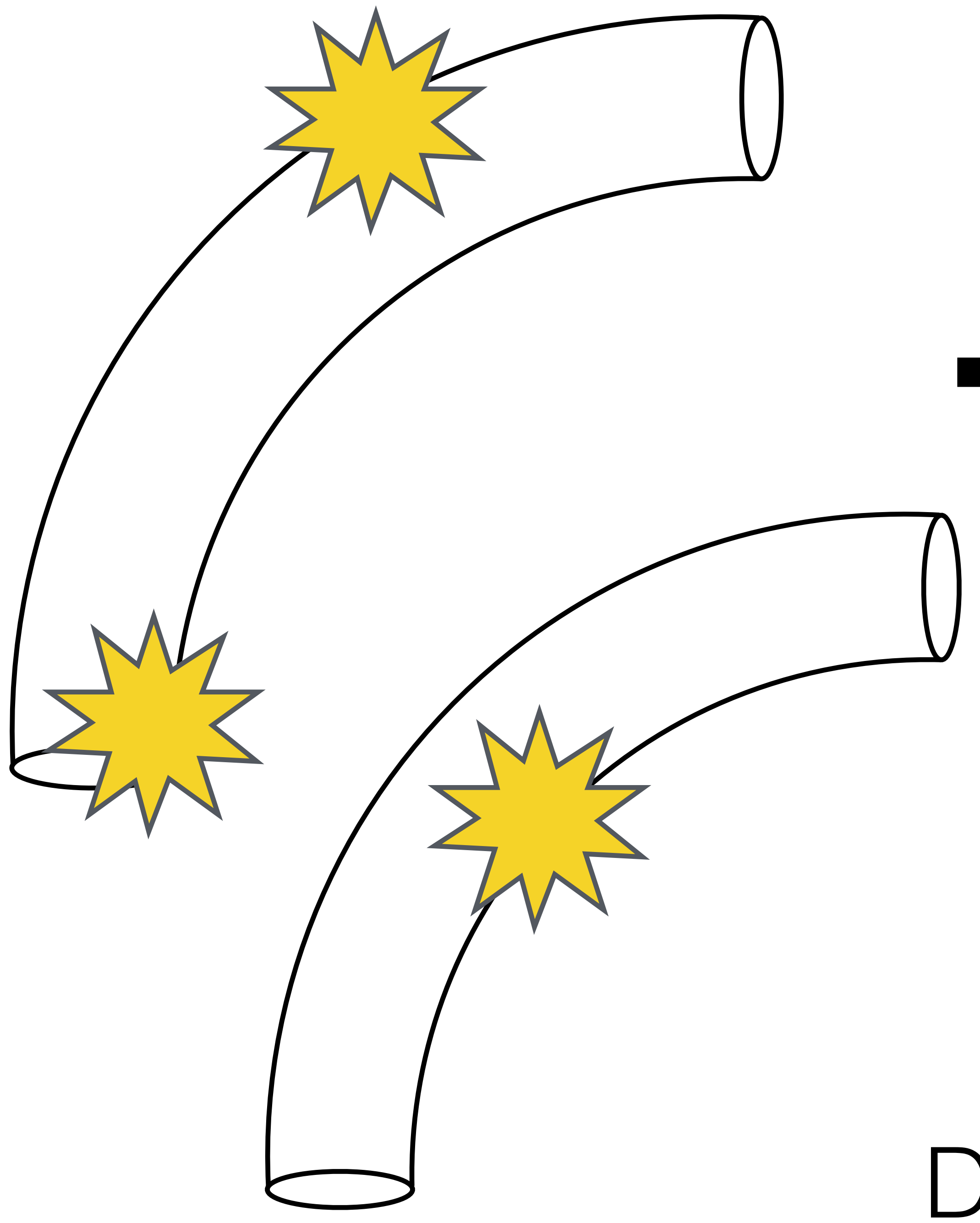
+

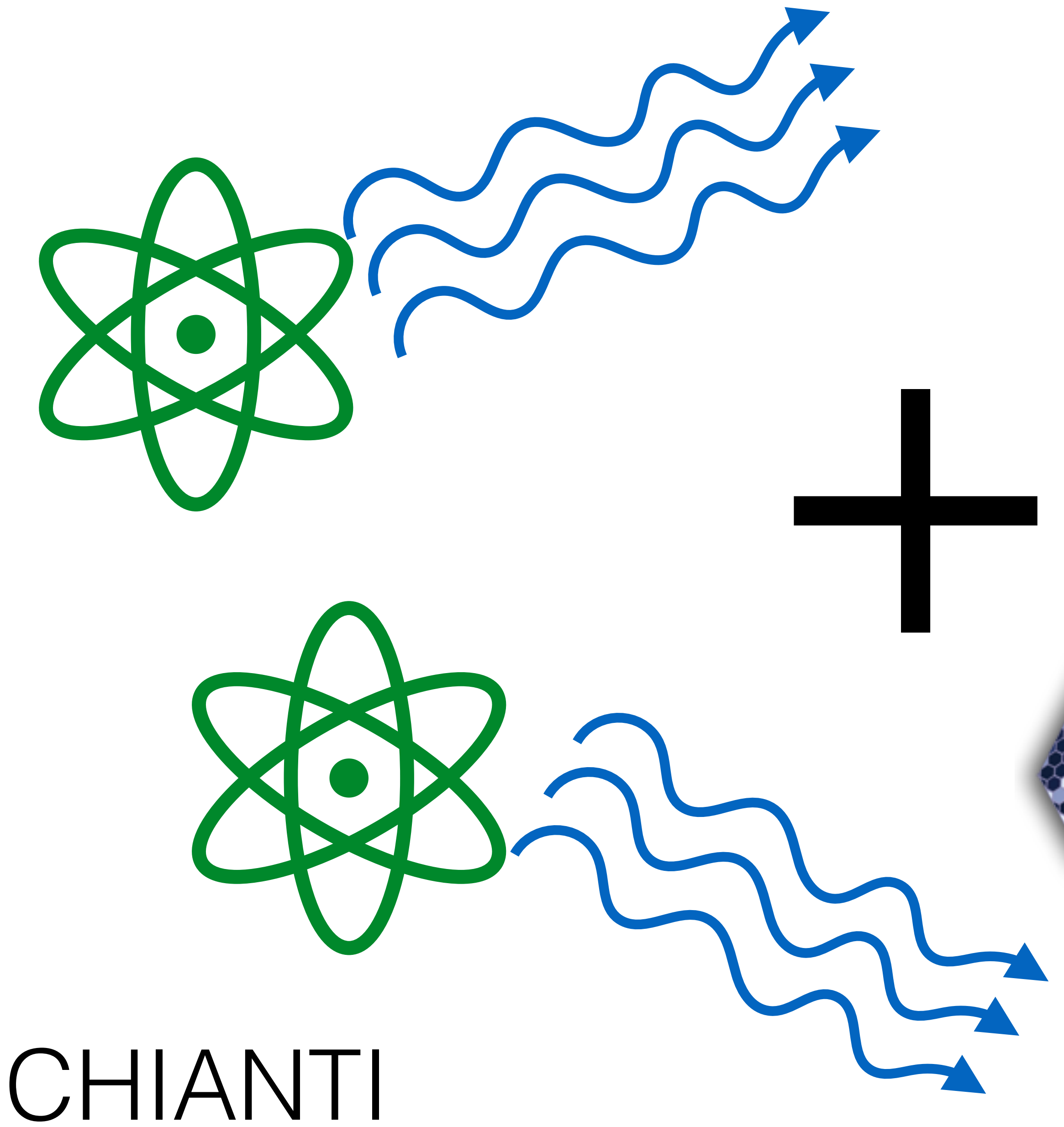
$R(\lambda)$



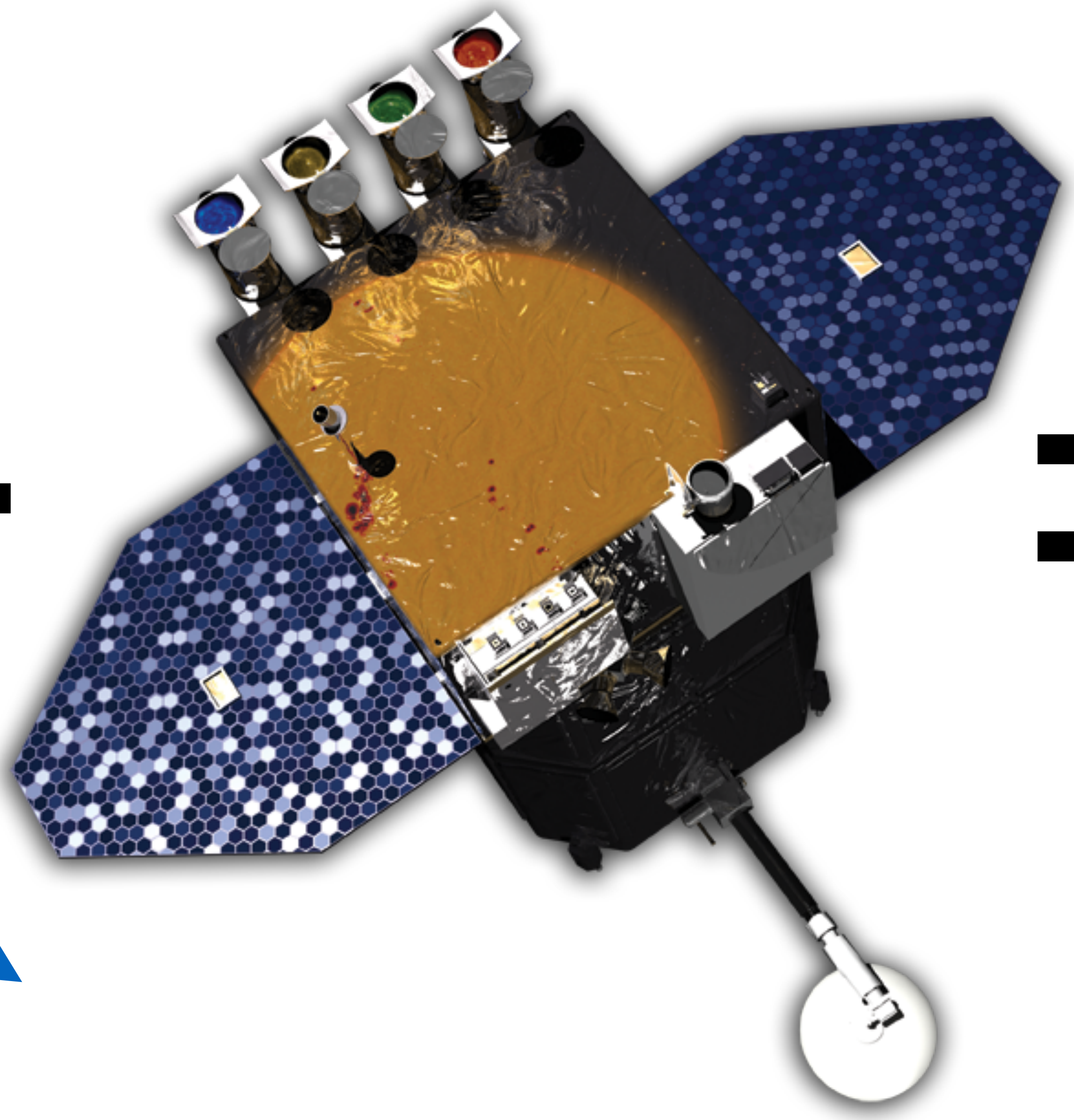
1. Modeling Plasma Dynamics and Emission

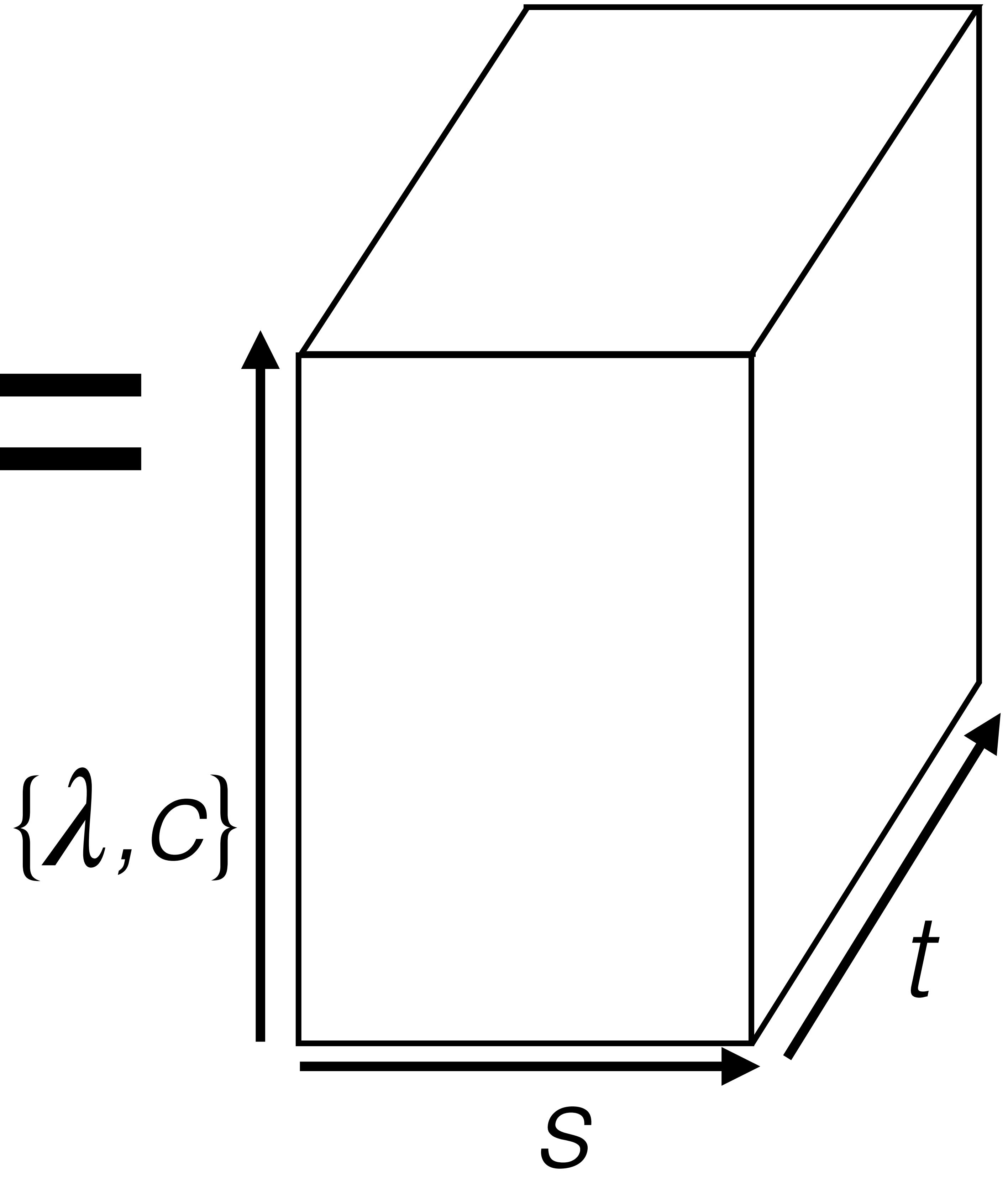
$T(s,t), n(s,t)$ + $\varepsilon(\lambda, n, T)$ + $R(\lambda)$ = $I(\lambda, s, t)$





CHIANTI
(Dere et al., 1997,
Del Zanna et al., 2015)





2. Diagnostics—EM Slope

Reduced representation of data that preserves signatures of heating frequency

2. Diagnostics—EM Slope

Reduced representation of data that preserves signatures of heating frequency

"Cool" EM Slope: $\mathrm{EM} \propto T^{a}$

$$\mathrm{EM} \sim n^2 \tau_r$$

$$\tau_r \sim T^{1-\alpha} n^{-1}$$

May change with *L*

$$\mathrm{EM} \sim n T^{1-\alpha}$$

$$T \propto n^2$$

$$\mathrm{EM} \sim T^{3/2-\alpha} \sim T^2$$

Jordan (1975,1976); Cargill (1994); Cargill and Klimchuk (2004); Cargill (2014)

2. Diagnostics—EM Slope

Reduced representation of data that preserves signatures of heating frequency

"Cool" EM Slope: $\mathrm{EM} \propto T^a$

$$\mathrm{EM} \sim n^2\tau_r$$

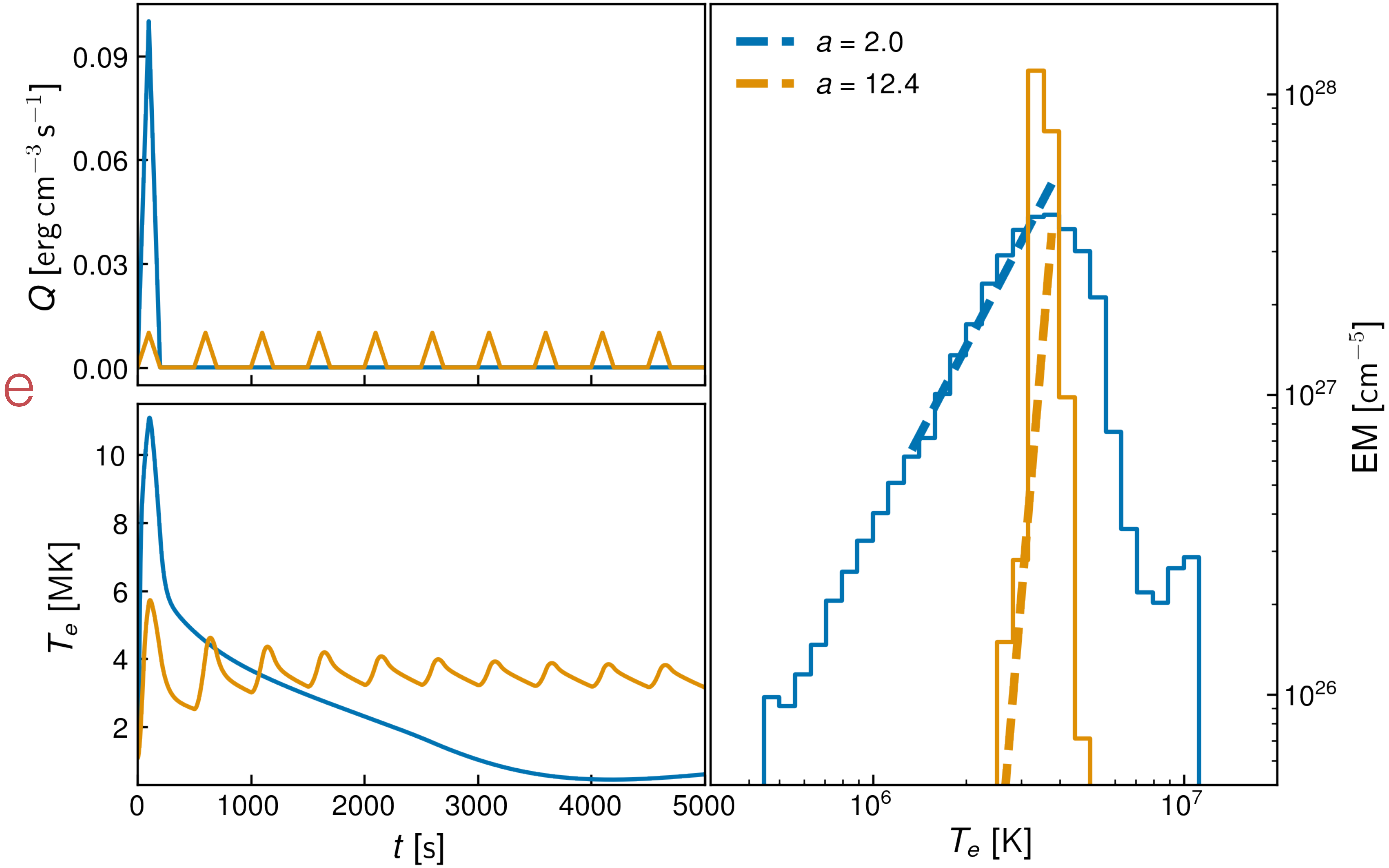
$$\tau_r \sim T^{1-\alpha}n^{-1}$$

$$\mathrm{EM} \sim nT^{1-\alpha}$$

$$T \propto n^2$$

May change with L

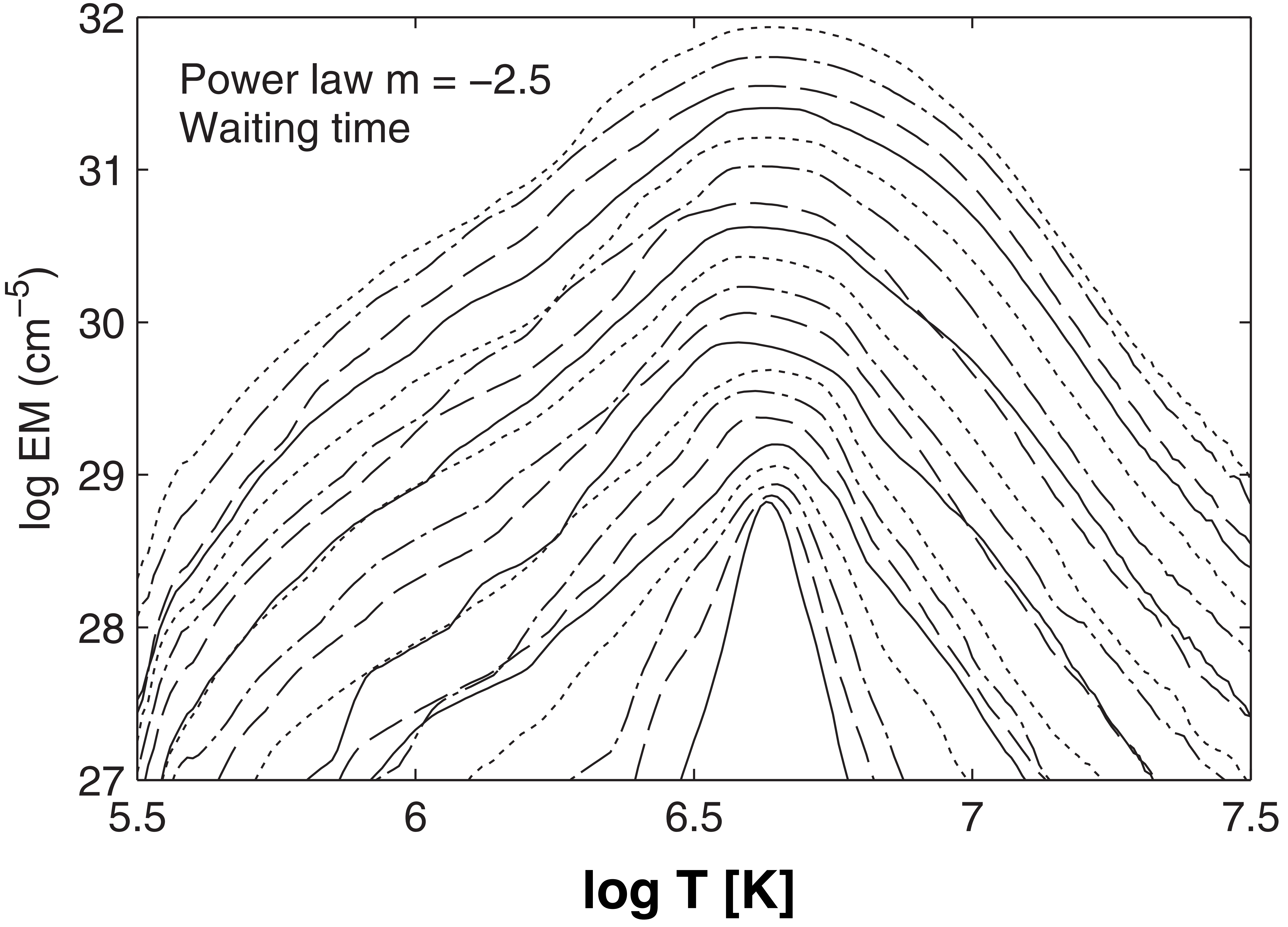
$$\mathrm{EM} \sim T^{3/2-\alpha} \sim T^2$$

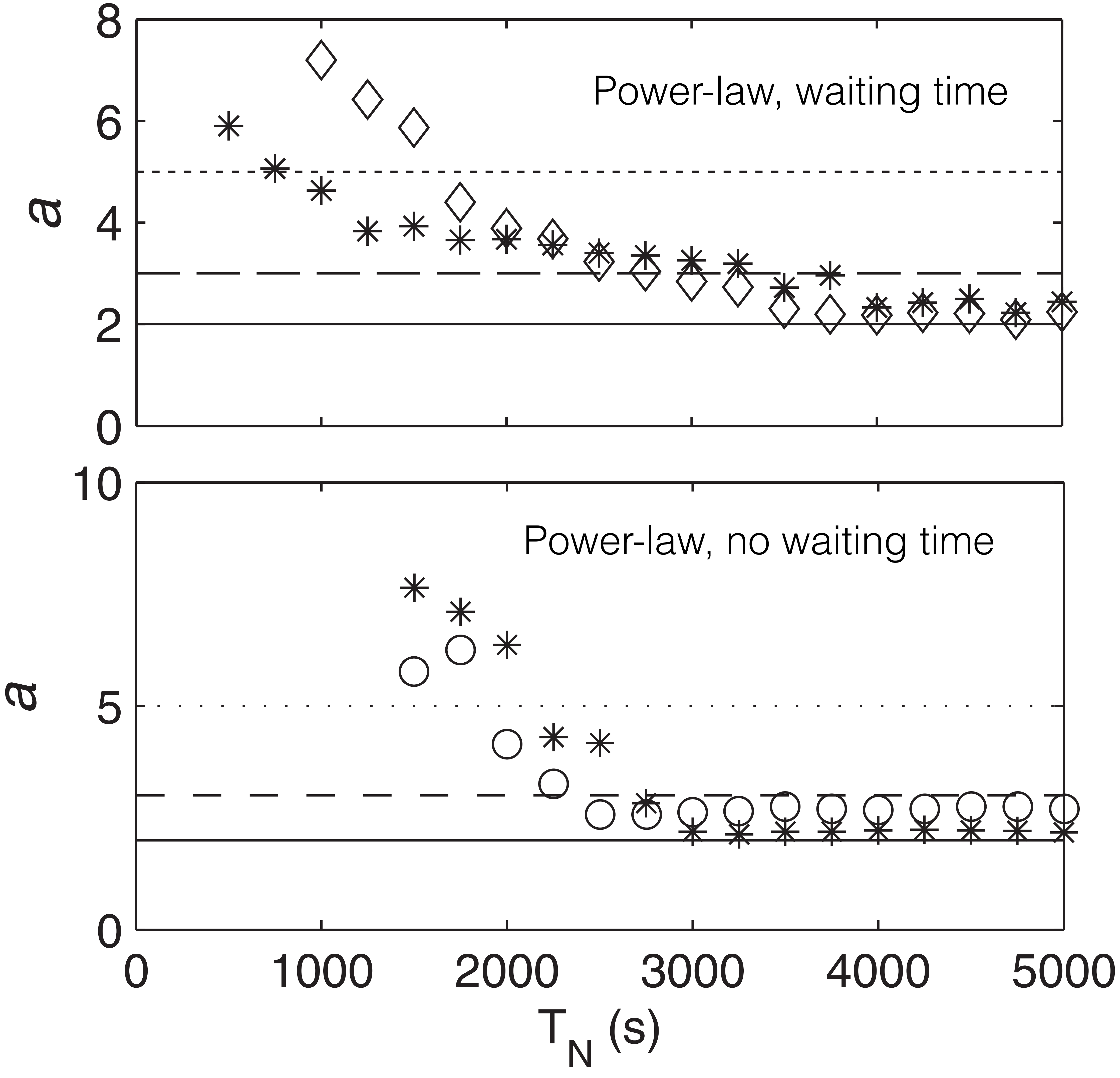


Jordan (1975,1976); Cargill (1994); Cargill and Klimchuk (2004); Cargill (2014)

2. Diagnostics—EM Slope

Cargill (2014)





2. Diagnostics—EM Slope

Table 3.2 Summary of observational and modeling studies that have used the emission measure slope, a, as a diagnostic for the underlying energy deposition. The approximate range of observed slopes is $2 \lesssim a \lesssim 5$. Adapted from Table 3 of Bradshaw et al. (2012).

