An Introduction to Wavelet Analysis with Application to Water Quality Time Series

Don Percival

Applied Physics Laboratory
Department of Statistics
University of Washington
Seattle, Washington, USA

`http://faculty.washington.edu/dbp`

Overview: I

- as a subject, wavelets are
  - relatively new (1983 to present)
  - a synthesis of old/new ideas
  - keyword in 45,875 articles and books since 1989 (7685 more since 2008: an inundation of material!!!)
- wavelets can help us understand
  - time series (i.e., observations collected over time)
  - images

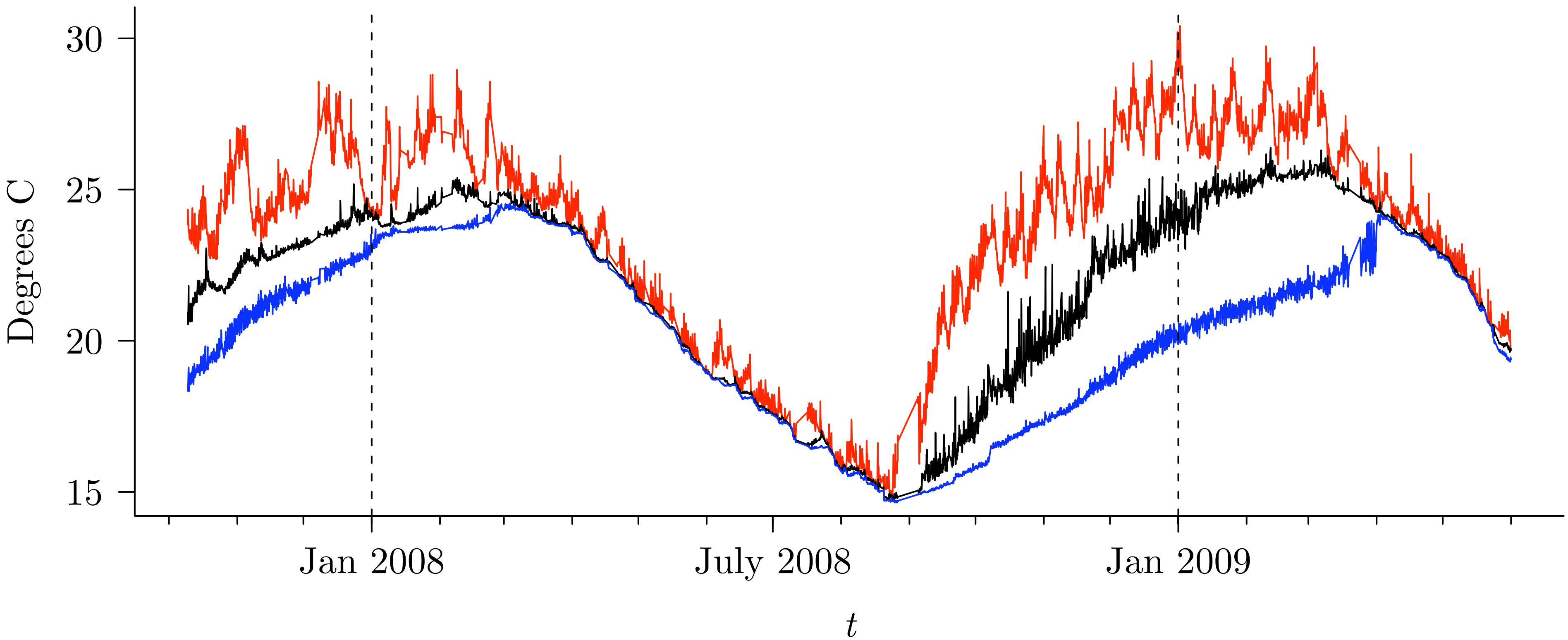
Overview: II

- wavelets capable of describing how
  - time series evolve over time on a given scale
  - images change from one place to the next on a given scale,

  where here 'scale' is either
  - an interval (span) of time (hour, year, . . .) or
  - a spatial area (square kilometer, acre, . . .)

Overview: III

- example: water temperatures recorded at wall of Wivenhoe Dam at depths of 1, 10 & 20 meters (9 October 2007 & on)



Overview: IV

- some questions that wavelets can help address:
  1. How does variance change across time?
  2. Are variations from one day to the next more prominent than variations from one month to the next?
  3. Temperatures at 10 and 20 meters are less variable than those at 1 meter, but are some of their other statistical properties similar?
  4. What are the pairwise relationships between these series on a scale-by-scale basis (e.g., day-to-day, month-to-month)?

Outline of Remainder of Talk: I

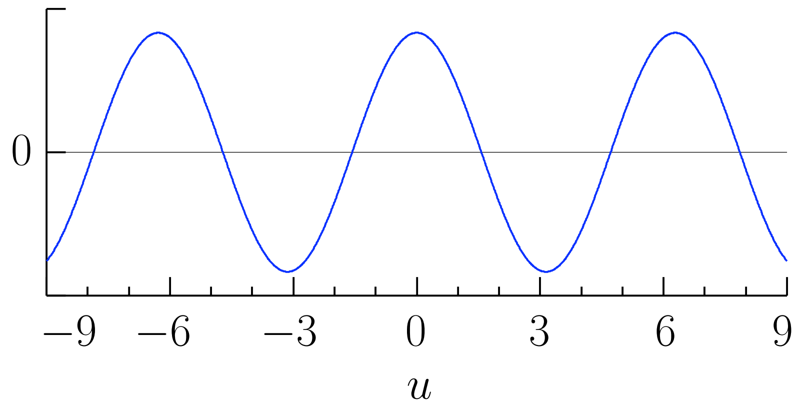
- discuss what exactly a wavelet is
- discuss wavelet analysis (emphasis on physical interpretation)
- point out two fundamental properties of the continuous wavelet transform (CWT):
  1. CWT is fully equivalent to the transformed time series
  2. CWT tells how 'energy' in time series is distributed across different scales and different times
- describe the discrete wavelet transform (DWT)
- point out two analogous fundamental properties of DWT

Outline of Remainder of Talk: II

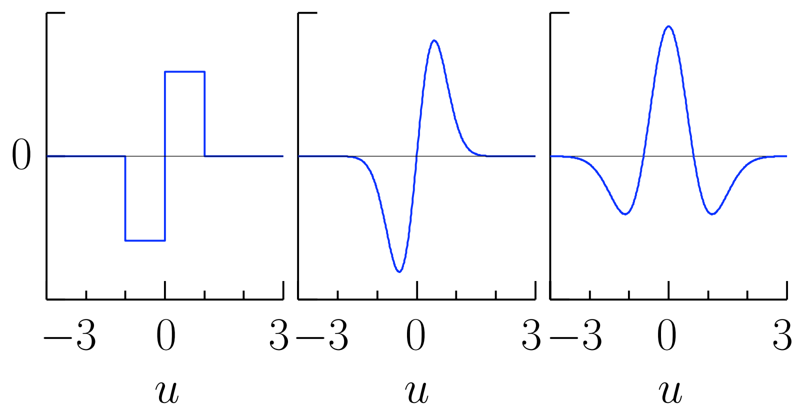
- look at some preliminary results from wavelet-based analysis of water temperature (on-going work with Sarah Lennox, You-Gan Wang and Ross Darnell)
- concluding remarks

What is a Wavelet?

- looking at $\cos(u)$ vs. u, a cosine is a 'big wave'