Reference	Type	Slope (a)	Temperature range [K]
Warren et al. (2011)	observation	3.26	10^6–$10^{6.6}$
	model	2.17	
Winebarger et al. (2011)	observation	3.2	10^6–$10^{6.5}$
Tripathi et al. (2011)	observation	2.08–2.47	$10^{5.5}$–$10^{6.55}$
		2.05–2.7[a]	
Mulu-Moore et al. (2011a)	model[b]	1.6–2	10^6–T_{peak}[c]
		2–2.3	
Warren et al. (2012)	observation	1.7–4.5	10^6–$10^{6.6}$
Schmelz & Pathak (2012)	observation	1.91–5.17	10^6–T_{peak}[d]
Bradshaw et al. (2012)	model	0.81–2.56	10^6–T_{peak}[e]
Reep et al. (2013)	model	0.88–4.56	10^6–T_{peak}[f]
Cargill (2014)	model	2–8	T_0–$10^{6.6}$[g]
Del Zanna et al. (2015b)	observation[h]	4.4 ± 0.4	10^6–3×10^6
		4.6 ± 0.4	

[a] DEM(T_e) computed from background-subtracted observations.
[b] Intensities were modeled using photospheric (first row) and coronal (second row) abundances.
[c] T_{peak} varied from $10^{6.6}$ K to $10^{6.8}$ K.
[d] T_{peak} varied from $10^{6.3}$ K to $10^{6.8}$ K.
[e] T_{peak} varied from $10^{5.85}$ K to $10^{7.35}$ K.
[f] T_{peak} varied from $10^{6.35}$ K to $10^{6.65}$ K.
[g] a is computed for 12 different values of T_0 between $10^6$and$10^{6.25}$ and averaged.
[h] The slope was computed in every pixel of active region NOAA 11193 once when it first appeared (first row) and then again after one rotation (second row).

Barnes PhD Thesis (2019)

2. Diagnostics—EM Slope

Table 3.2 Summary of observational and modeling studies that have used the emission measure slope, a, as a diagnostic for the underlying energy deposition. The approximate range of observed slopes is $2 \lesssim a \lesssim 5$. Adapted from Table 3 of Bradshaw et al. (2012).

Reference	Type	Slope (a)	Temperature range [K]
Warren et al. (2011)	observation	3.26	10^6–$10^{6.6}$
	model	2.17	
Winebarger et al. (2011)	observation	3.2	10^6–$10^{6.5}$
Tripathi et al. (2011)	observation	2.08–2.47	$10^{5.5}$–$10^{6.55}$
		2.05–2.7[a]	
Mulu-Moore et al. (2011a)	model[b]	1.6–2	10^6–T_{peak}[c]
		2–2.3	
Warren et al. (2012)	observation	1.7–4.5	10^6–$10^{6.6}$
Schmelz & Pathak (2012)	observation	1.91–5.17	10^6–T_{peak}[d]
Bradshaw et al. (2012)	model	0.81–2.56	10^6–T_{peak}[e]
Reep et al. (2013)	model	0.88–4.56	10^6–T_{peak}[f]
Cargill (2014)	model	2–8	T_0–$10^{6.6}$[g]
Del Zanna et al. (2015b)	observation[h]	4.4 ± 0.4	10^6–3×10^6
		4.6 ± 0.4	

[a] DEM(T_e) computed from background-subtracted observations.
[b] Intensities were modeled using photospheric (first row) and coronal (second row) abundances.
[c] T_{peak} varied from $10^{6.6}$ K to $10^{6.8}$ K.
[d] T_{peak} varied from $10^{6.3}$ K to $10^{6.8}$ K.
[e] T_{peak} varied from $10^{5.85}$ K to $10^{7.35}$ K.
[f] T_{peak} varied from $10^{6.35}$ K to $10^{6.65}$ K.
[g] a is computed for 12 different values of T_0 between $10^6$and$10^{6.25}$ and averaged.
[h] The slope was computed in every pixel of active region NOAA 11193 once when it first appeared (first row) and then again after one rotation (second row).

$$1.7 \lesssim a \lesssim 5$$

Barnes PhD Thesis (2019)

2. Diagnostics — EM Slope

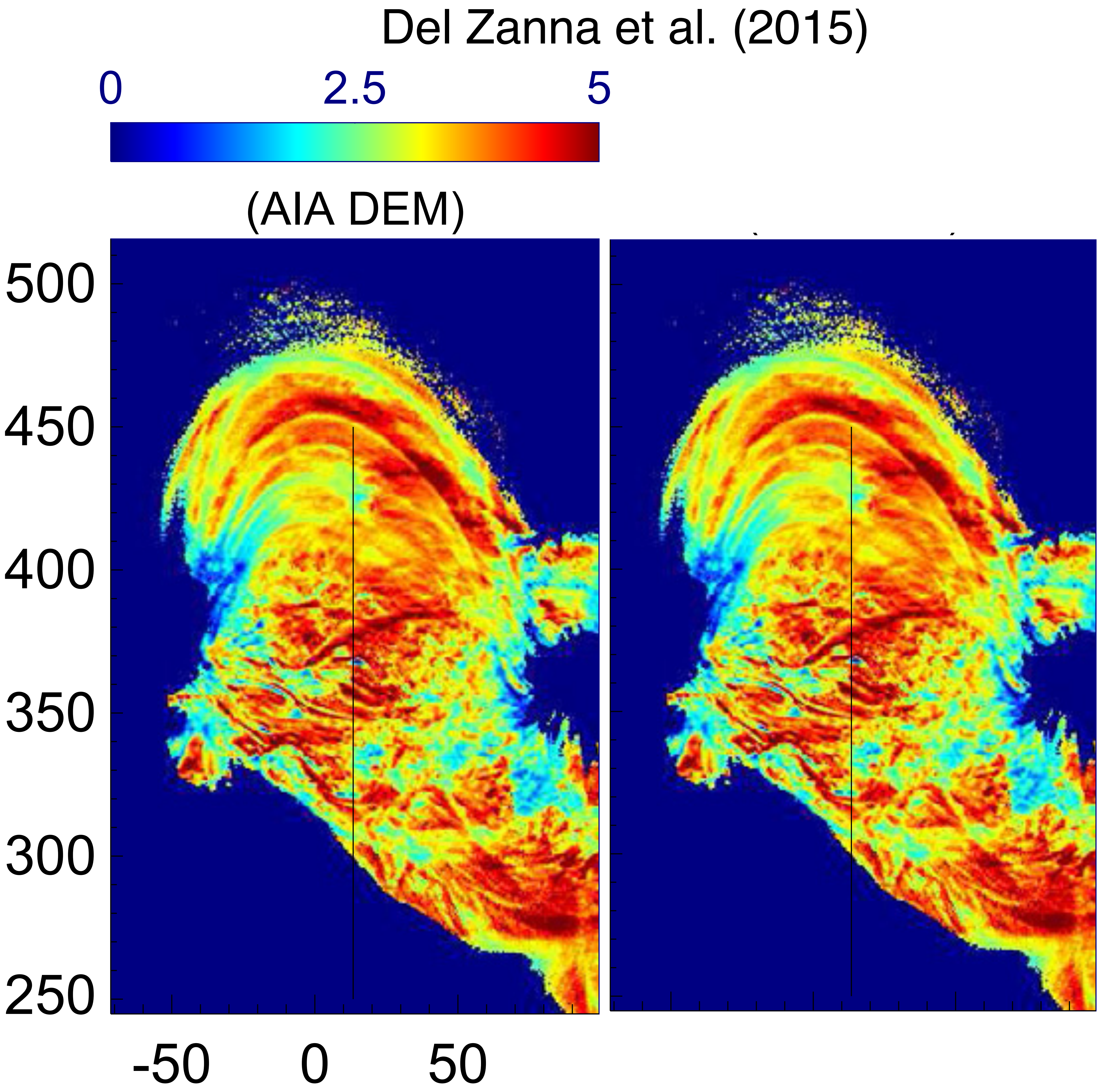
Table 3.2 Summary of observational and modeling studies that have used the emission measure slope, a, as a diagnostic for the underlying energy deposition. The approximate range of observed slopes is $2 \lesssim a \lesssim 5$. Adapted from Table 3 of Bradshaw et al. (2012).

Reference	Type	Slope (a)	Temperature range [K]
Warren et al. (2011)	observation	3.26	10^6–$10^{6.6}$
	model	2.17	
Winebarger et al. (2011)	observation	3.2	10^6–$10^{6.5}$
Tripathi et al. (2011)	observation	2.08–2.47	$10^{5.5}$–$10^{6.55}$
		2.05–2.7[a]	
Mulu-Moore et al. (2011a)	model[b]	1.6–2	10^6–T_{peak}[c]
		2–2.3	
Warren et al. (2012)	observation	1.7–4.5	10^6–$10^{6.6}$
Schmelz & Pathak (2012)	observation	1.91–5.17	10^6–T_{peak}[d]
Bradshaw et al. (2012)	model	0.81–2.36	10^6–T_{peak}[e]
Reep et al. (2013)	model	0.88–4.56	10^6–T_{peak}[f]
Cargill (2014)	model	2–8	T_0–$10^{6.6}$[g]
Del Zanna et al. (2015b)	observation[h]	4.4 ± 0.4	10^6–3×10^6
		4.6 ± 0.4	

[a] DEM(T_e) computed from background-subtracted observations.
[b] Intensities were modeled using photospheric (first row) and coronal (second row) abundances.
[c] T_{peak} varied from $10^{6.6}$ K to $10^{6.8}$ K.
[d] T_{peak} varied from $10^{6.3}$ K to $10^{6.8}$ K.
[e] T_{peak} varied from $10^{5.85}$ K to $10^{7.35}$ K.
[f] T_{peak} varied from $10^{6.35}$ K to $10^{6.65}$ K.
[g] a is computed for 12 different values of T_0 between $10^6$and$10^{6.25}$ and averaged.
[h] The slope was computed in every pixel of active region NOAA 11193 once when it first appeared (first row) and then again after one rotation (second row).

$$1.7 \lesssim a \lesssim 5$$

Barnes PhD Thesis (2019)



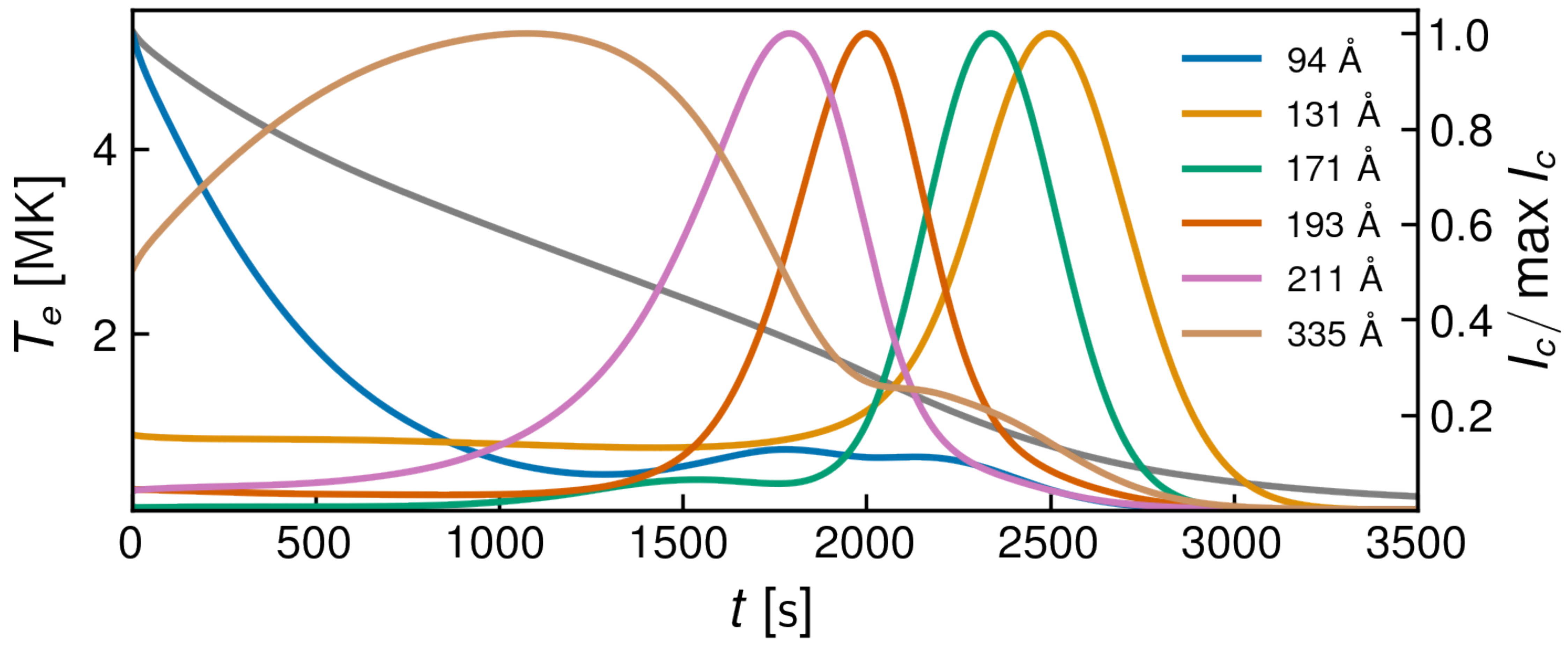
2. Diagnostics—Time Lag

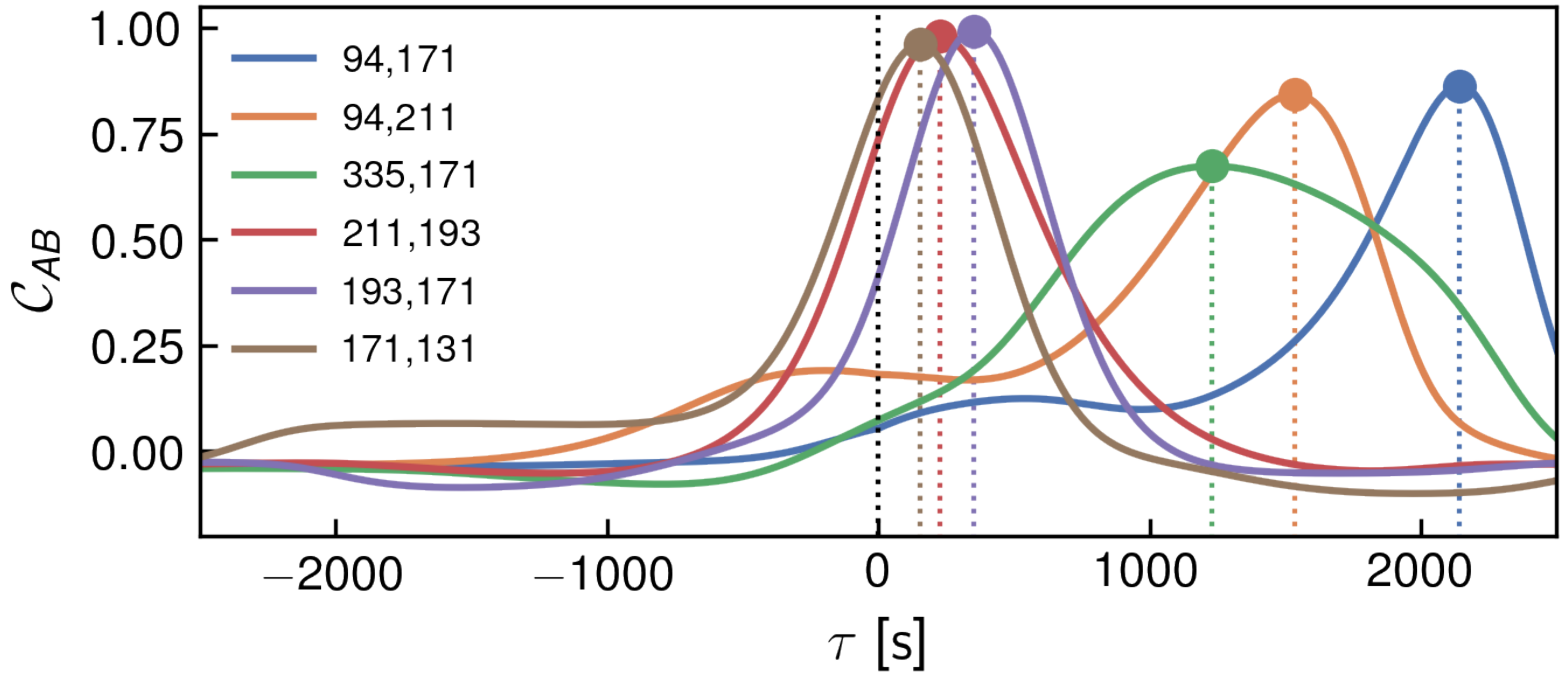
Reduced representation of data that preserves signatures of heating frequency

2. Diagnostics—Time Lag

Reduced representation of data that preserves signatures of heating frequency

Time Lag



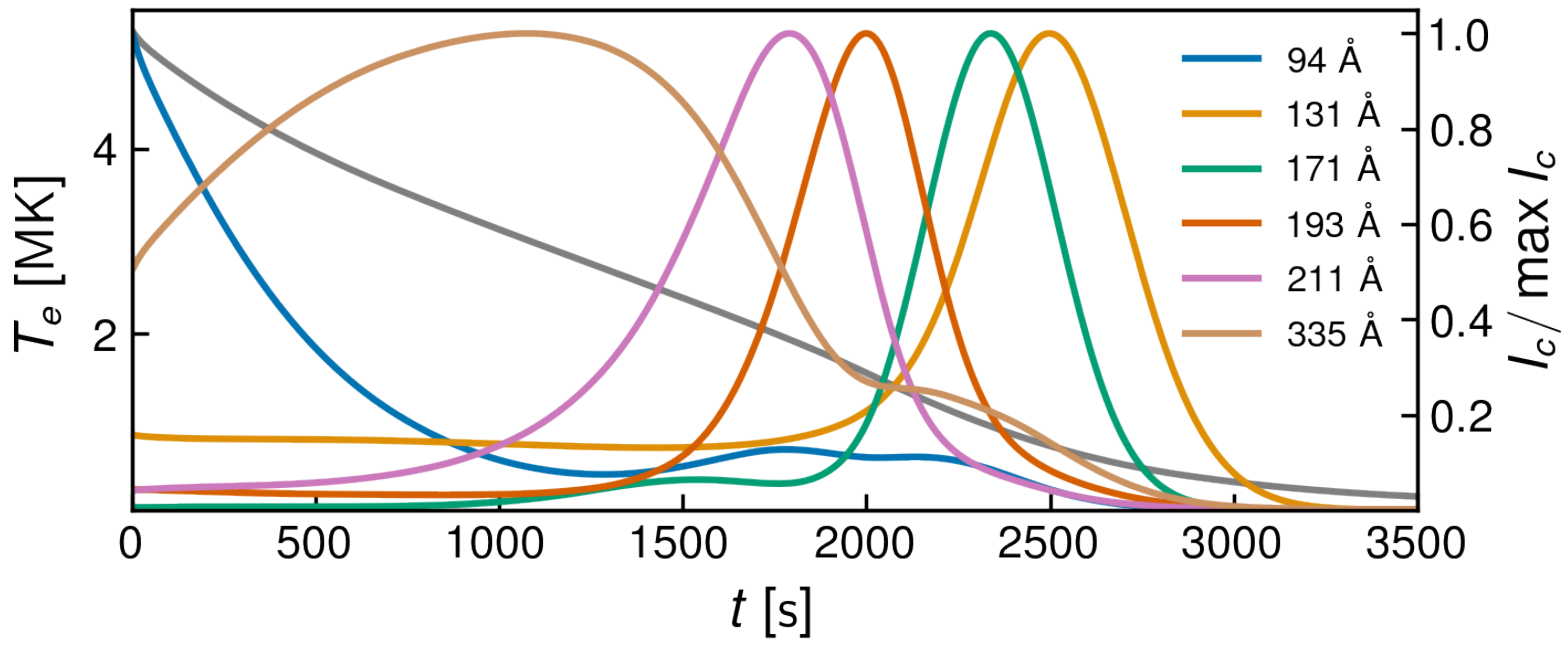


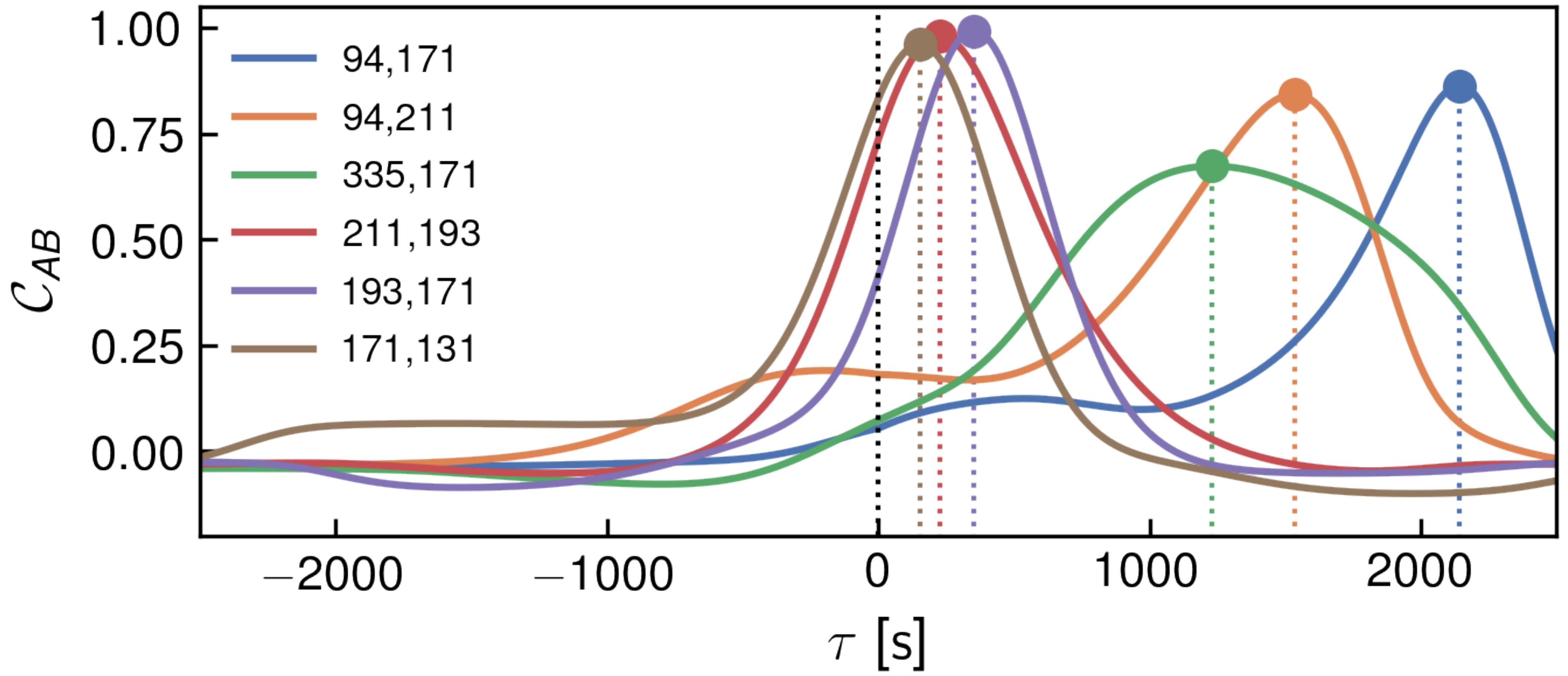
See talk by Viall

2. Diagnostics—Time Lag

Reduced representation of data that preserves signatures of heating frequency

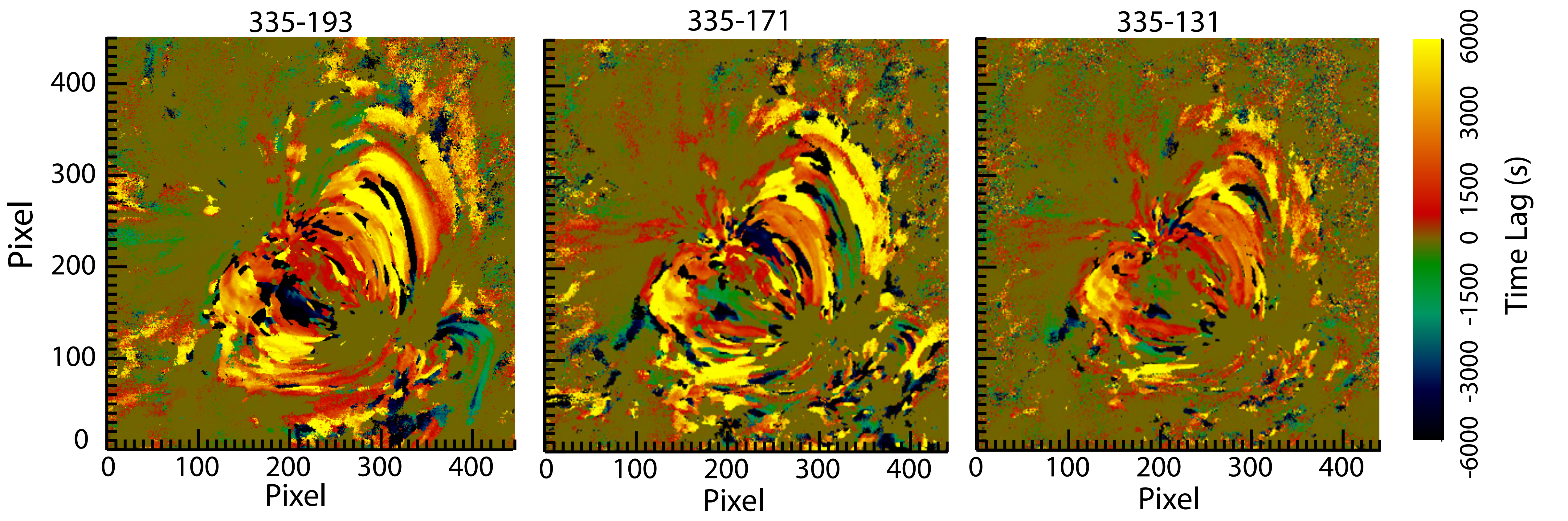
Time Lag





See talk by Viall

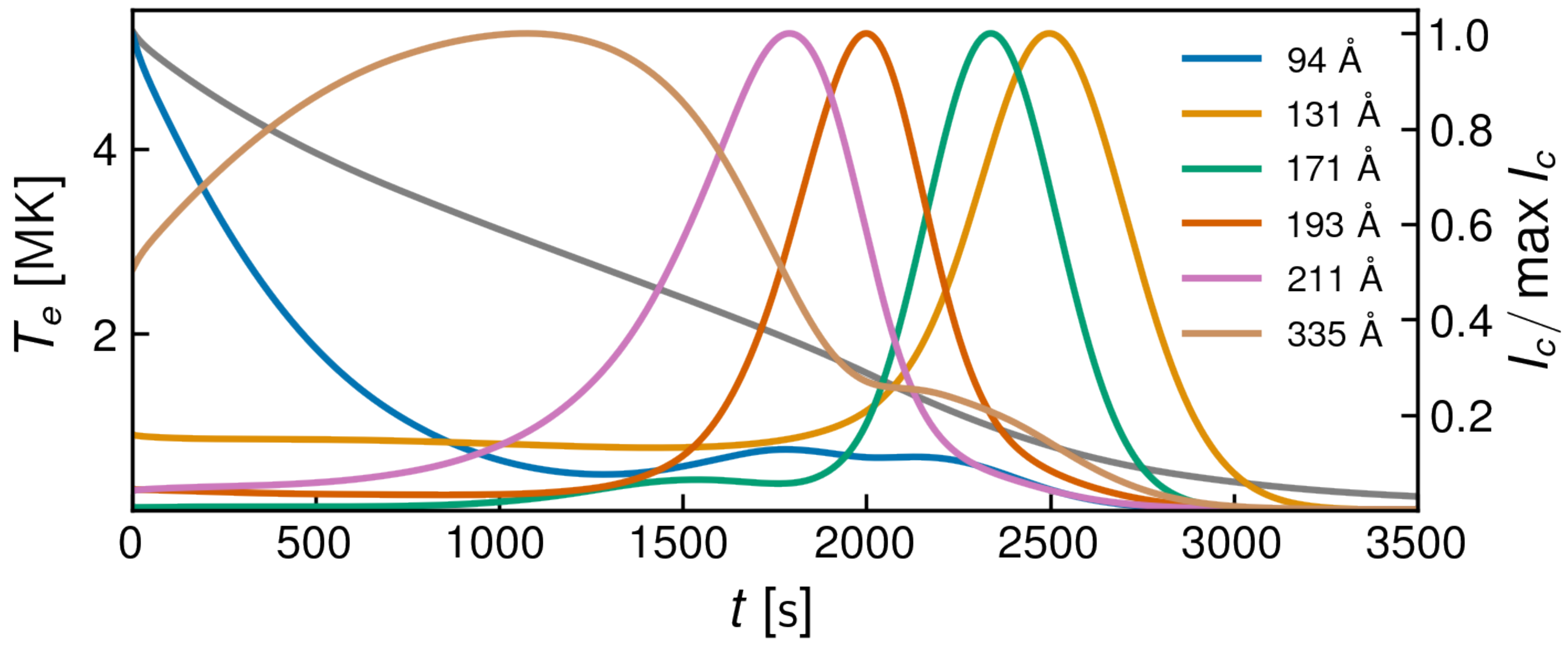
Viall and Klimchuk (2012)

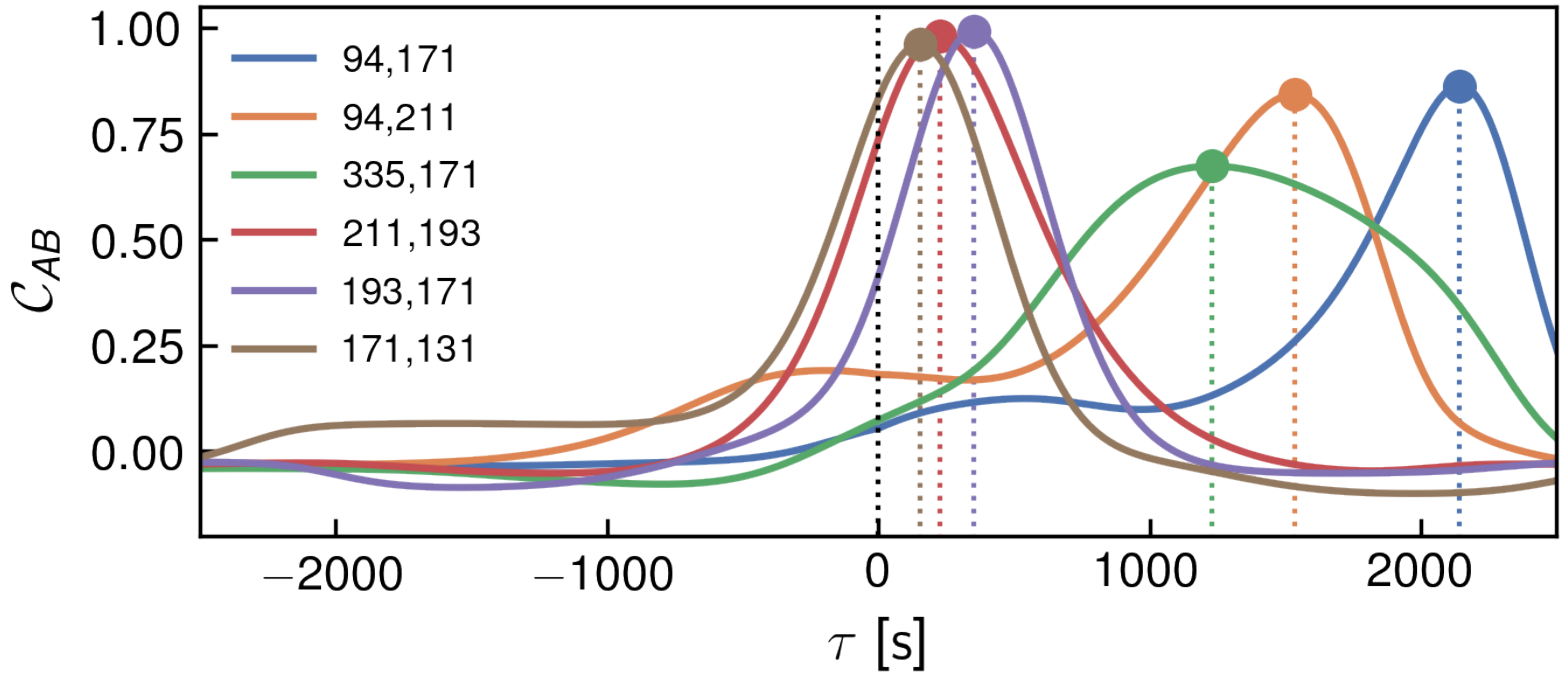


2. Diagnostics—Time Lag

Reduced representation of data that preserves signatures of heating frequency

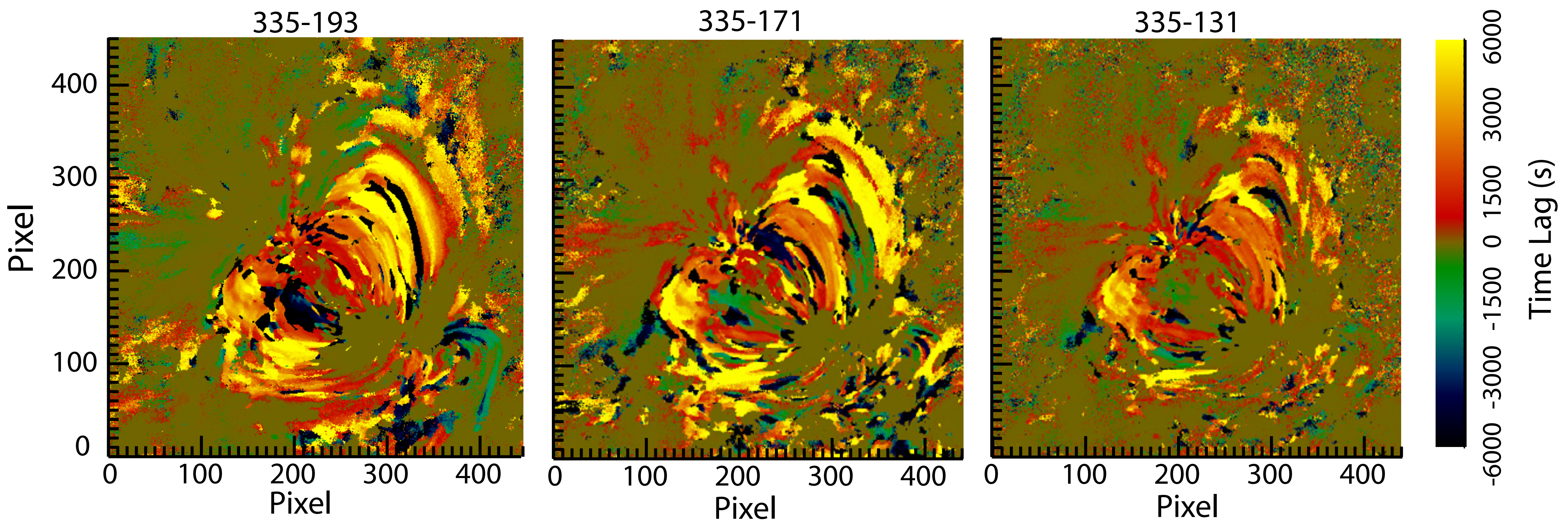
Time Lag



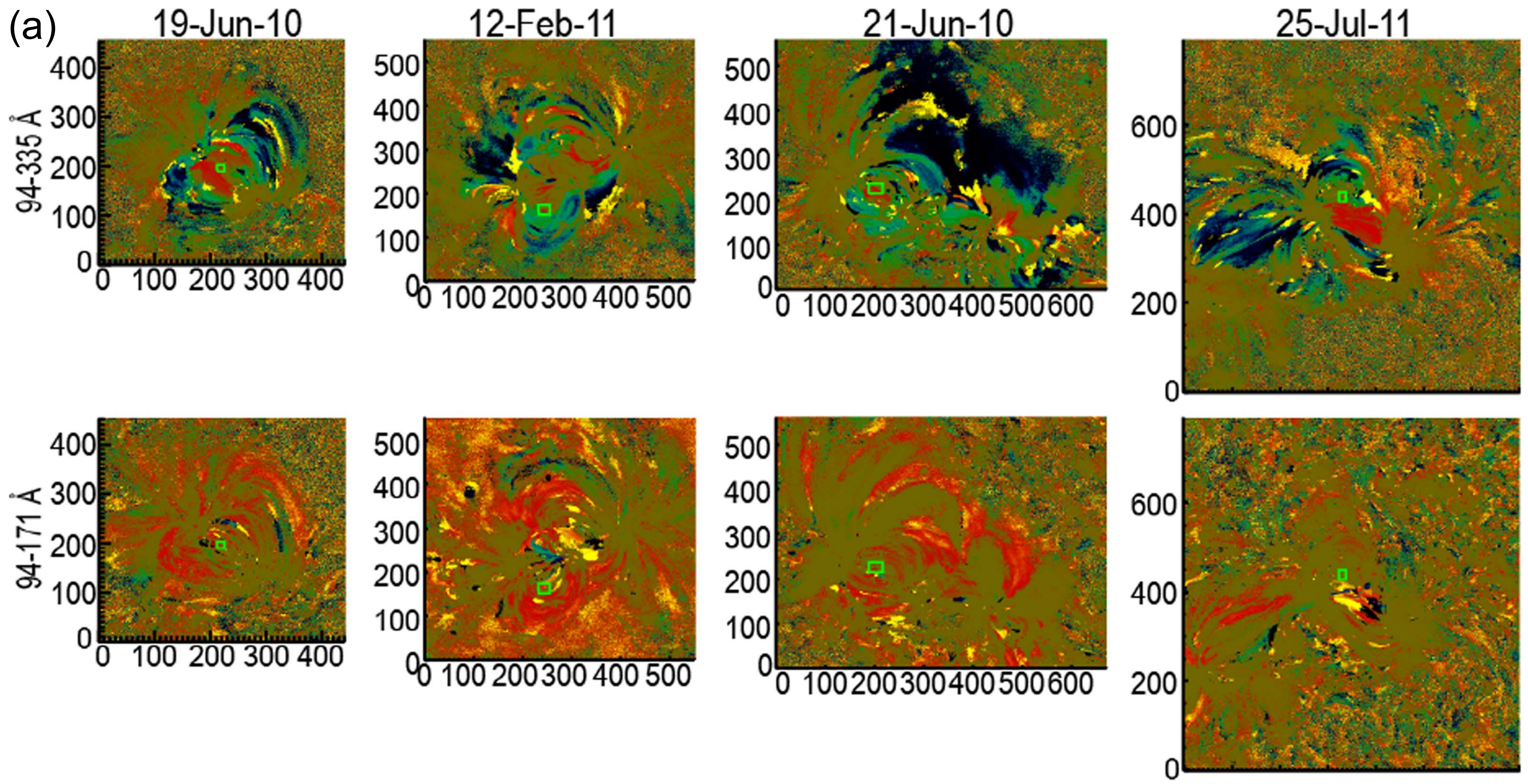


See talk by Viall

Viall and Klimchuk (2012)

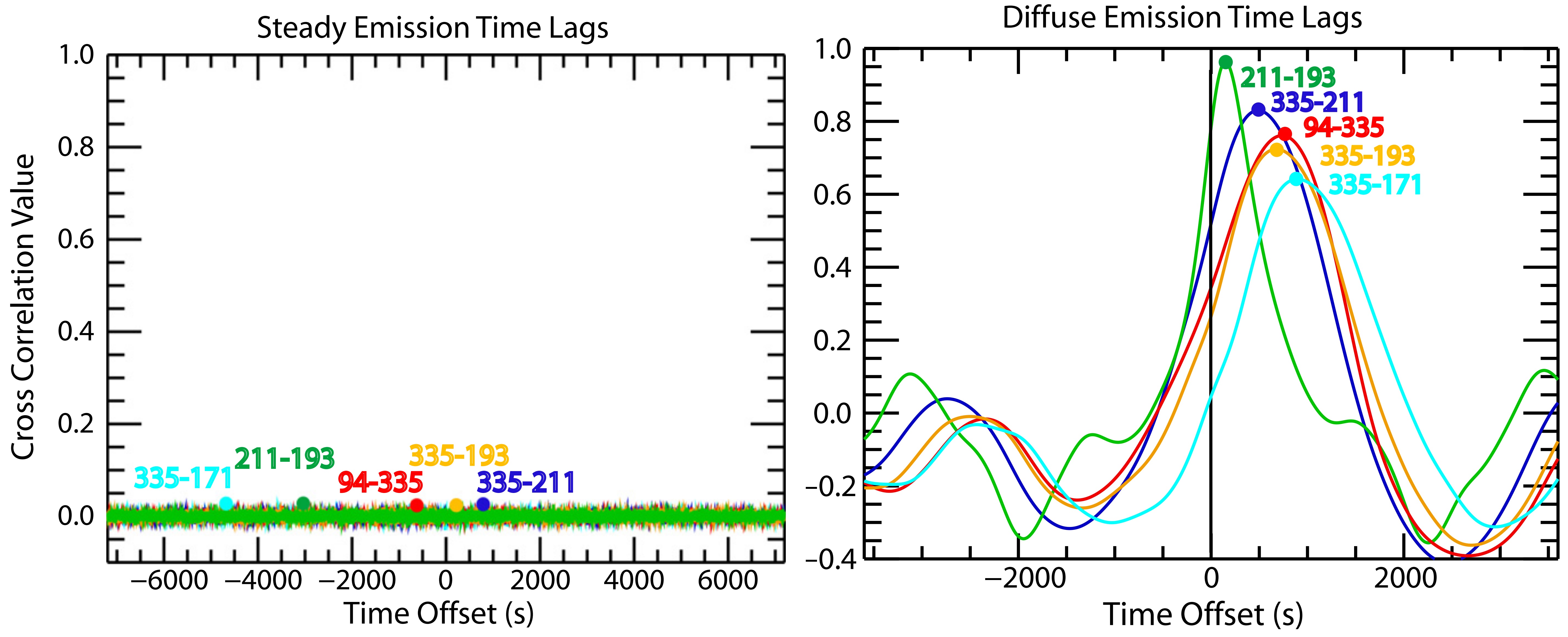


Viall and Klimchuk (2017)



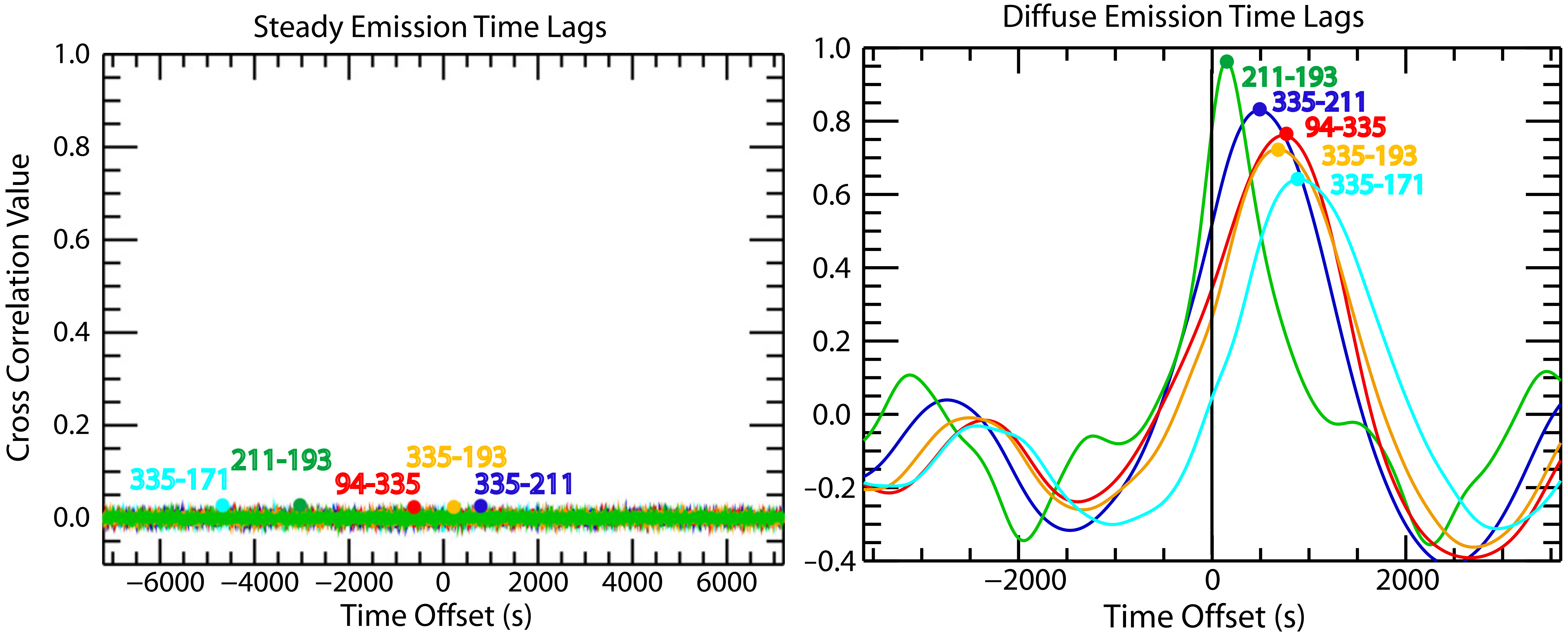
2. Diagnostics—Time Lag

Viall and Klimchuk (2013)

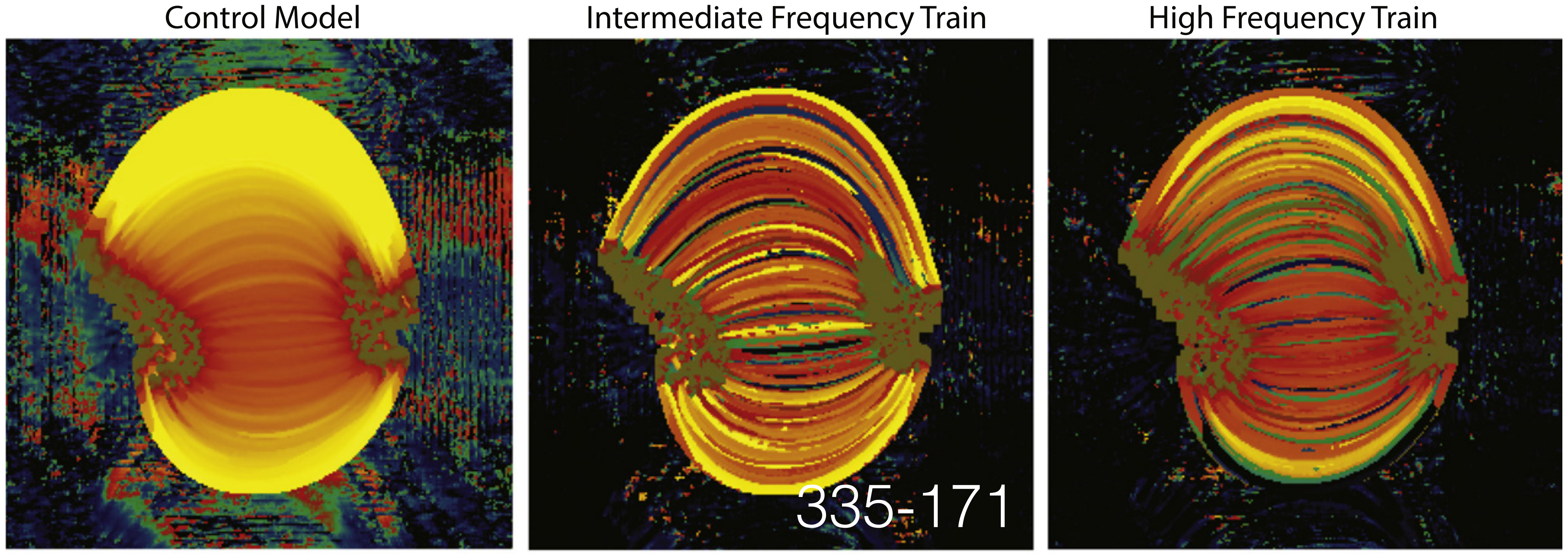


2. Diagnostics—Time Lag

Viall and Klimchuk (2013)

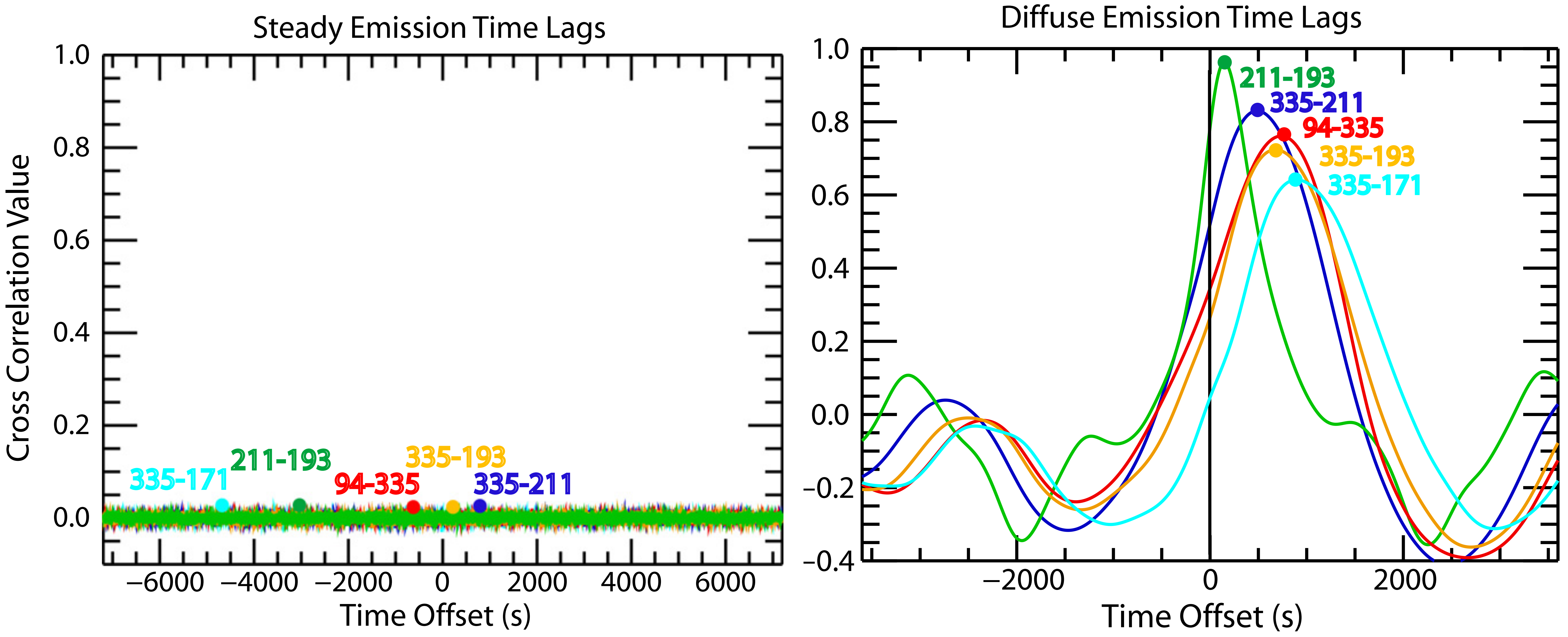


Bradshaw and Viall (2016)

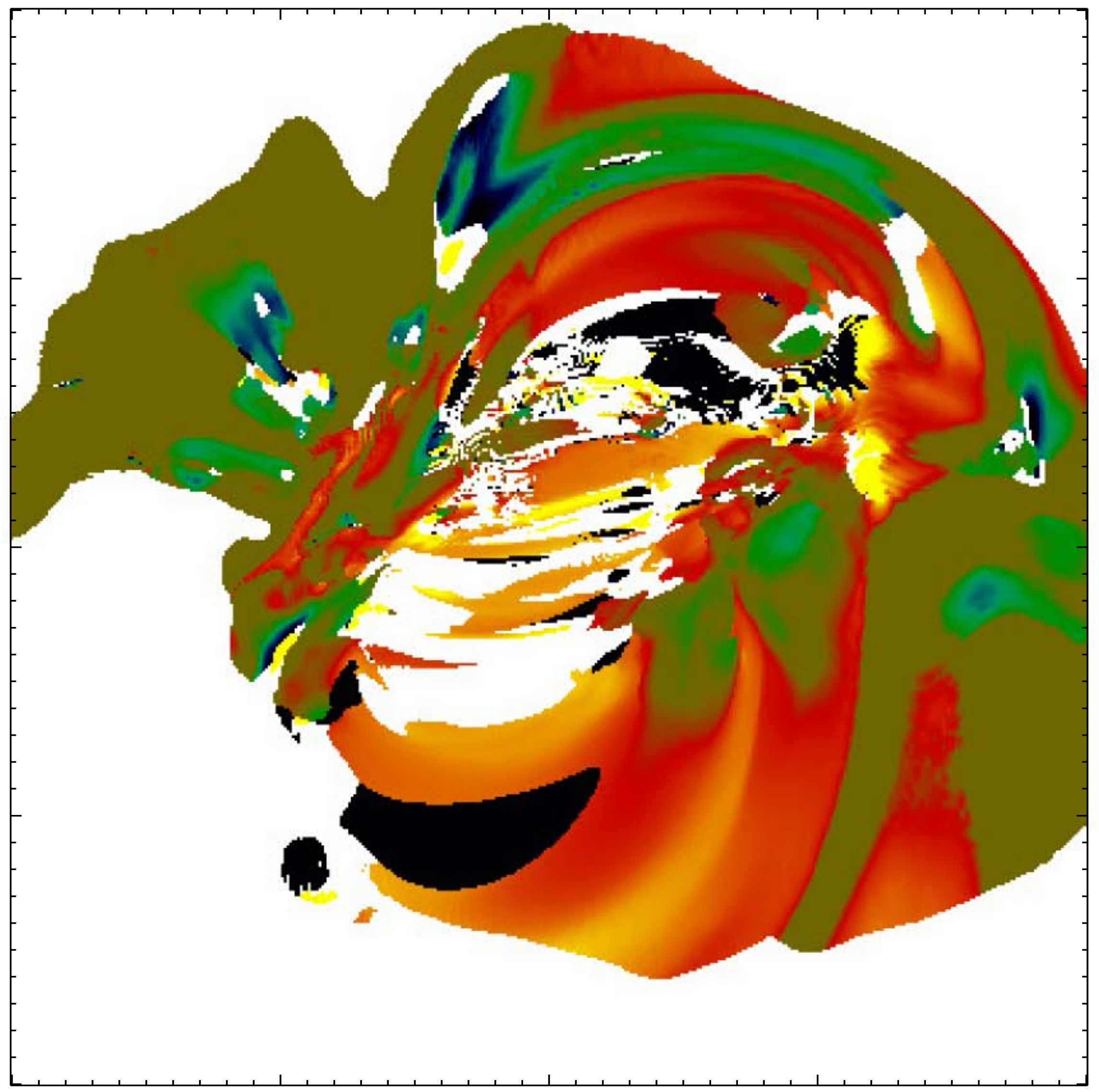


2. Diagnostics—Time Lag

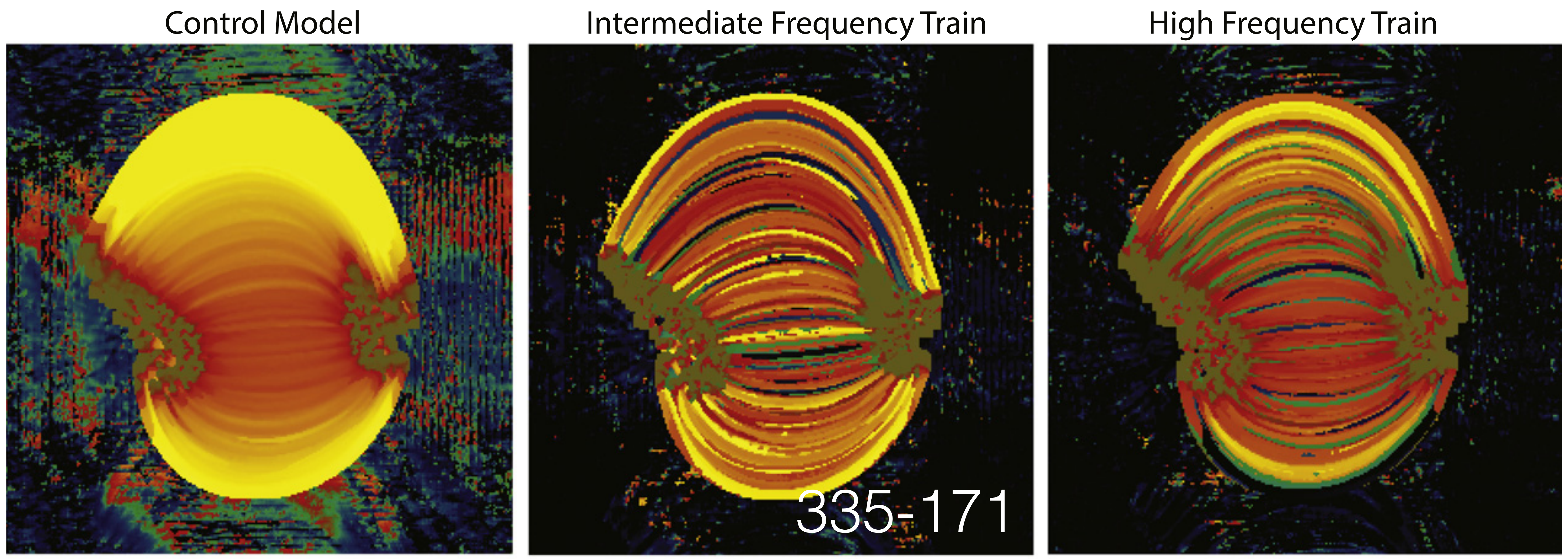
Viall and Klimchuk (2013)



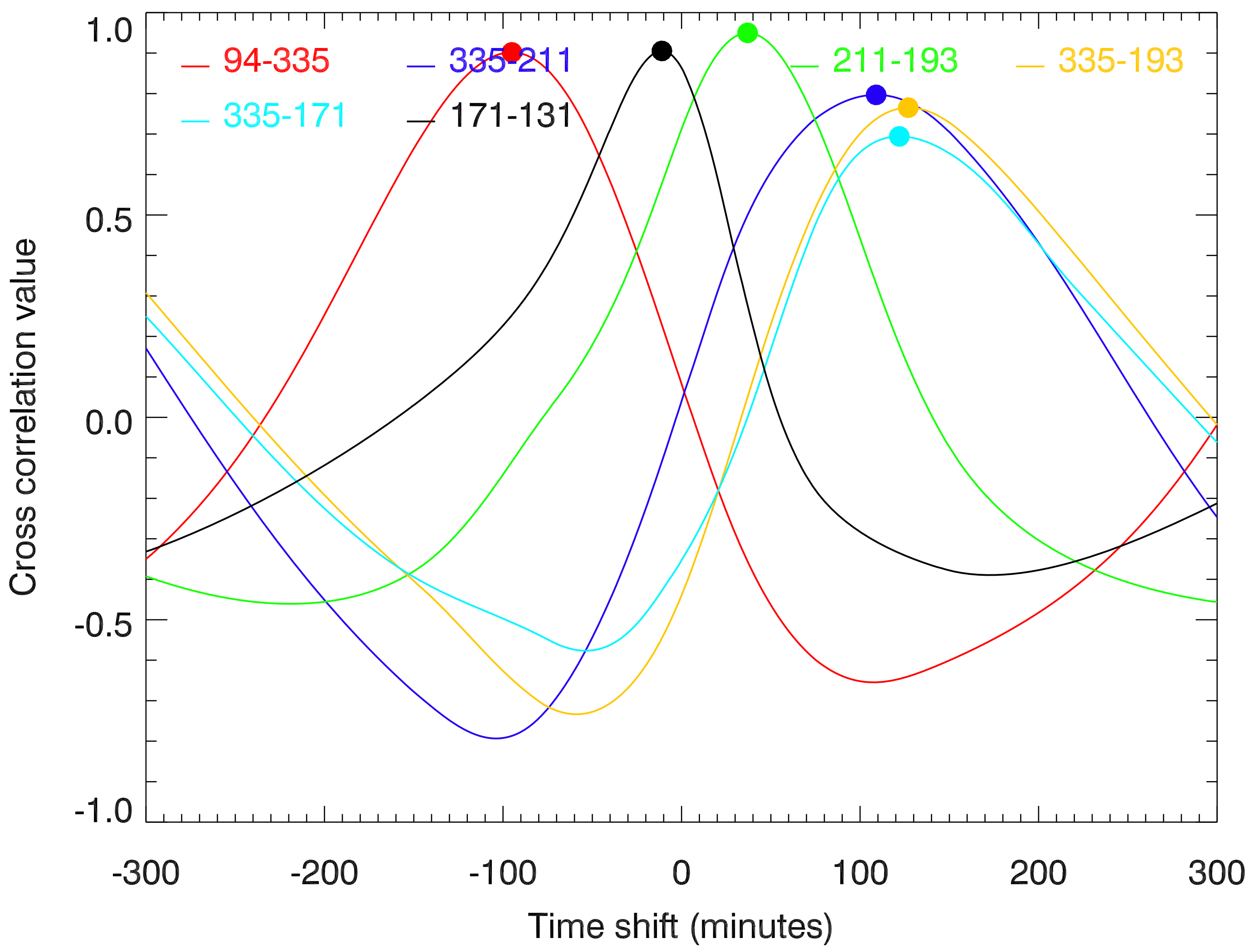
Winebarger et al. (2016)



Bradshaw and Viall (2016)



Froment et al. (2017)

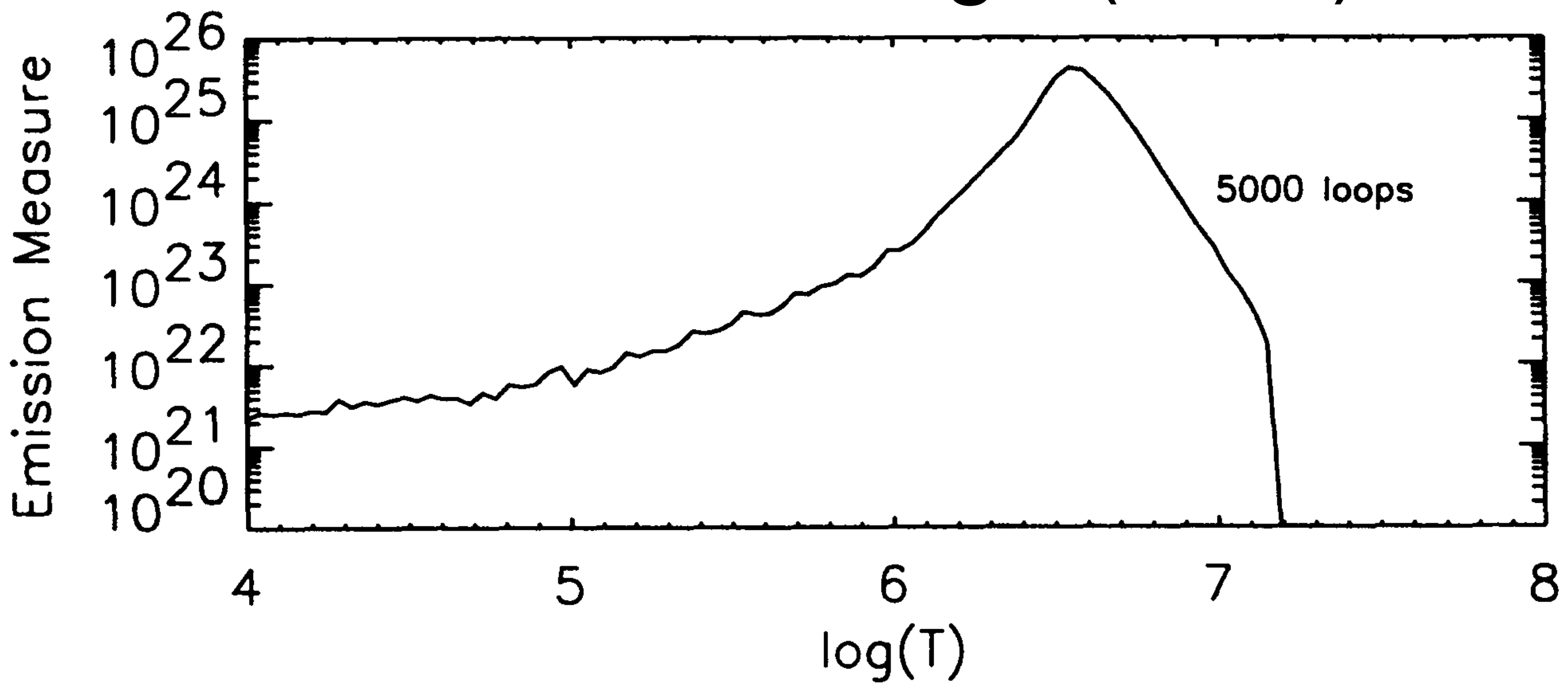


2. Diagnostics—Other

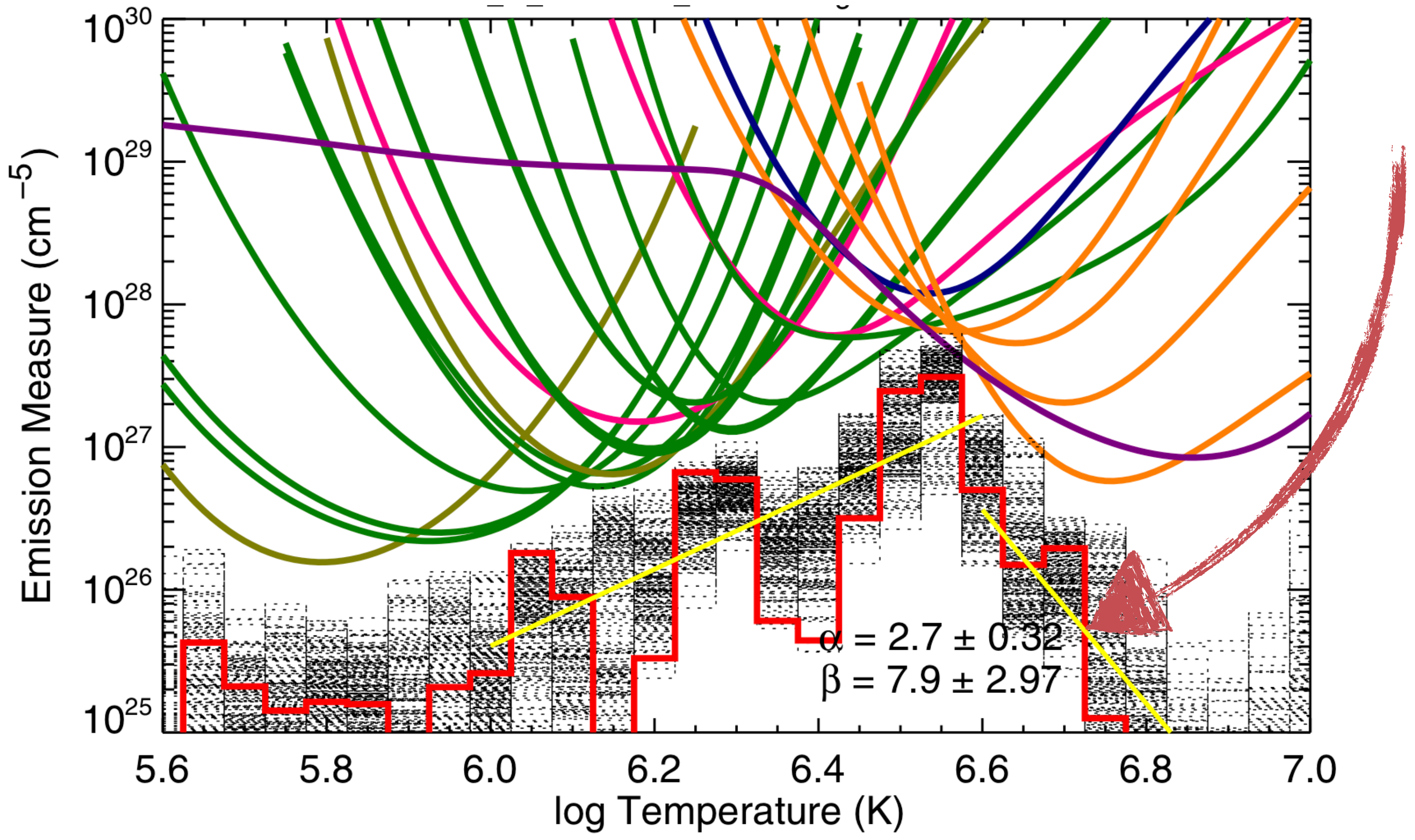
2. Diagnostics—Other

Hot Plasma

Cargill (1994)



Warren et al. (2012)

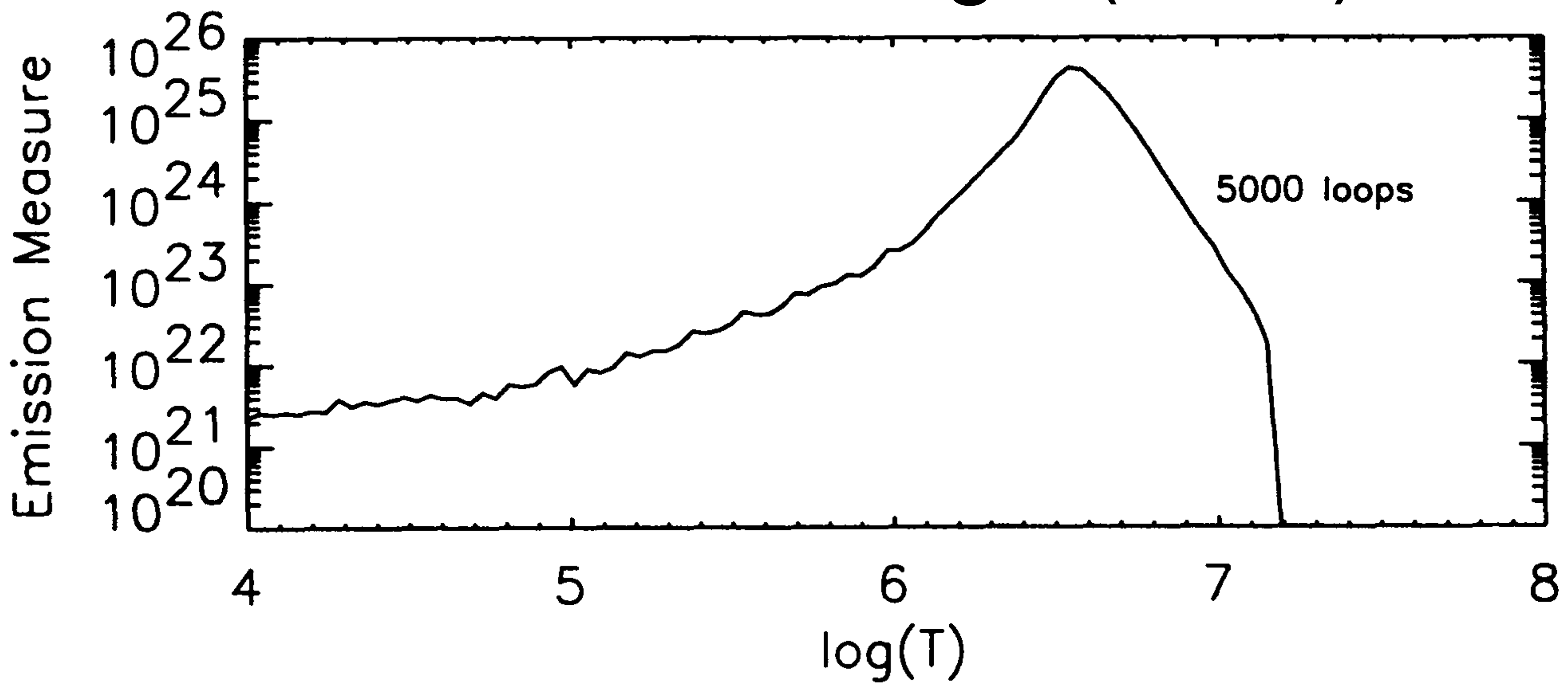


See talk by Winebarger

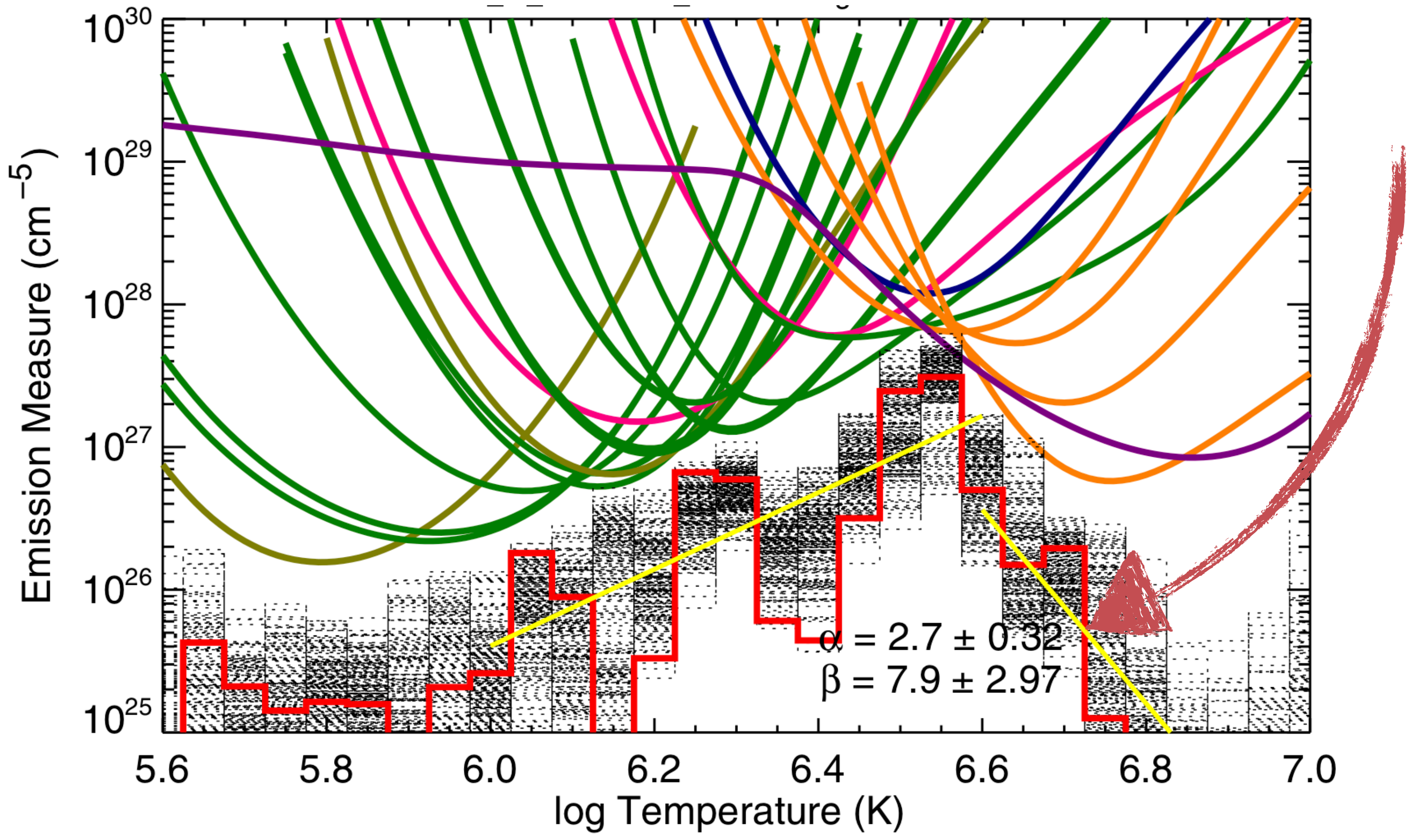
2. Diagnostics—Other

Hot Plasma

Cargill (1994)



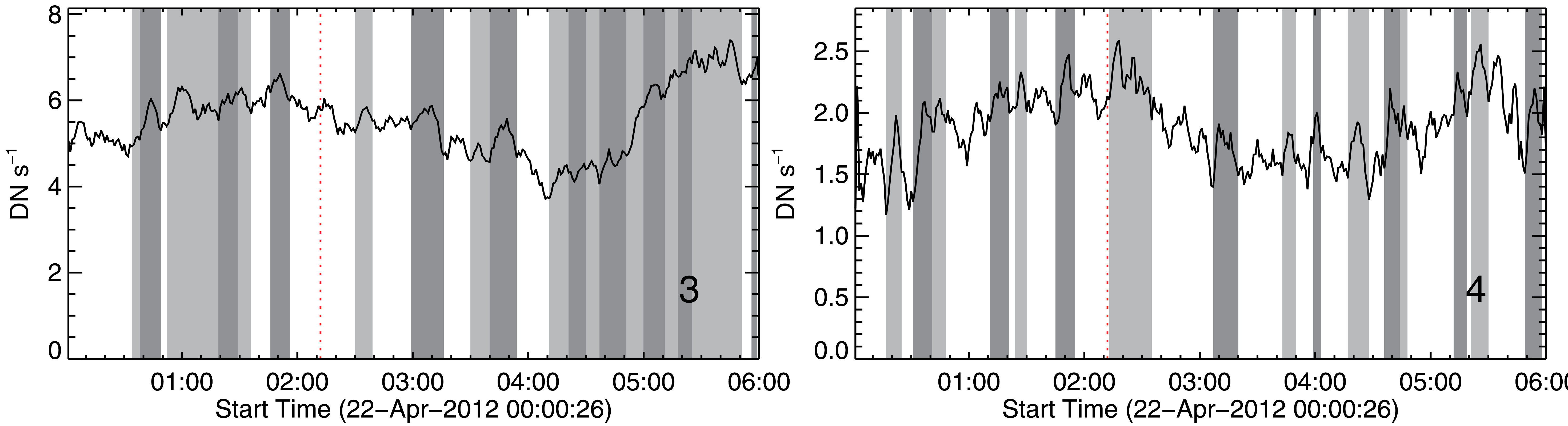
Warren et al. (2012)

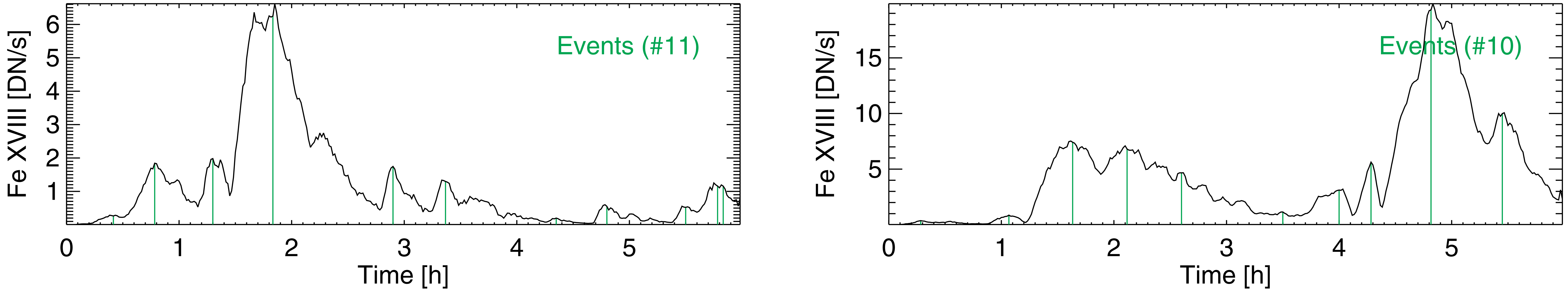


See talk by Winebarger

Event Detection

Ugarte-Urra and Warren (2014)





See poster by Plowman
See talk by Warren

3. Quantitative Comparisons of Models and Observations

3. Quantitative Comparisons of Models and Observations

$$d(x) = |M(x) - O|$$

3. Quantitative Comparisons of Models and Observations

Model diagnostic

Some distance metric

Observed diagnostic

$$d(x) = |M(x) - O|$$

Heating frequency

3. Quantitative Comparisons of Models and Observations

Model diagnostic

Some distance metric

Observed diagnostic

$$d(x) = |M(x) - O|$$

Heating frequency

$$x^\star = \min_x d(x)$$

3. Quantitative Comparisons of Models and Observations

Model diagnostic

Some distance metric

Observed diagnostic

$$d(x) = |M(x) - O|$$

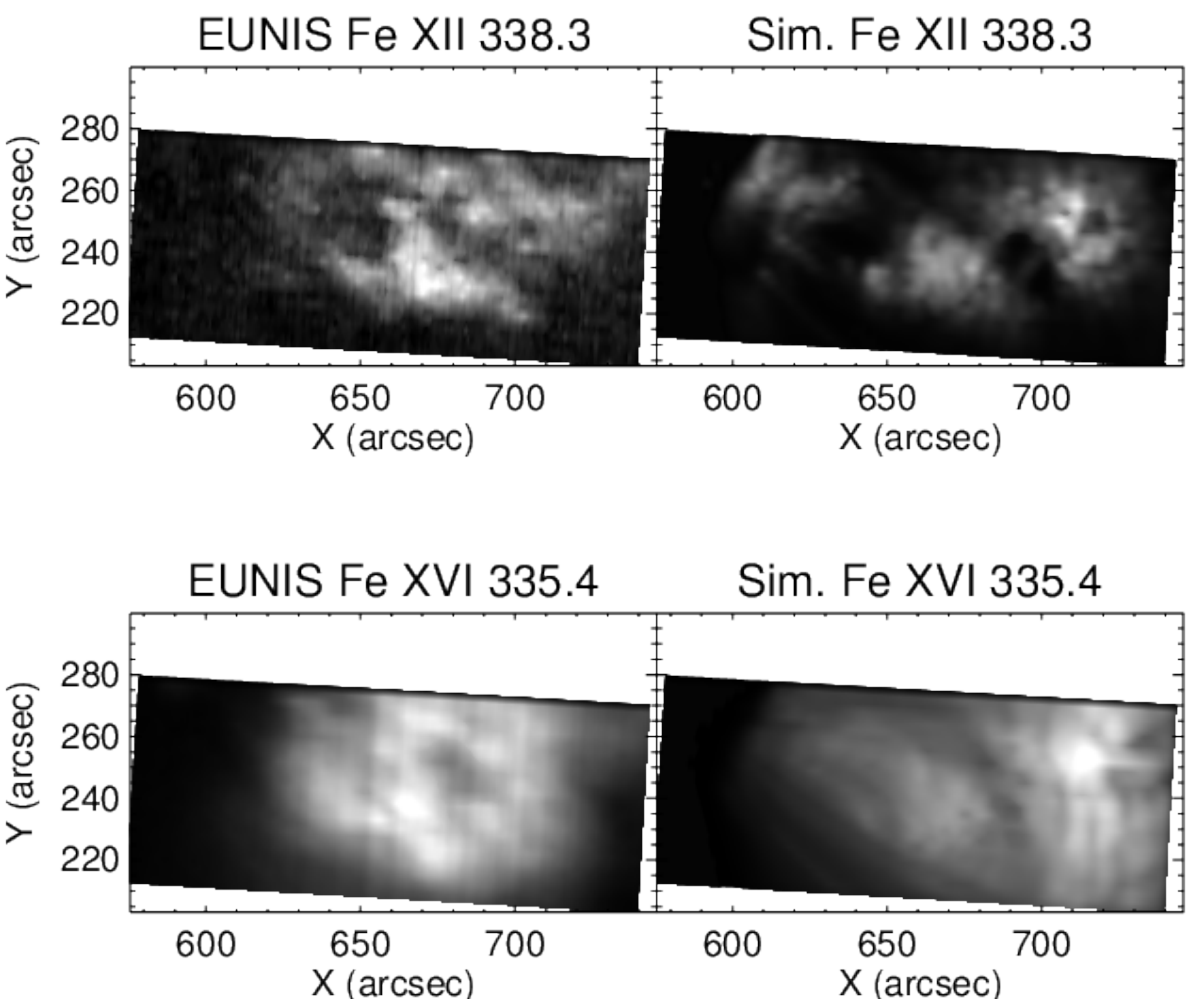
Heating frequency

$$x^{\star} = \min_{x} d(x)$$

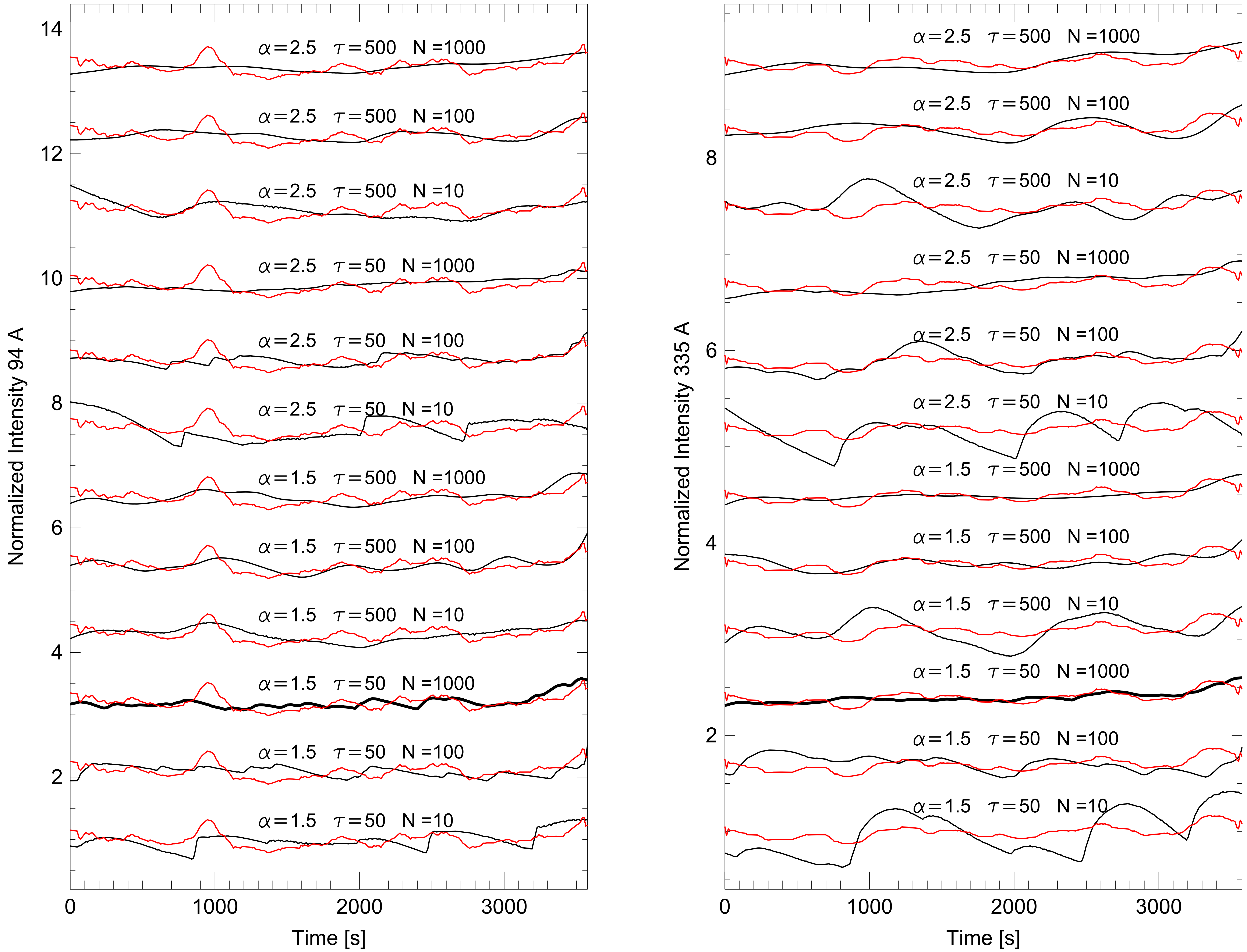
Heating frequency
most consistent
with data

3. Quantitative Comparisons of Models and Observations

Allred et al. (2018)

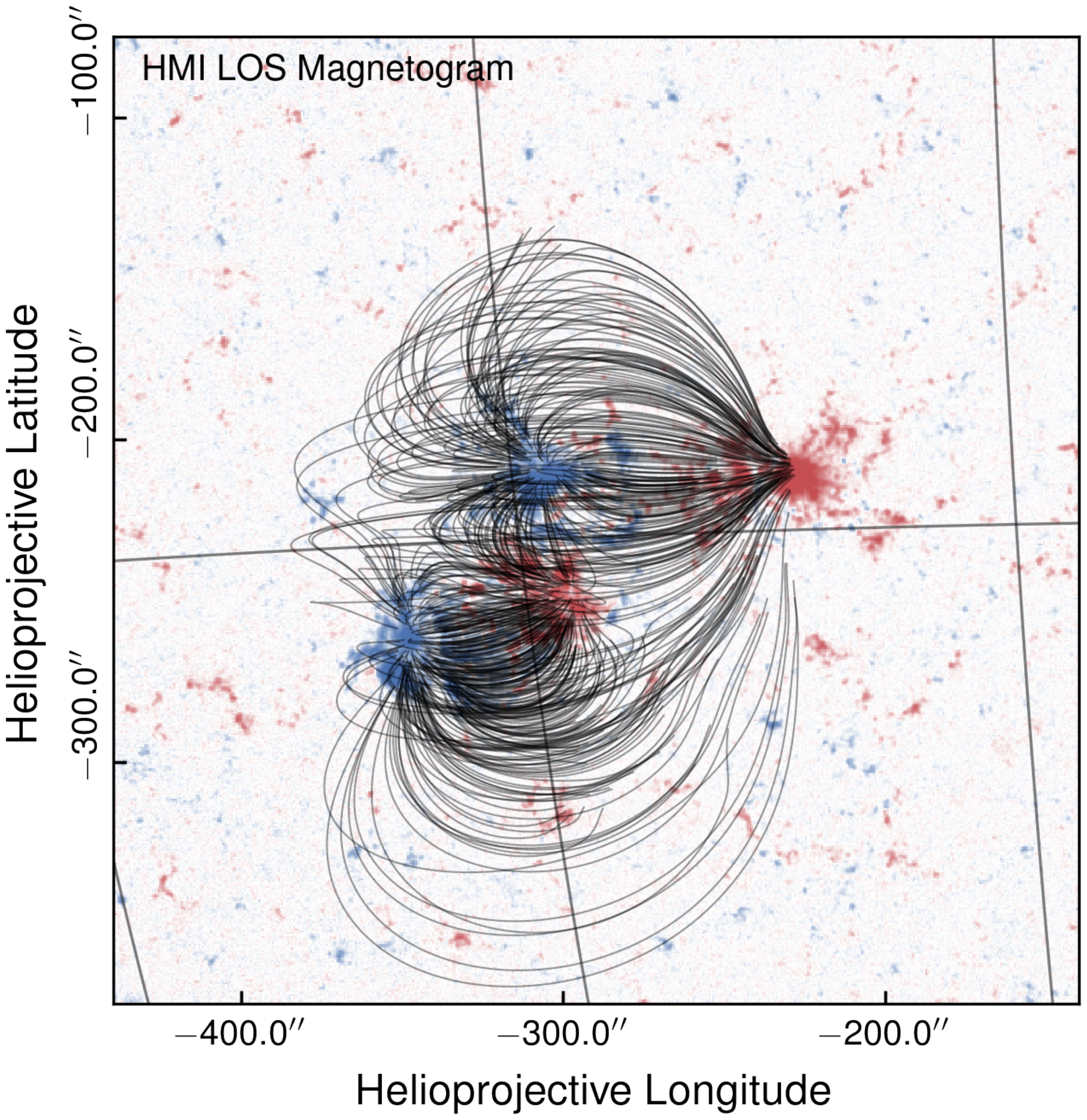


Tajfirouze et al. (2016)



See talk by Schonfeld

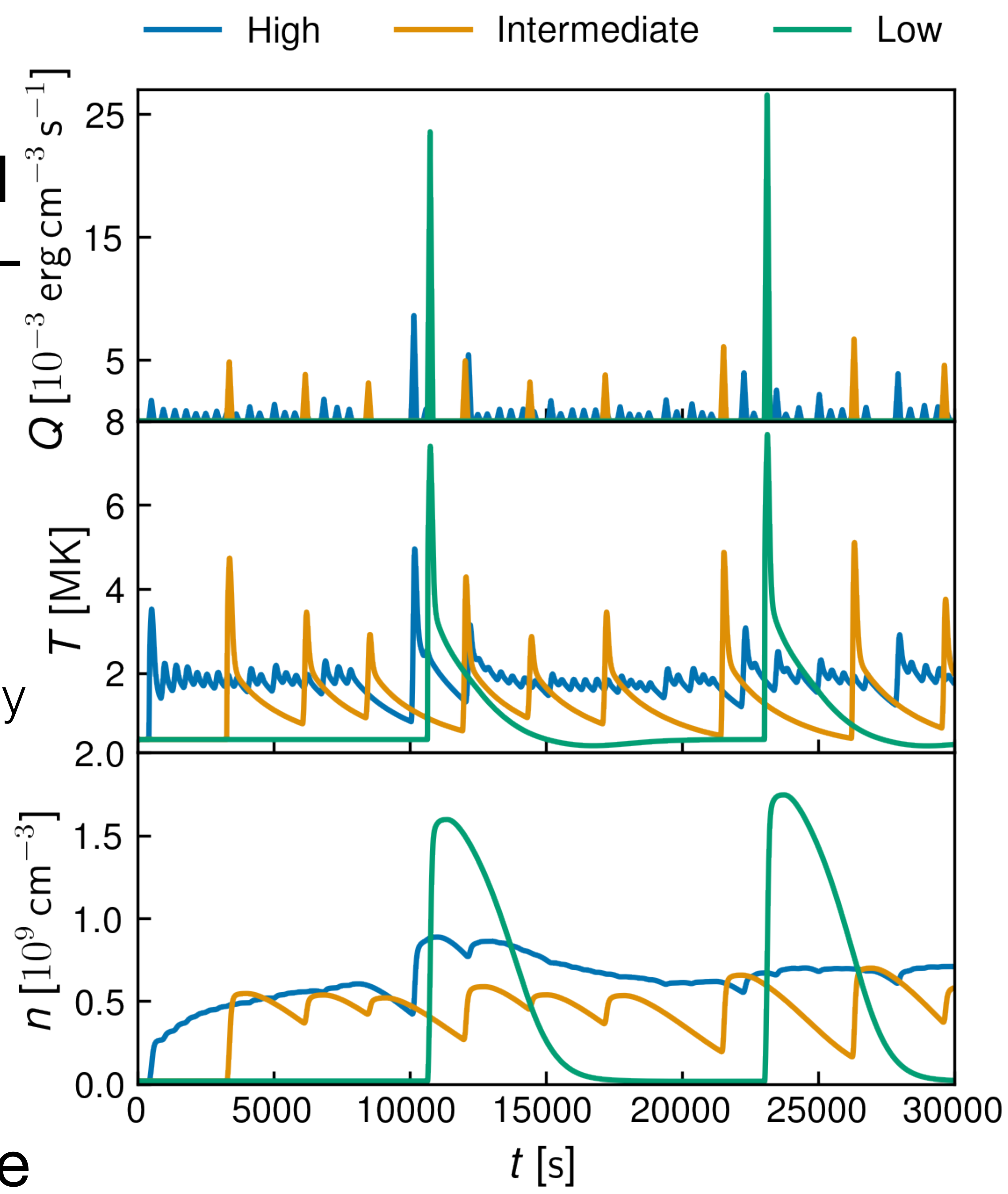
Modeling Emission from NOAA 1158



- Trace 5000 field lines from potential field extrapolation
- Simulate *T(t), n(t)* of each field line using the two-fluid EBTEL model
- Discrete events on each strand with frequency,

$$\varepsilon = \frac{\langle t_{\mathrm{wait}} \rangle}{\tau_{\mathrm{cool}}} \begin{cases} < 1, & \text{high frequency} \\ \sim 1, & \text{intermediate frequency} \\ > 1 & \text{low frequency} \end{cases}$$

- Waiting time proportional to heating rate
- Constrain total flux over whole AR to be 10^7 erg cm^{-2} s^{-1}



Warren et al. (2012), Barnes et al. (2016a)

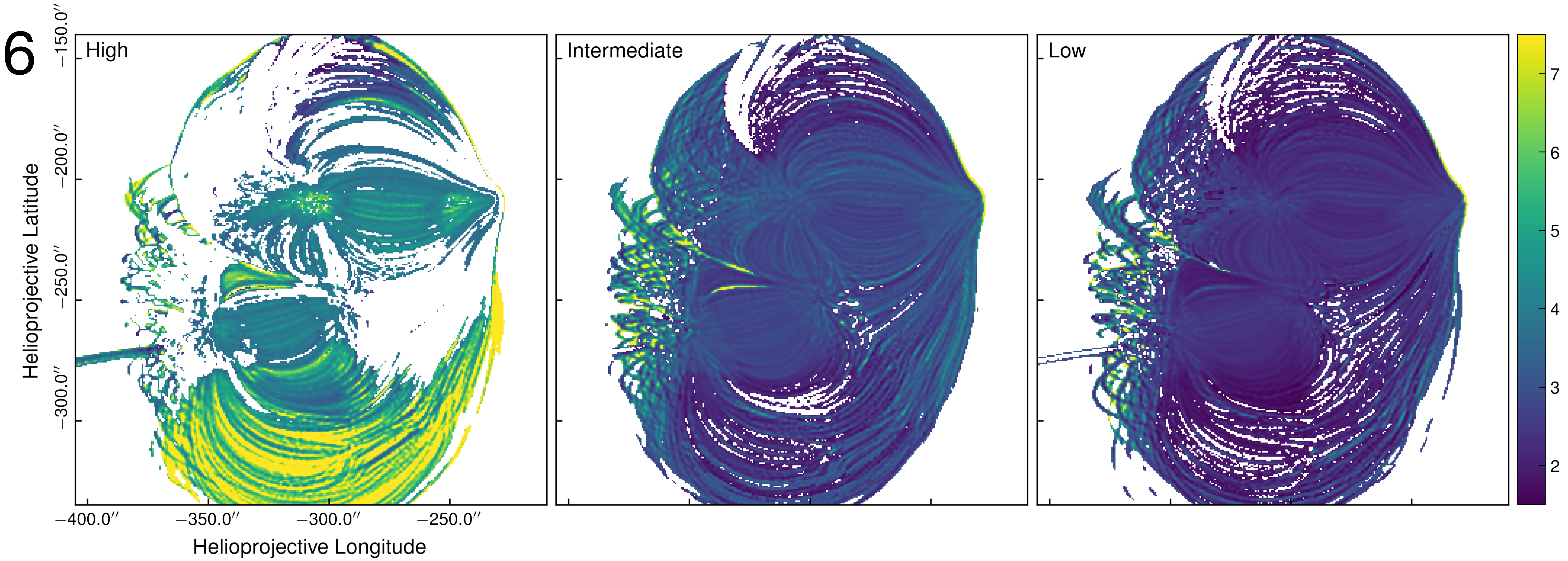
Modeling EM Slopes from NOAA 1158

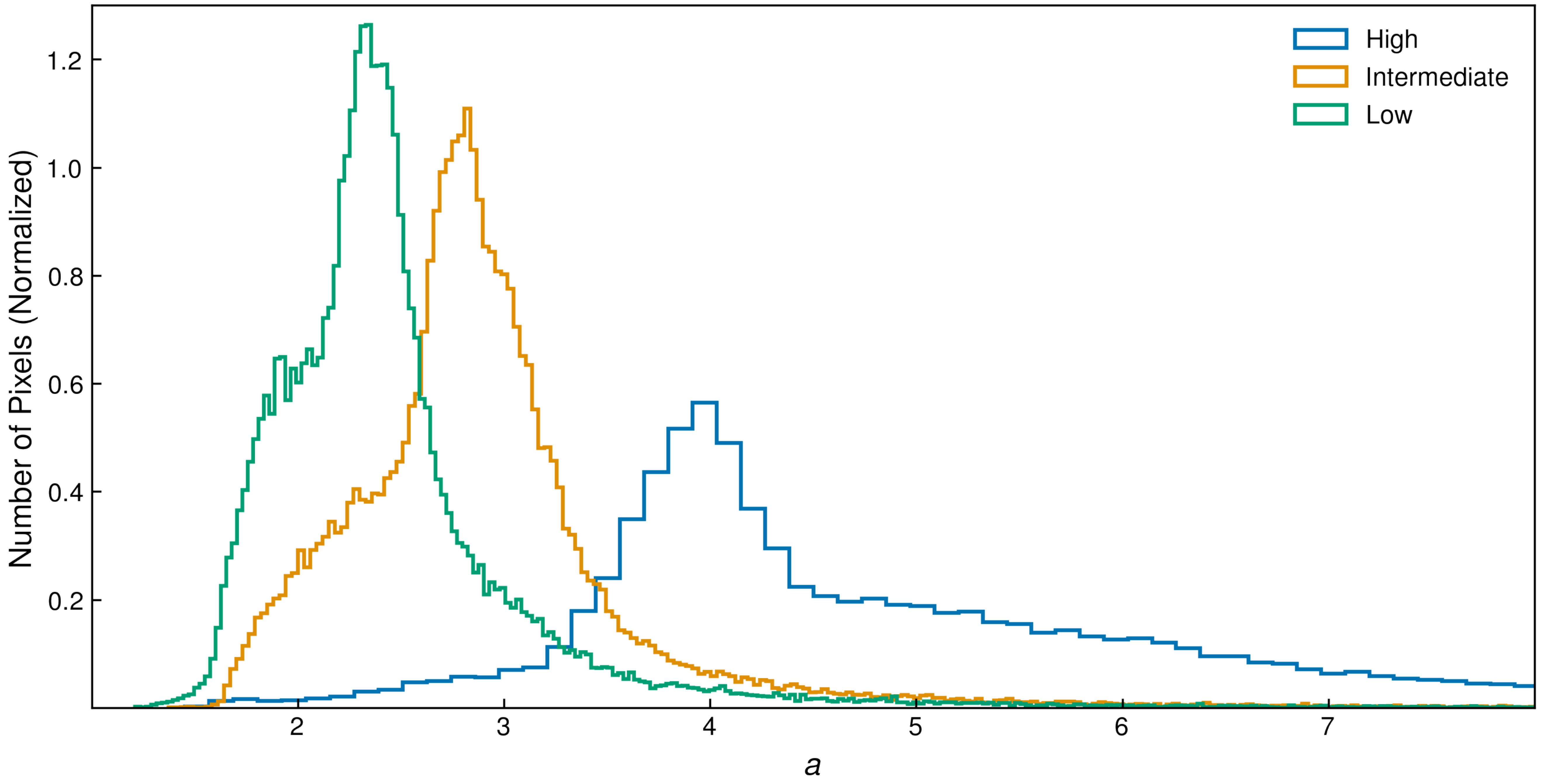
- Synthesize emission from 6 AIA EUV channels for all frequencies

- Compute EM(T) using method of Hannah and Kontar (2012)

- Bin temperature in range,
$$5.5 \leq \log T \leq 7.2$$
with bin widths
$$\Delta \log T = 0.1$$

- Fit EM slope in each pixel over temperature range,
$$8 \times 10^5\,\mathrm{K} < T < T_{\mathrm{peak}}$$



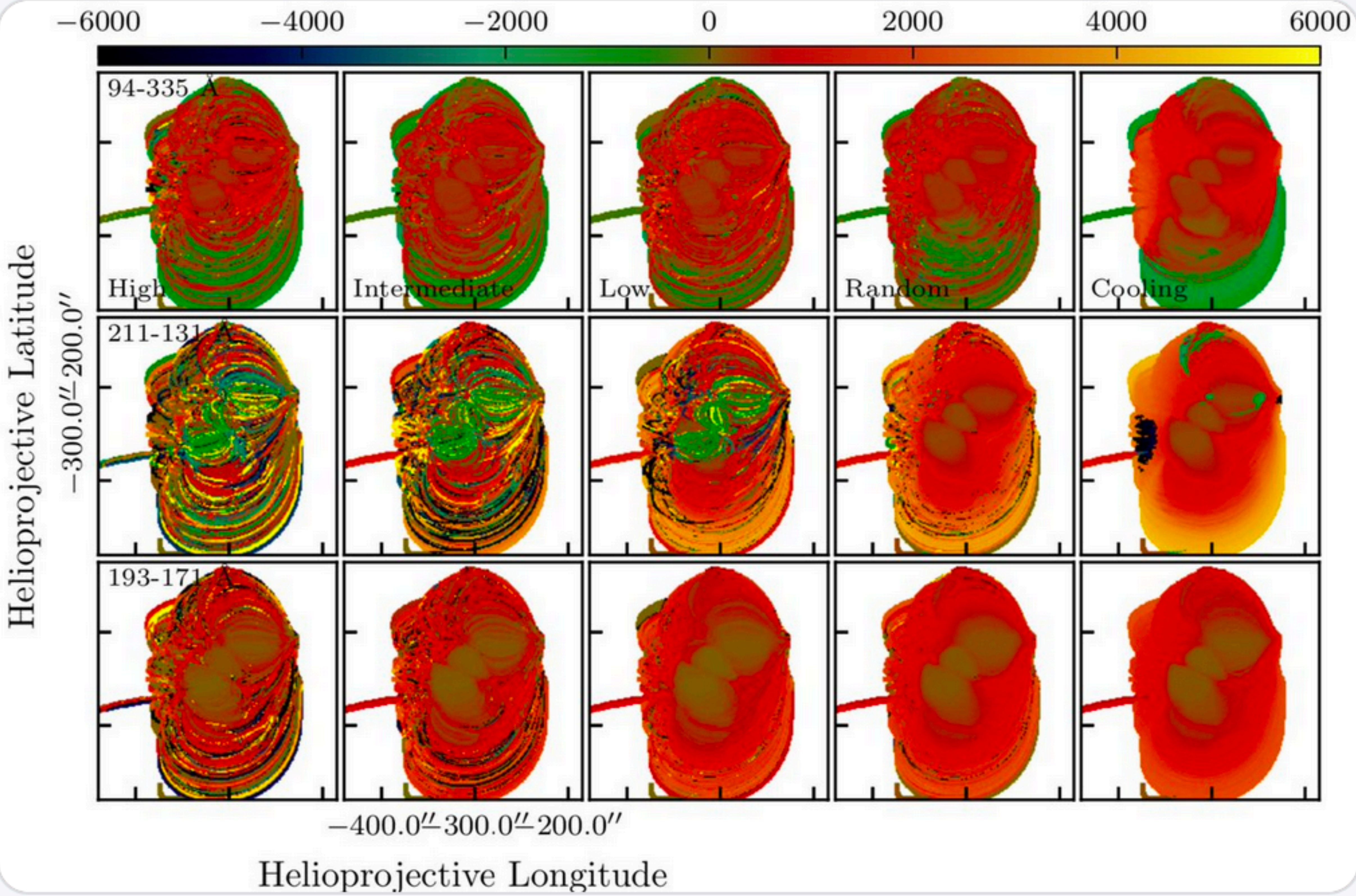


Modeling Time Lags from NOAA 1158

Modeling Diagnostics from NOAA 1158

Will Barnes @wtbarnes_ · 9h

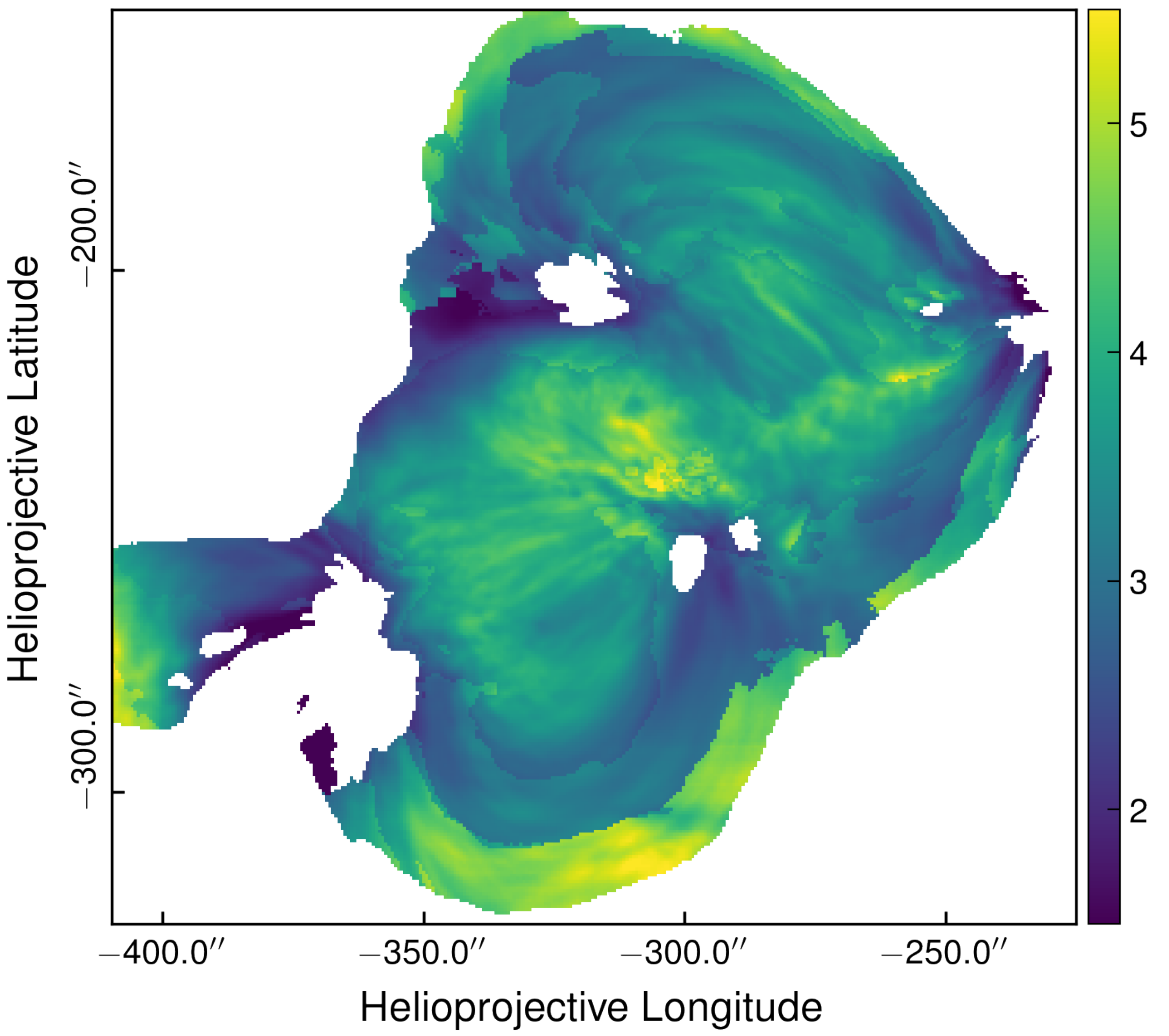
Just in time for @coronal_loops9, our paper on modeling signatures of nanoflare heating has been accepted to ApJ! The preprint is available on the arXiv now: arxiv.org/abs/1906.03350. Relevant code and notebooks available here: github.com/rice-solar-phy…

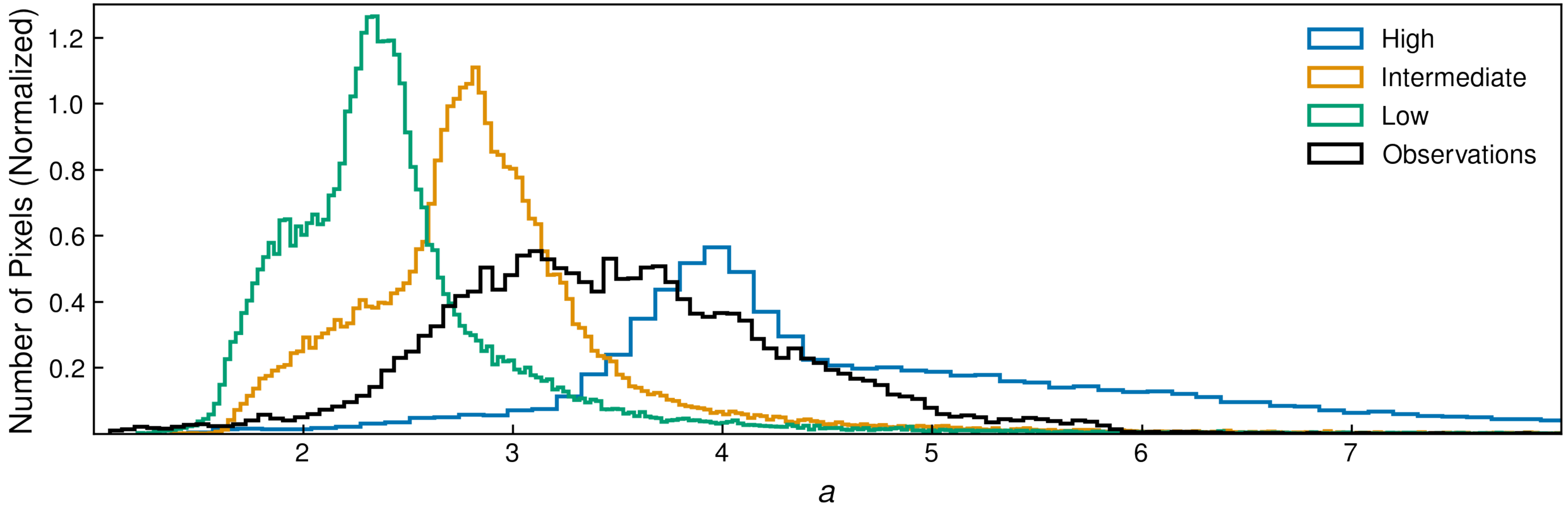


1 2 10

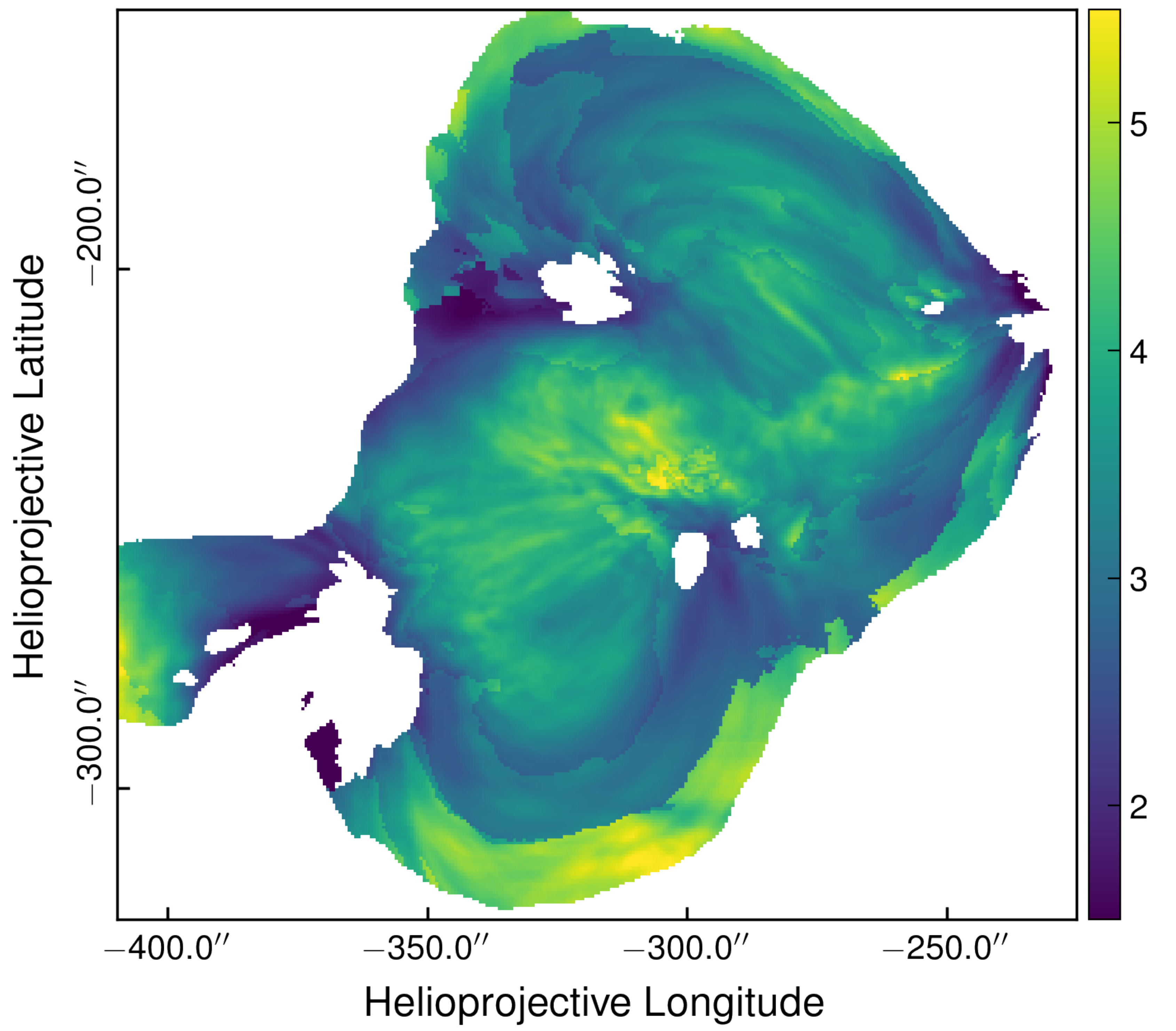
Show this thread

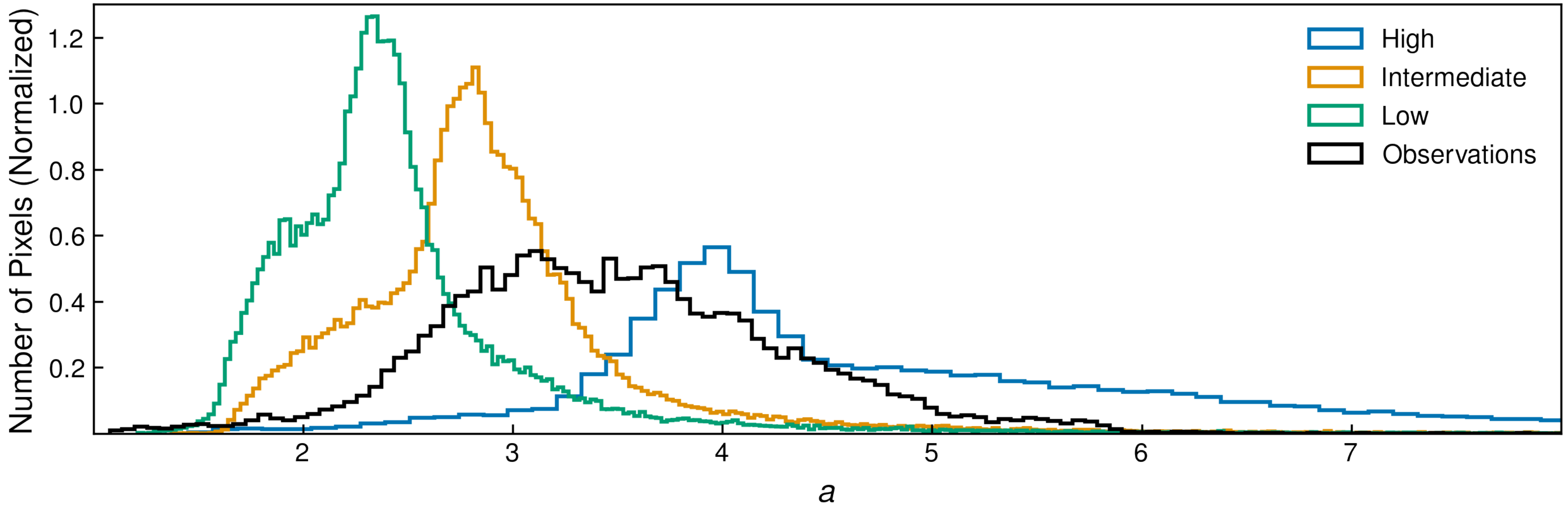
Observed Diagnostics from NOAA 1158

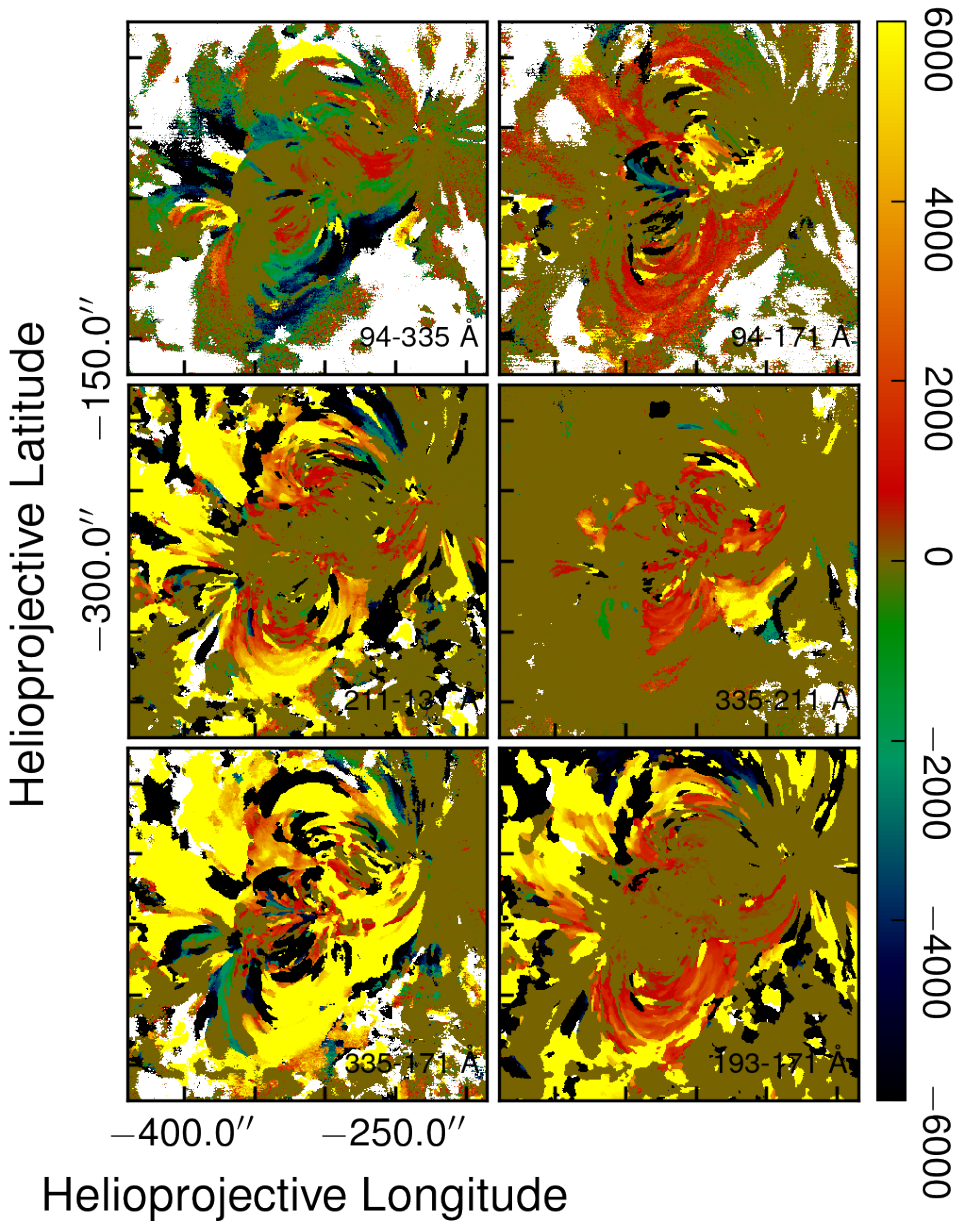




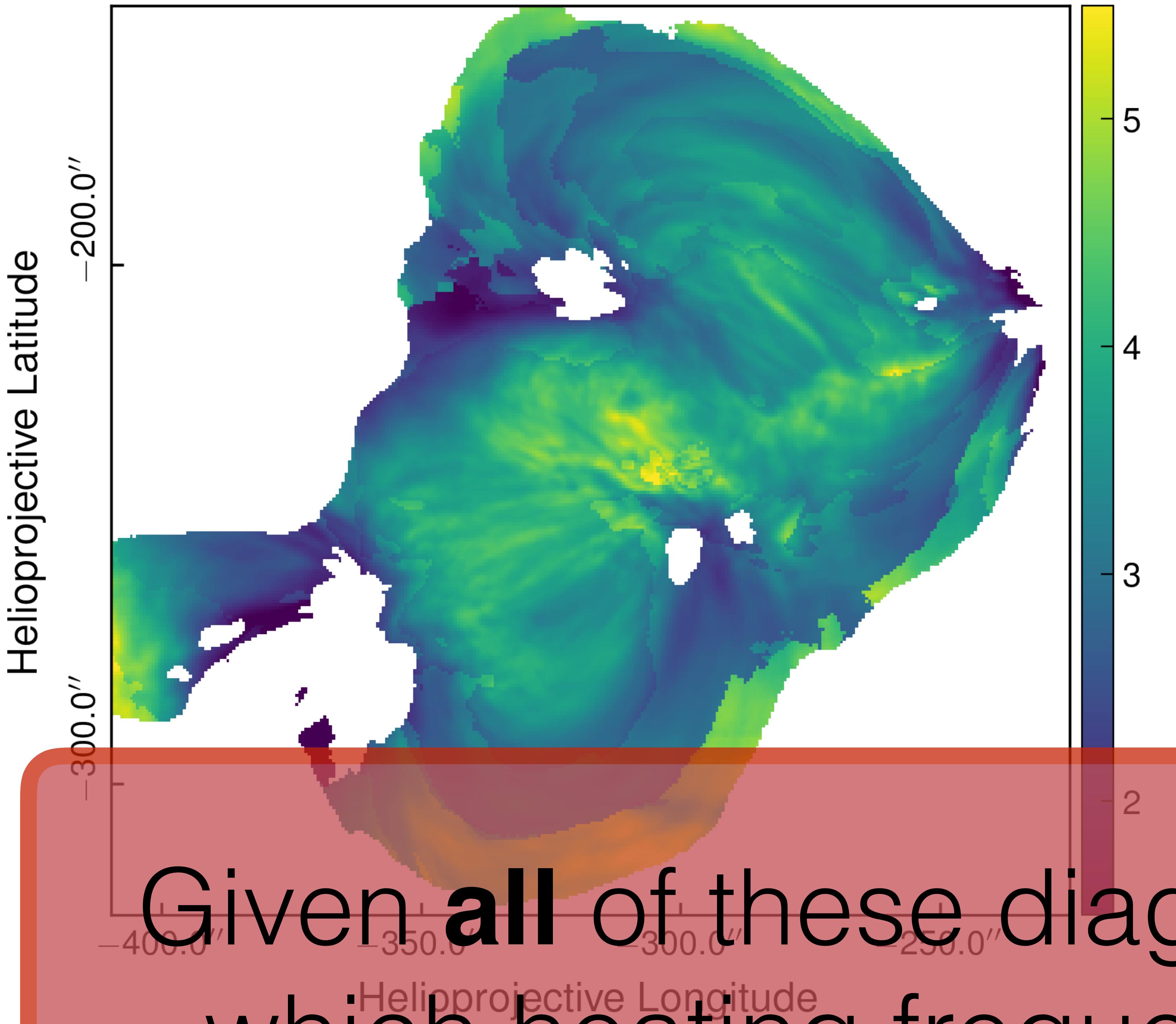
Observed Diagnostics from NOAA 1158

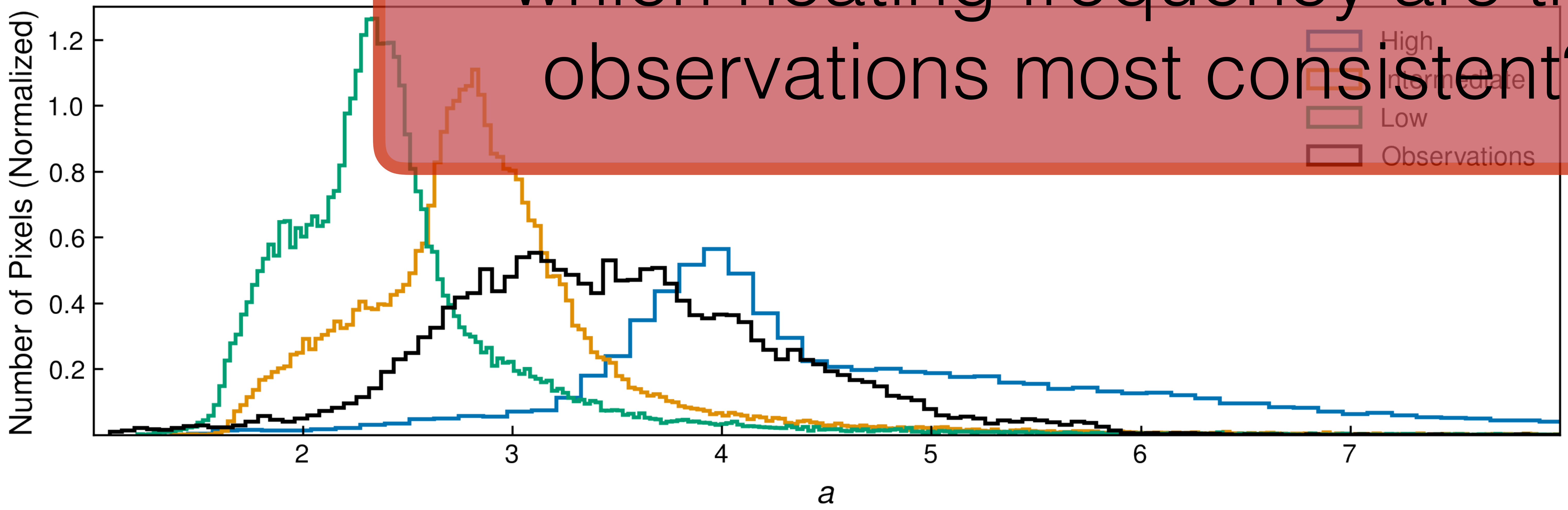


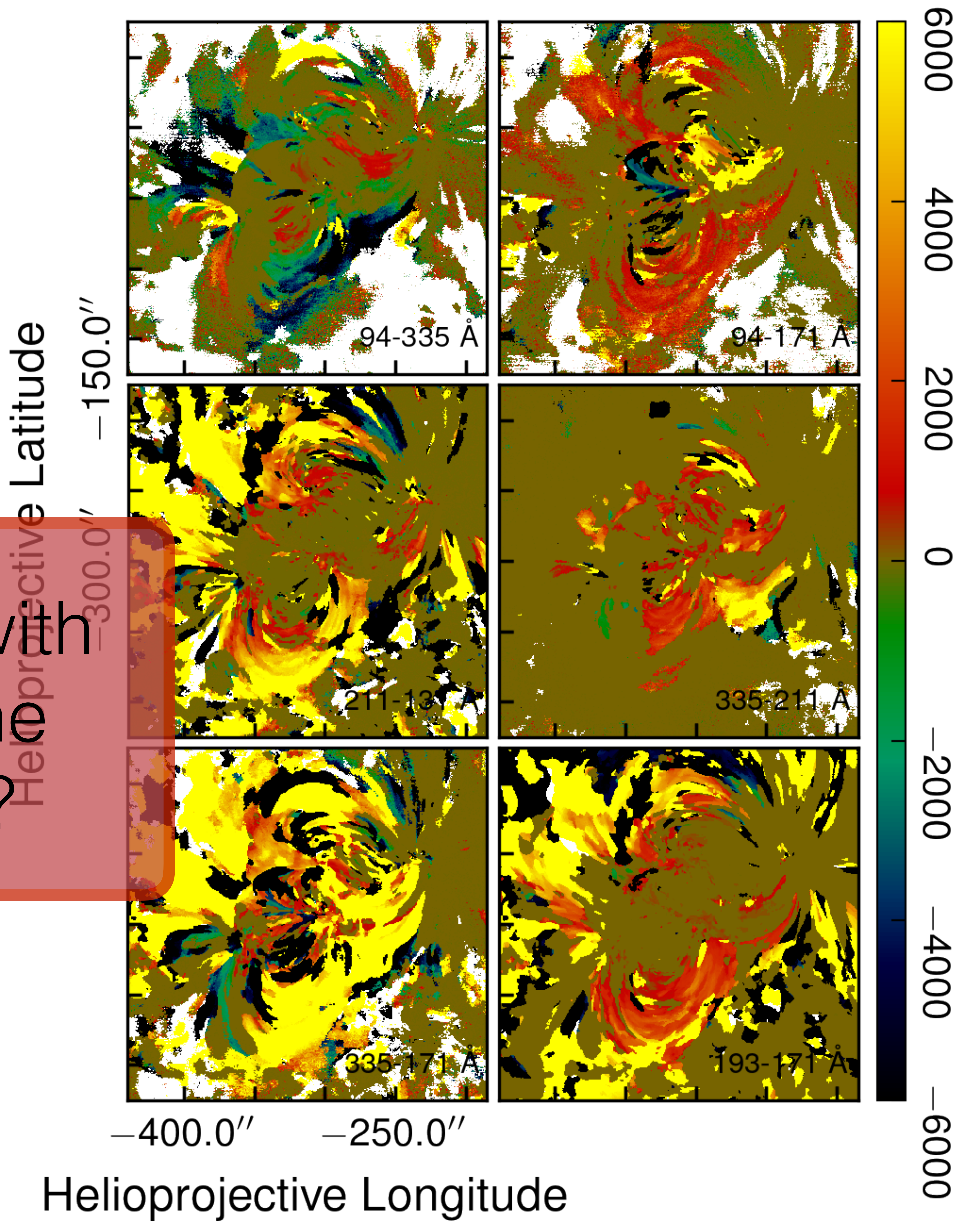




Observed Diagnostics from NOAA 1158





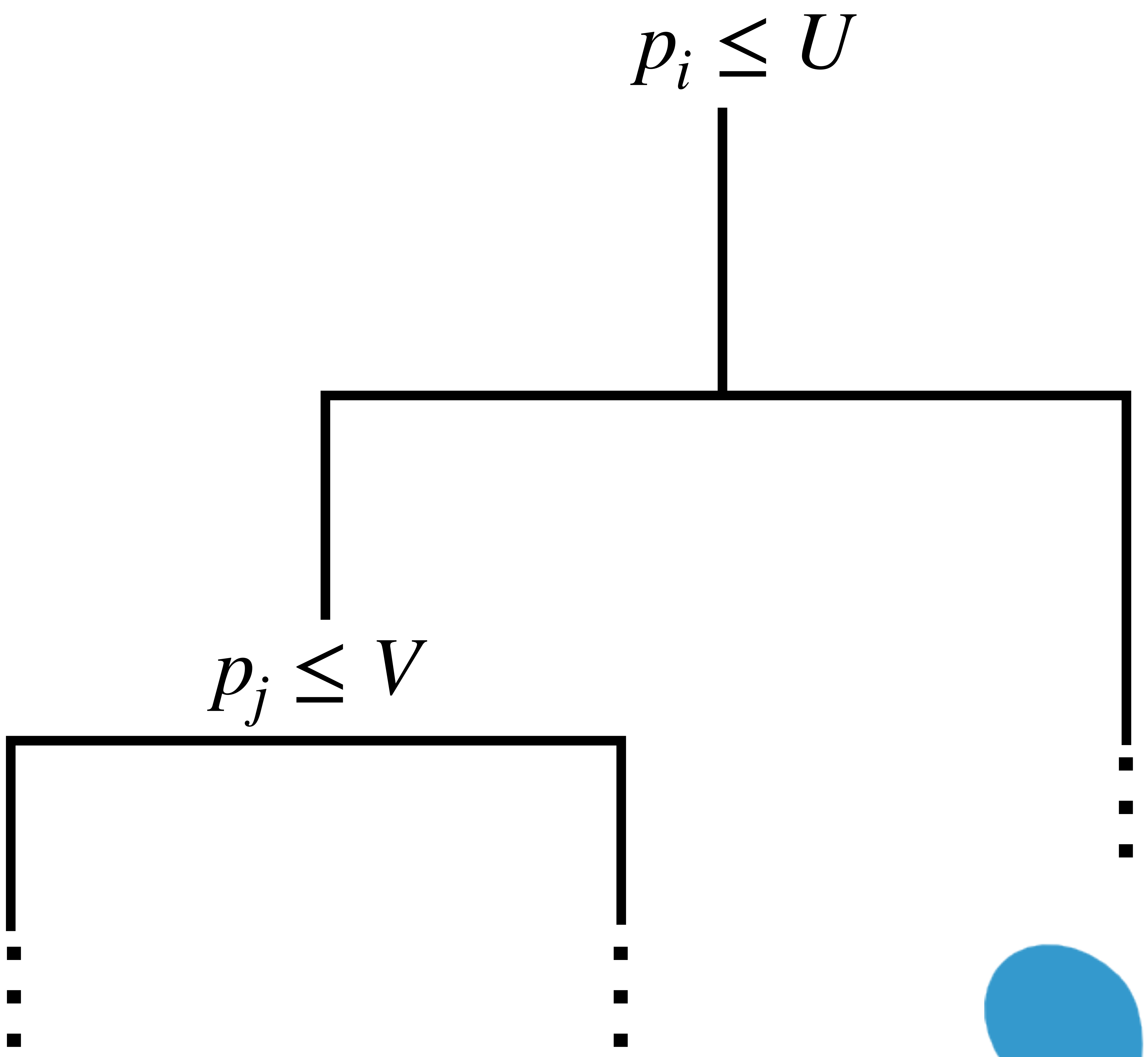


Given **all** of these diagnostics, with which heating frequency are the observations most consistent?

Comparing Models and Observations

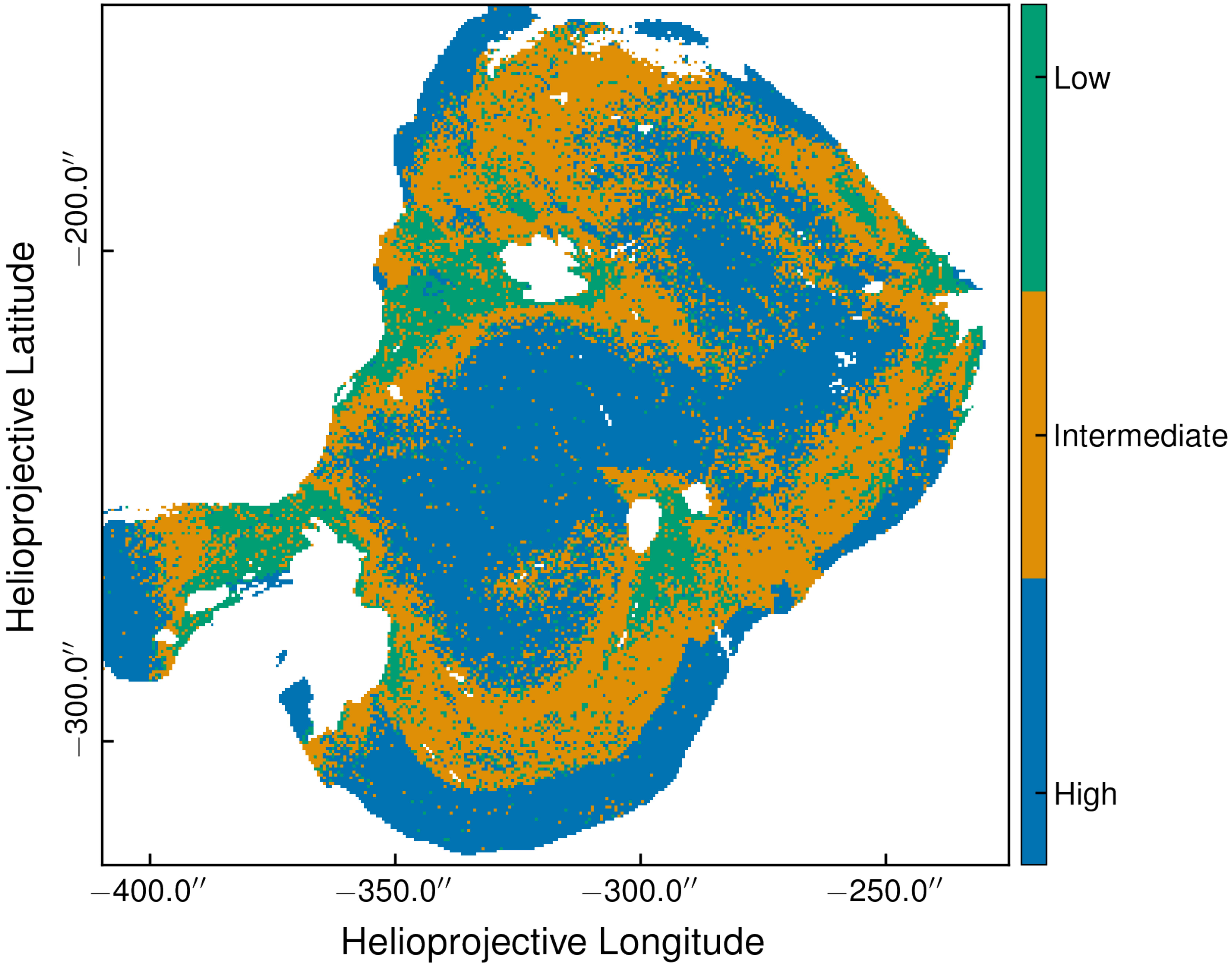
- **Question: With which heating model are the observations most consistent?**
- Classification problem—decision tree
- 31 total “features” (EM slope, 15 time lags, 15 maximum cross-correlations
- 3 discrete classes: high, intermediate, low
- Model results = training data
- Observations = unlabeled data
- Combine multiple decision trees in a *random forest*

$$\{a, \tau_{94,335}, \tau_{94,171}, \cdots, \max \mathscr{C}_{95,335}, \max \mathscr{C}_{95,171}, \cdots\}$$



Hastie et al. (2009), James et al. (2013)

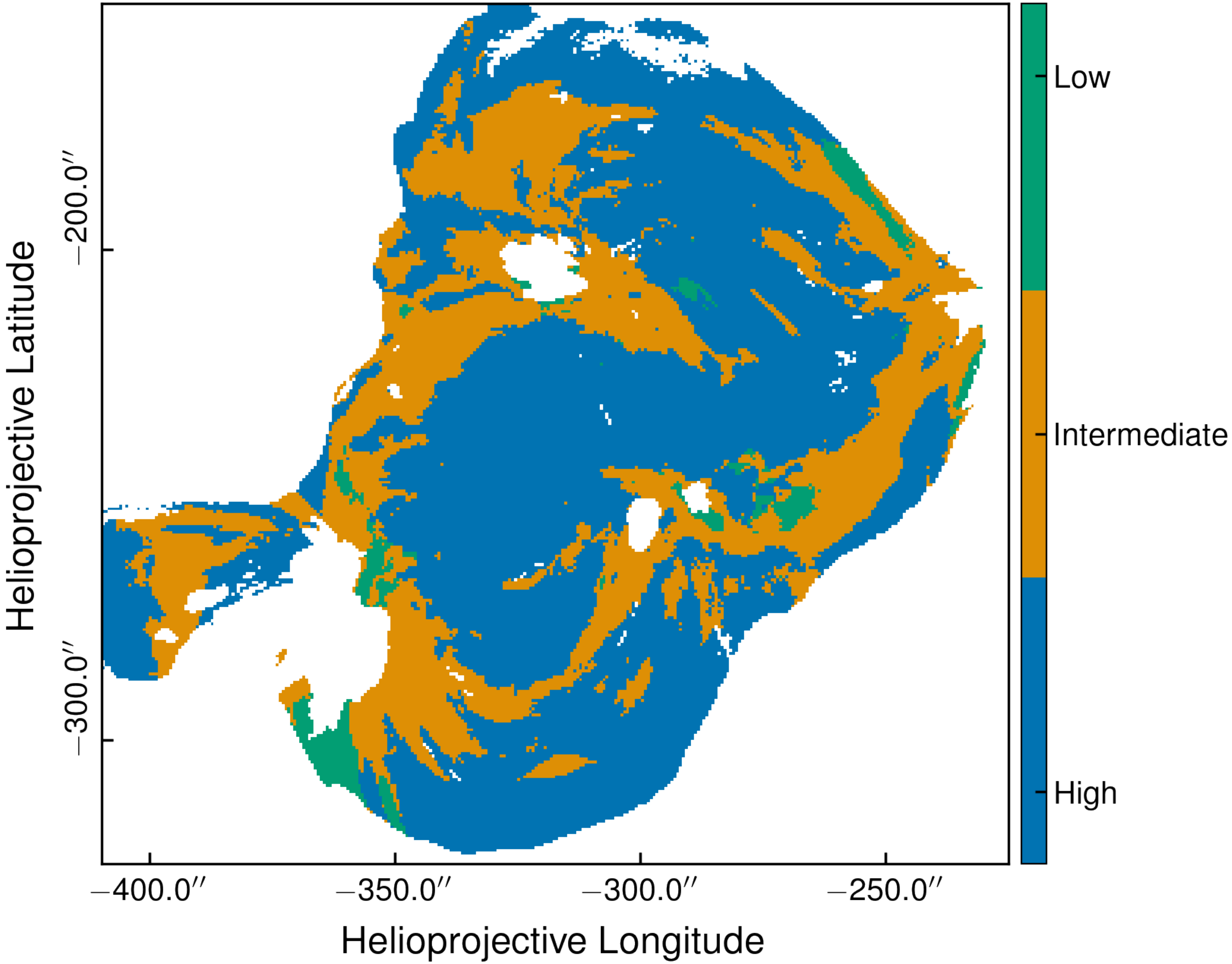
Comparing Models and Observations



EM slope only

p	1
Test Error	0.3
% High	0.48
% Intermediate	0.37
% Low	0.15

Comparing Models and Observations



EM Slope, Time Lag, Maximum cross-correlation

p	31
Test Error	0.03
% High	0.64
% Intermediate	0.33
% Low	0.03

Looking Forward

- Constraints on heating properties with **multiple** diagnostics
- **Quantitative** comparisons between models and data
- Understand distribution of frequencies **across** active region
- How does the distribution of frequencies over **multiple** active regions?
- How does the distribution of frequencies vary with **age**?

LOCKHEED MARTIN

- Publications
  - Barnes, W. T., Bradshaw, S. J., Viall, N. M., 2019, “Understanding Heating in Active Region Cores through Machine Learning I. Numerical Modeling and Predicted Observables”, ApJ (accepted), https://arxiv.org/abs/1906.03350
  - Barnes, W. T., Bradshaw, S. J., Viall, N. M., 2019, “Understanding Heating in Active Region Cores through Machine Learning II. Observations”, in prep
- Acknowledgment
  - SOC
  - LOC
  - Steve Bradshaw
  - Nicki Viall

SOLAR AND ASTROPHYSICS LAB

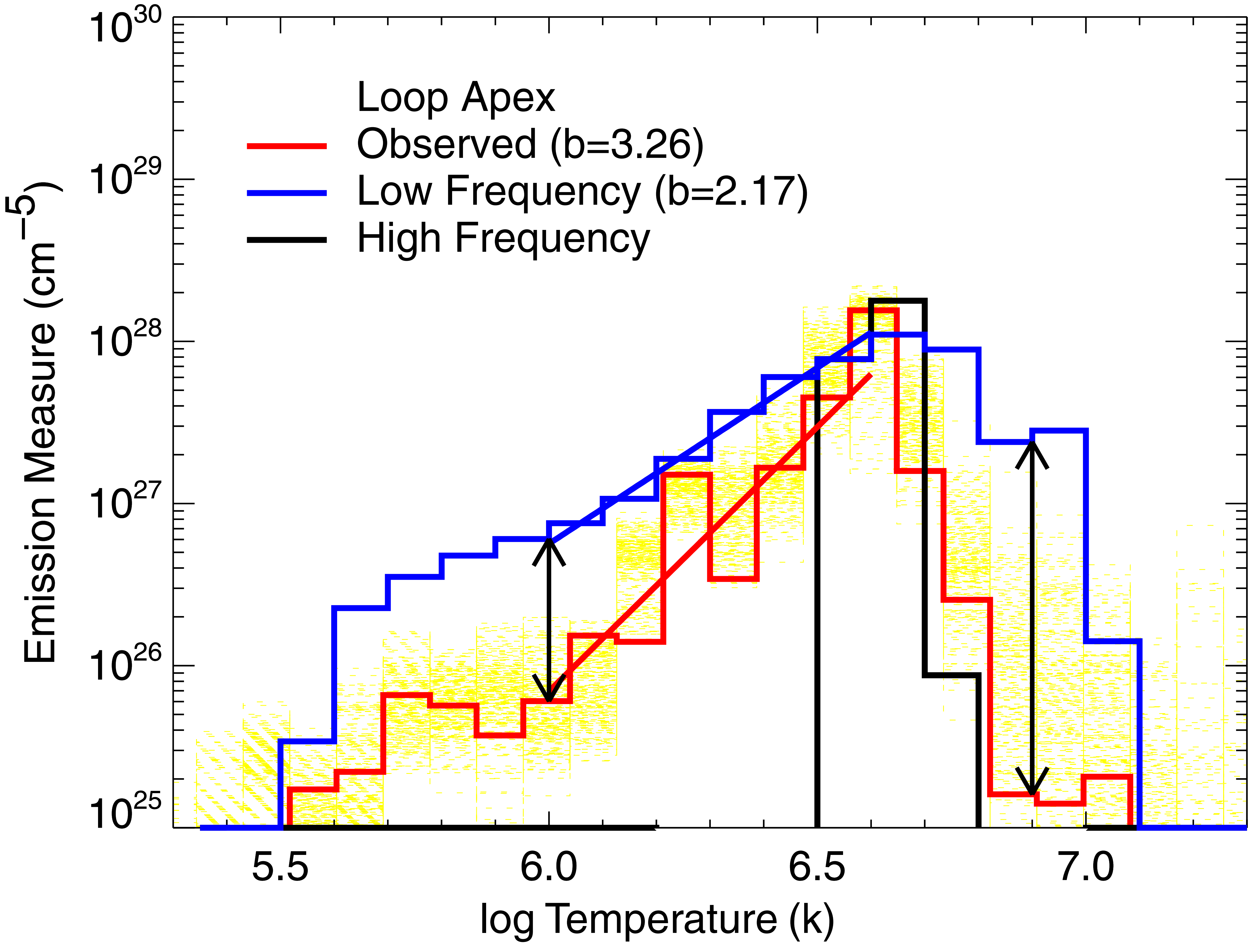
sunpy powered

Supplementary Slides

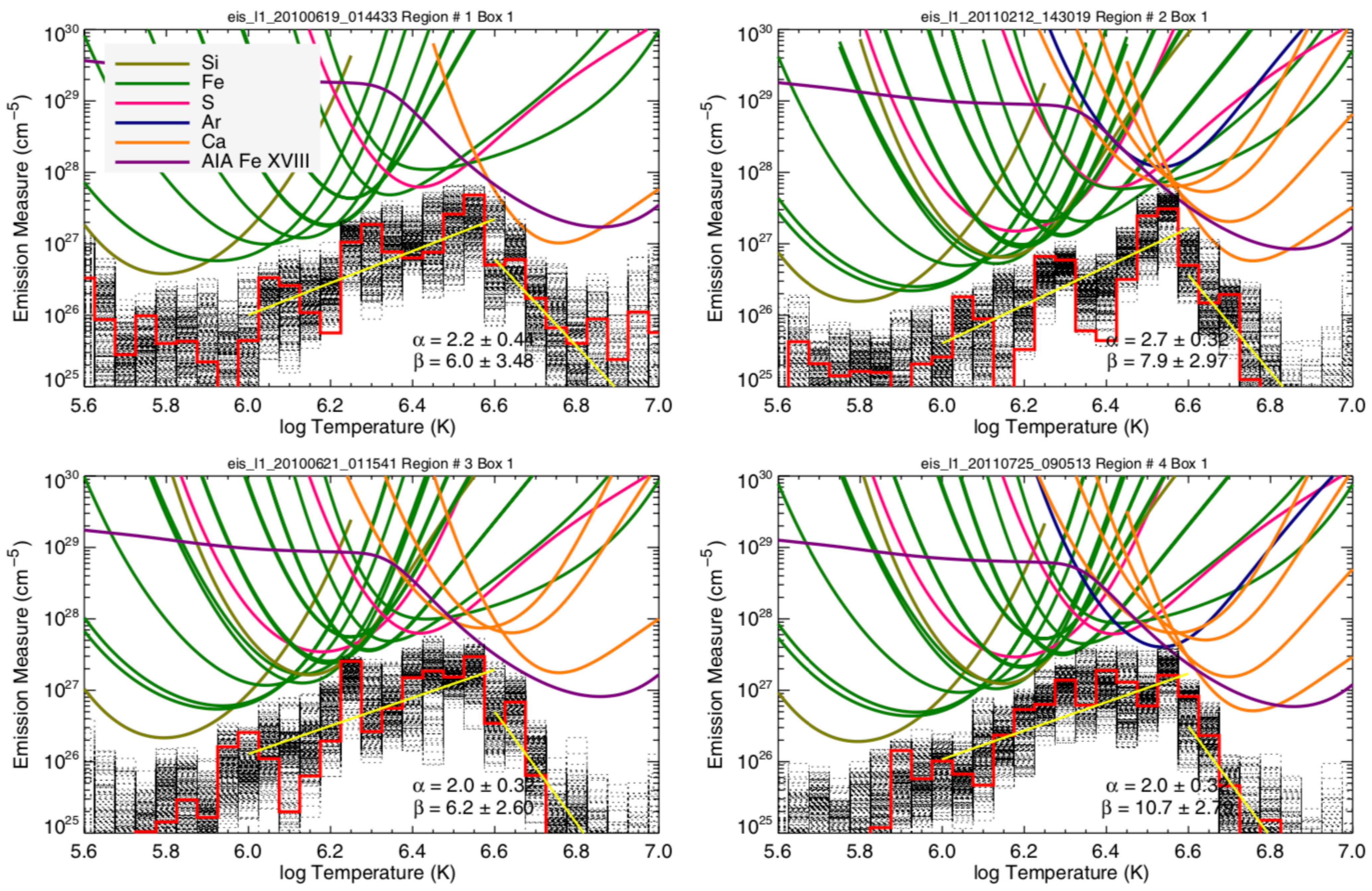
2. Diagnostics of Heating Frequency

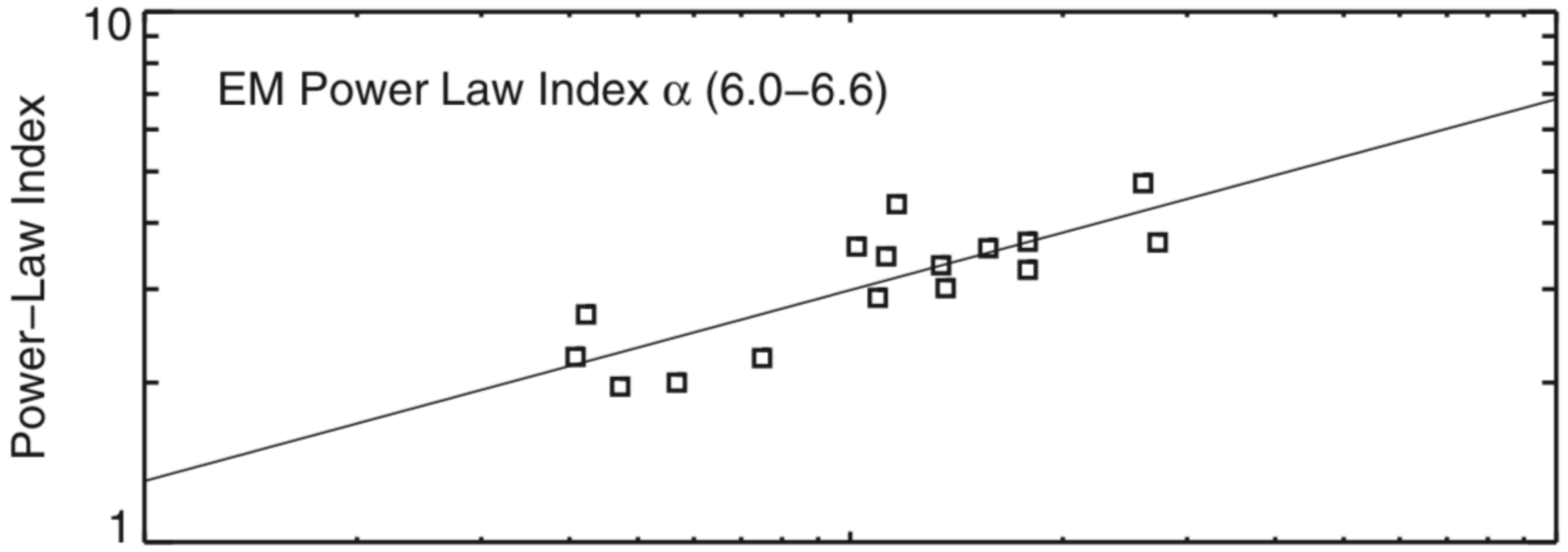
Reduced representation of data that preserves signatures of heating frequency

Warren et al. (2011)



Warren et al. (2012)

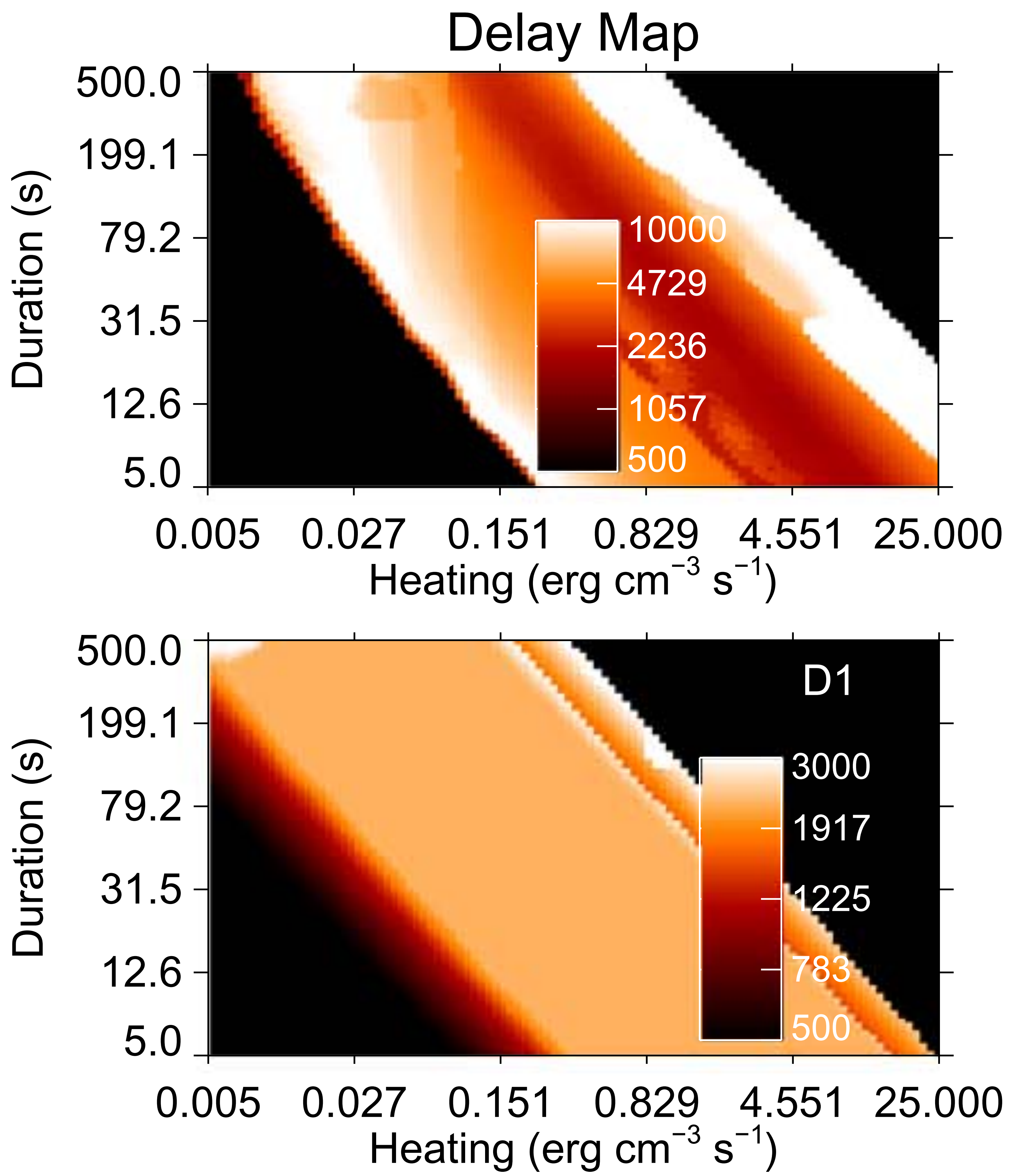




3. Quantitative Comparisons of Models and Observations

3. Quantitative Comparisons of Models and Observations

Marsh et al. (2018)



- Single loop modeled with EBTEL
- Varied duration, magnitude, heating rate
- Computed likelihood between modeled and observed NuSTAR and FOXSI-2 HXR spectra

Computing Intensities

$$P_c(s,t) = \sum_{\{ij\}} P_{ij} R_c(\lambda_{ij})$$

$$P_{ij} = \frac{n_h}{n_e} \mathrm{Ab}(X) f_{X,k}(T_e) N_j(n_e, T_e) A_{ij} \Delta E_{ij} n_e$$

Mason and Monsignori Fossi (1994), Bradshaw and Raymond (2013), Del Zanna and Mason (2018)

Computing Intensities

$$P_c(s,t) = \sum_{\{ij\}} P_{ij} R_c(\lambda_{ij})$$

$$P_{ij} = \frac{n_h}{n_e} \mathrm{Ab}(X) f_{X,k}(T_e) N_j(n_e, T_e) A_{ij} \Delta E_{ij} n_e$$

EBTEL Model

Klimchuk et al. (2008)
Cargill et al. (2012a,b)
Barnes et al. (2016a)

Mason and Monsignori Fossi (1994), Bradshaw and Raymond (2013), Del Zanna and Mason (2018)

Computing Intensities

$$P_c(s,t) = \sum_{\{ij\}} P_{ij} R_c(\lambda_{ij})$$

Nonequilibrium Ionization

Bradshaw (2009)
Macneice et al. (1984)

$$P_{ij} = \frac{n_h}{n_e} \mathrm{Ab}(X) f_{X,k}(T_e) N_j(n_e, T_e) A_{ij} \Delta E_{ij} n_e$$

CHIANTI

Dere et al. (1997)
Young et al. (2016)

EBTEL Model

Klimchuk et al. (2008)
Cargill et al. (2012a,b)
Barnes et al. (2016a)

Mason and Monsignori Fossi (1994), Bradshaw and Raymond (2013), Del Zanna and Mason (2018)

Computing Intensities

Instrument

Boerner et al. (2012)
Bradshaw and Klimchuk (2011)

$$P_c(s,t) = \sum_{\{ij\}} P_{ij} R_c(\lambda_{ij})$$

Nonequilibrium Ionization

Bradshaw (2009)
Macneice et al. (1984)

$$P_{ij} = \frac{n_h}{n_e} \mathrm{Ab}(X) f_{X,k}(T_e) N_j(n_e, T_e) A_{ij} \Delta E_{ij} n_e$$

CHIANTI

Dere et al. (1997)
Young et al. (2016)

EBTEL Model

Klimchuk et al. (2008)
Cargill et al. (2012a,b)
Barnes et al. (2016a)

Mason and Monsignori Fossi (1994), Bradshaw and Raymond (2013), Del Zanna and Mason (2018)

Computing Intensities

Instrument

Boerner et al. (2012)
Bradshaw and Klimchuk (2011)

$$P_c(s,t) = \sum_{\{ij\}} P_{ij} R_c(\lambda_{ij})$$

Nonequilibrium Ionization

Bradshaw (2009)
Macneice et al. (1984)

$$P_{ij} = \frac{n_h}{n_e}\mathrm{Ab}(X) f_{X,k}(T_e) N_j(n_e, T_e) A_{ij} \Delta E_{ij} n_e$$

CHIANTI

Dere et al. (1997)
Young et al. (2016)

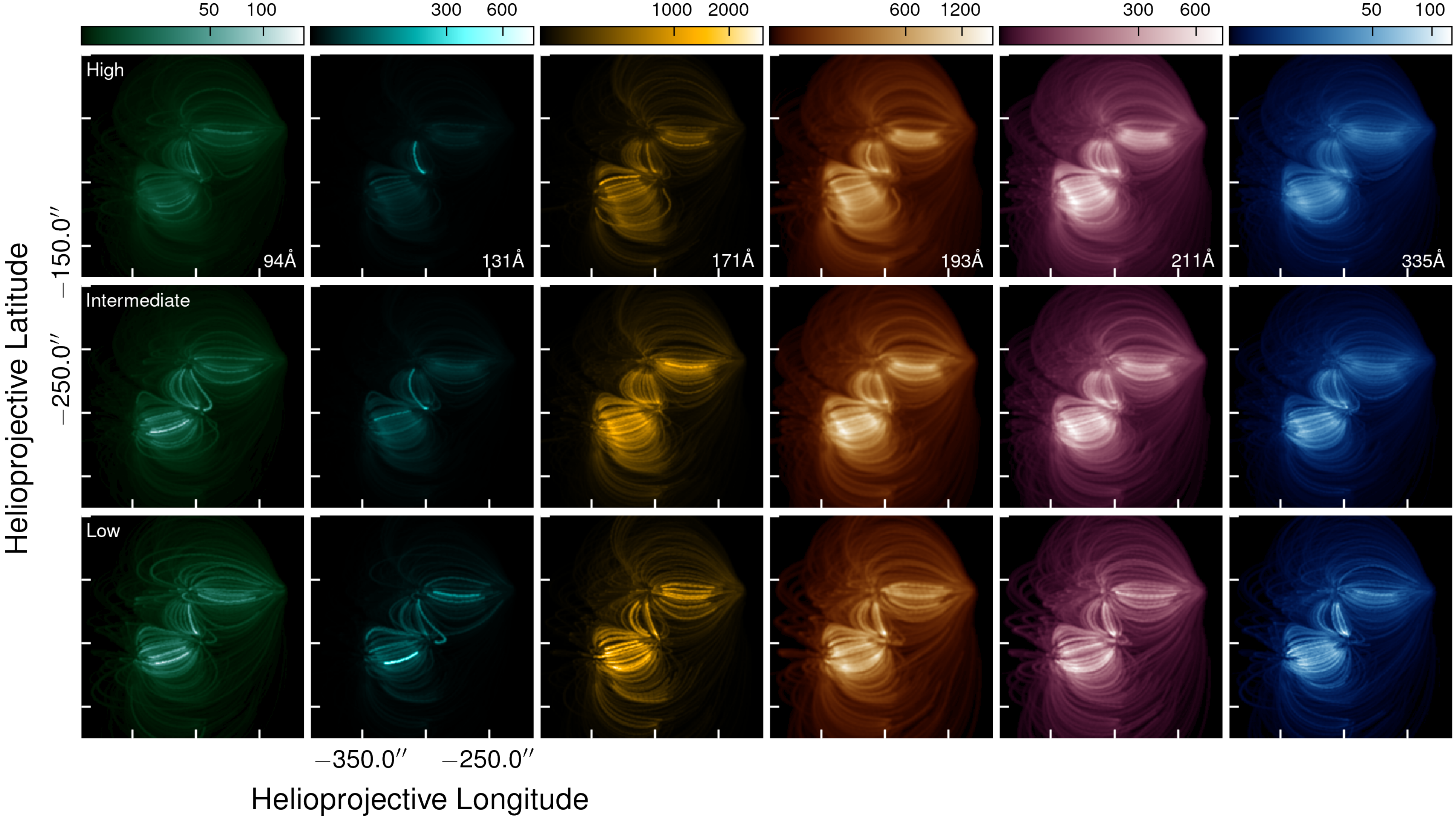
Element	# of Ions	# of Transitions
O	8	11892
Mg	11	31965
Si	13	30047
S	16	33091
Ca	17	42823
Fe	25	553541
Ni	19	83517

EBTEL Model

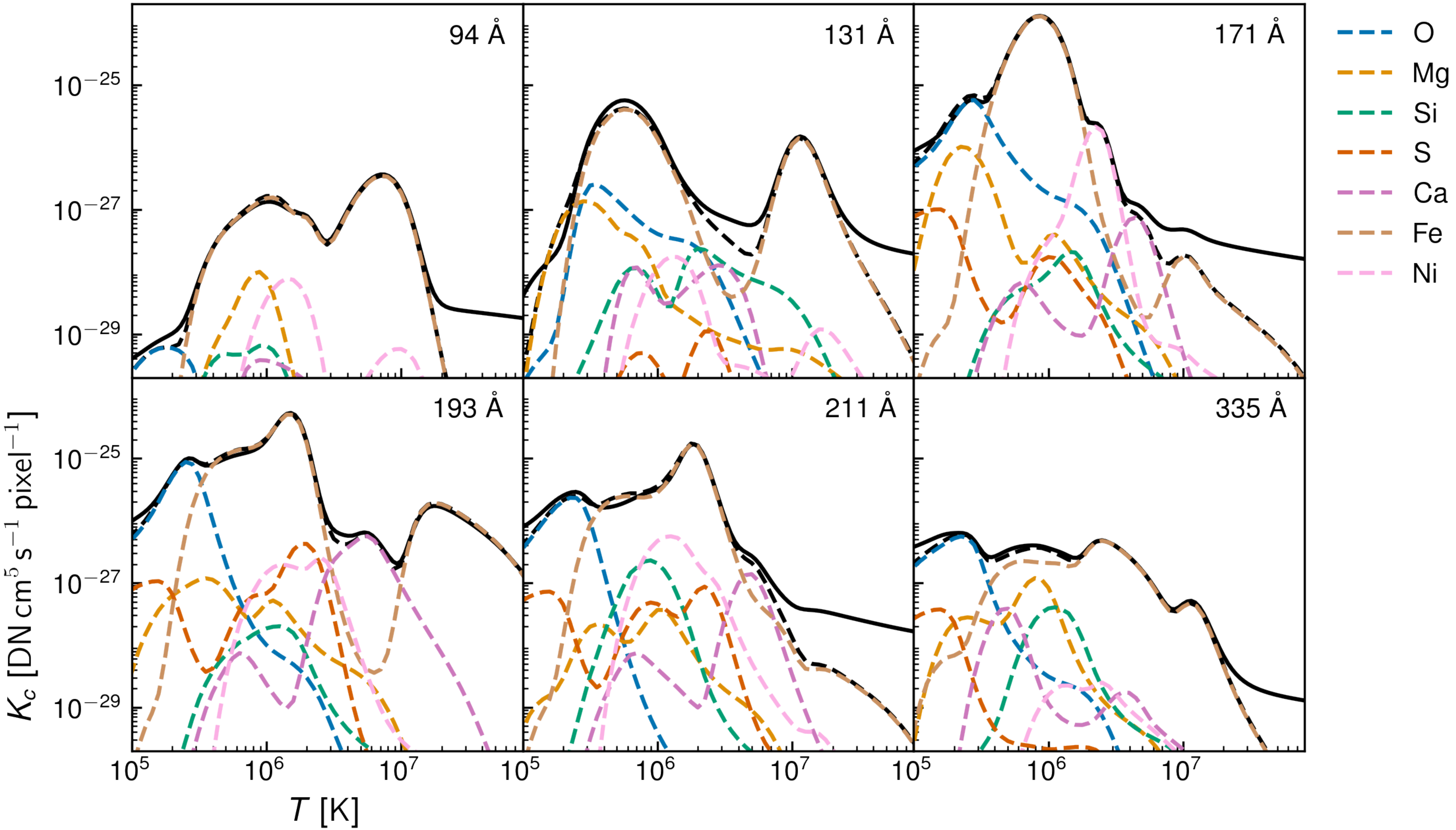
Klimchuk et al. (2008)
Cargill et al. (2012a,b)
Barnes et al. (2016a)

Mason and Monsignori Fossi (1994), Bradshaw and Raymond (2013), Del Zanna and Mason (2018)

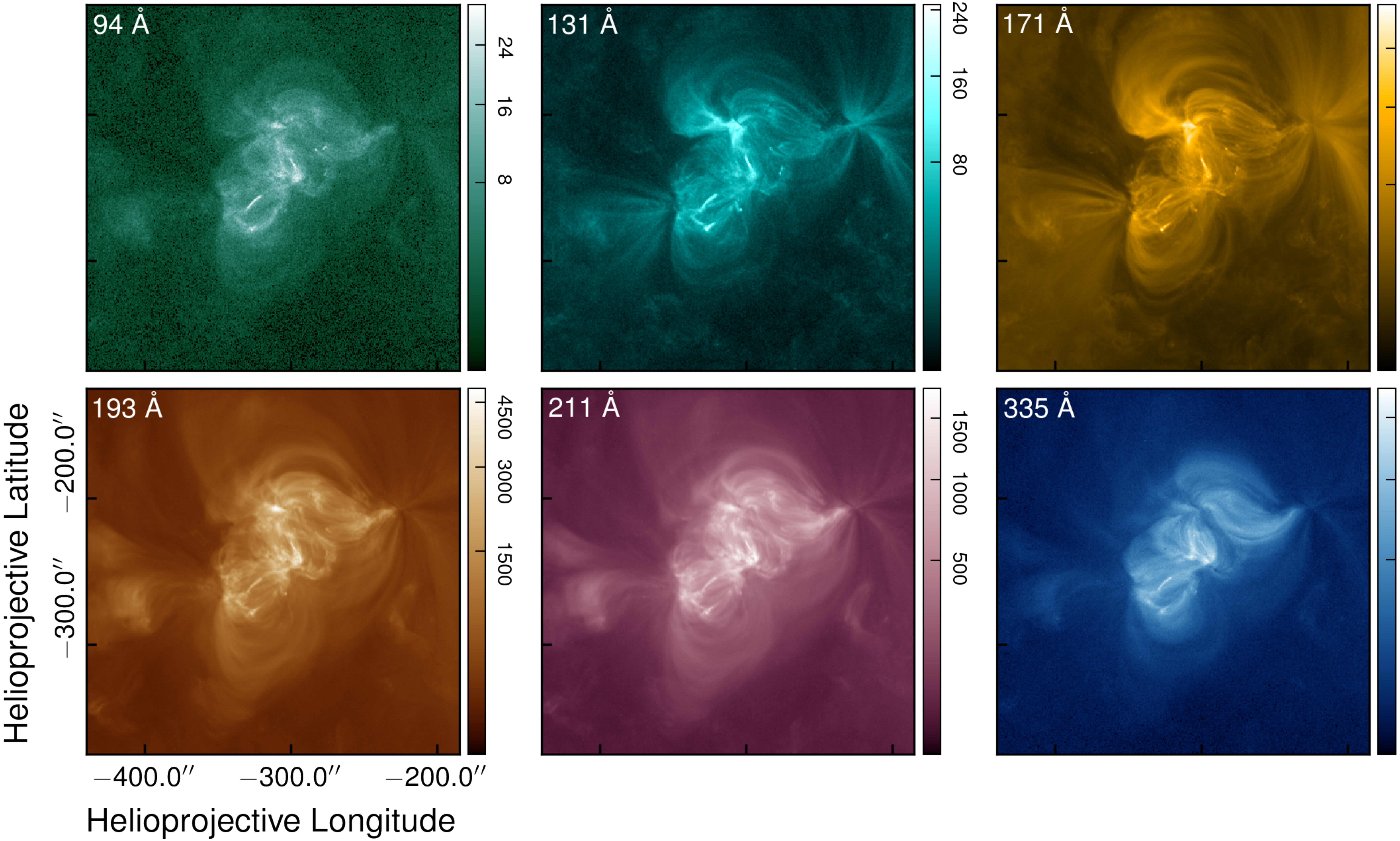
Modeling Emission from NOAA 1158



Effective AIA Response Functions



Observed Intensities



Random Forest—Data Preparation

1. Flatten all images (all frequencies) into "data" and "feature" matrices

$$\mathbf{X}_{model} = \underbrace{\begin{bmatrix} a_0 & \tau_{00} & \dots & \tau_{0N} & \mathscr{C}_{00} & \dots & \mathscr{C}_{0N} \\ a_1 & \tau_{10} & \dots & \tau_{1N} & \mathscr{C}_{10} & \dots & \mathscr{C}_{1N} \\ \vdots & & & \ddots & & & \vdots \\ a_M & \tau_{M0} & \dots & \tau_{MN} & \mathscr{C}_{M0} & \dots & \mathscr{C}_{MN} \end{bmatrix}}_{\text{\# of channel pairs (x2) + 1}} \text{Total \# of pixels}$$

$$\mathbf{Y}_{model} = \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_M \end{bmatrix} \text{frequency labels}$$

$$\mathbf{X}_{observed} = \begin{bmatrix} a_0 & \tau_{00} & \dots & \tau_{0N} & \mathscr{C}_{00} & \dots & \mathscr{C}_{0N} \\ a_1 & \tau_{10} & \dots & \tau_{1N} & \mathscr{C}_{10} & \dots & \mathscr{C}_{1N} \\ \vdots & & & \ddots & & & \vdots \\ a'_M & \tau_{M'0} & \dots & \tau_{M'N} & \mathscr{C}_{M'0} & \dots & \mathscr{C}_{M'N} \end{bmatrix}$$

$$\mathbf{Y}_{observed} = [?]$$

2. Split model data ($\mathbf{X}_{\text{model}}$) into training (2/3) and test set (1/3)

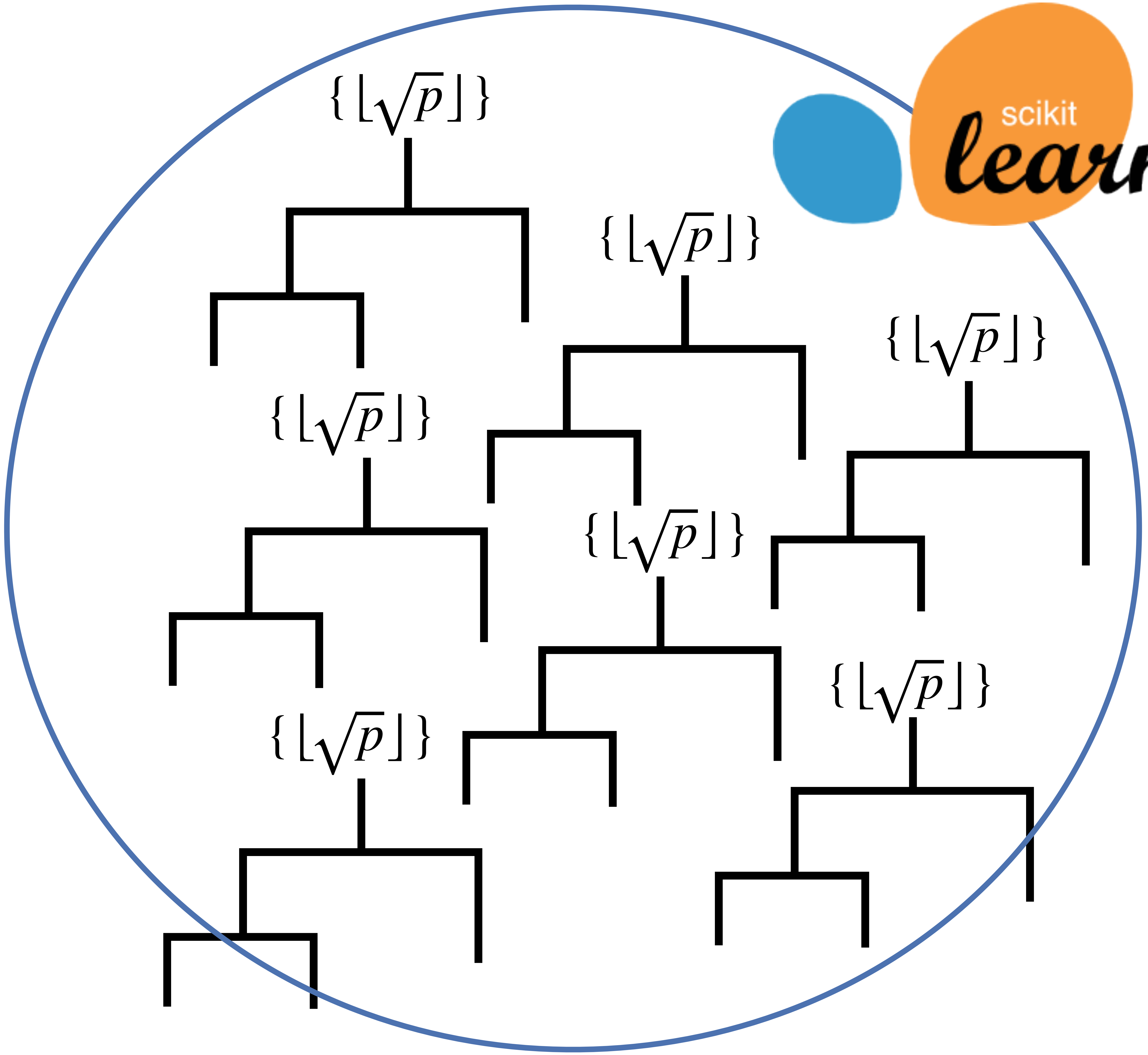
3. Train random forest on training set

4. Evaluate trained model performance on "unseen" test data

5. Classify observed pixels using trained model (and map back to coordinates)

Random Forests

- **Question: With which heating model are the observations most consistent?**
- Single decision tree = "weak learner"
- Combine multiple decision trees in a *random forest*
- Robust, efficient, easy to train
- Train on subsets of data, split on subsets of total features
- 500 total trees, maximum depth of 30



Breiman (2001), Hastie et al. (2009), Pedregosa et al. (2011)

Probability Maps — EM Slope Only

Probability Maps—EM Slope Only

Feature Importance

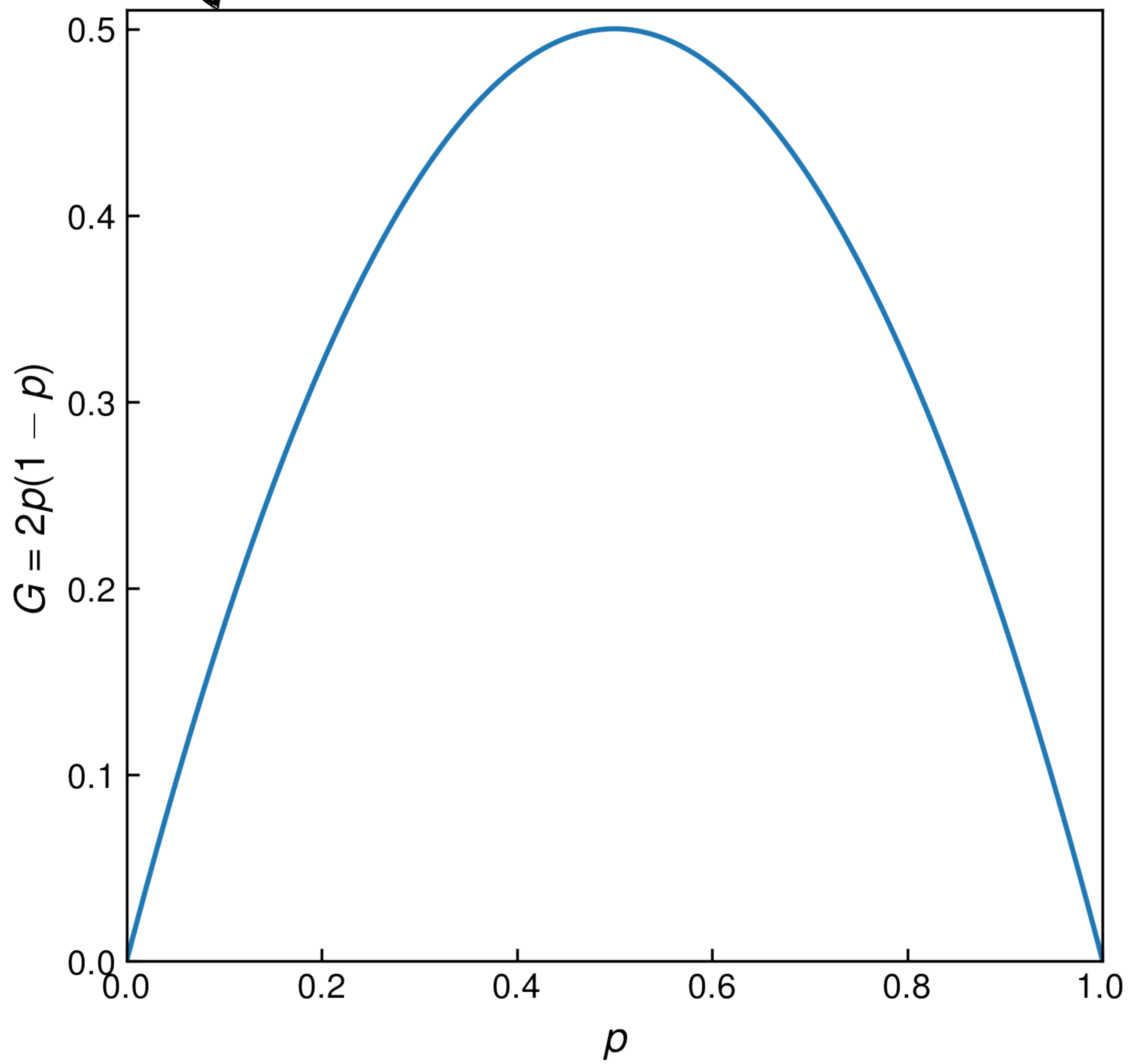
Gini Index

$$G_m = \sum_k \hat{p}_{mk}(1 - \hat{p}_{mk})$$

$$\hat{p}_{mk} = \frac{1}{M_m} \sum_{x_i \in R_m} I(y_i = k)$$

Reduction in Gini impurity

$$\Delta G_m = \frac{M_m}{M}\left(G_m - \frac{M_{m,R}}{M_m} G_{m,R} - \frac{M_{m,L}}{M_m} G_{m,L}\right)$$



Rank	Name
1	a
2	$\mathscr{C}_{211,193}$
3	$\mathscr{C}_{193,171}$
4	$\mathscr{C}_{211,171}$

Hastie et al. (2009)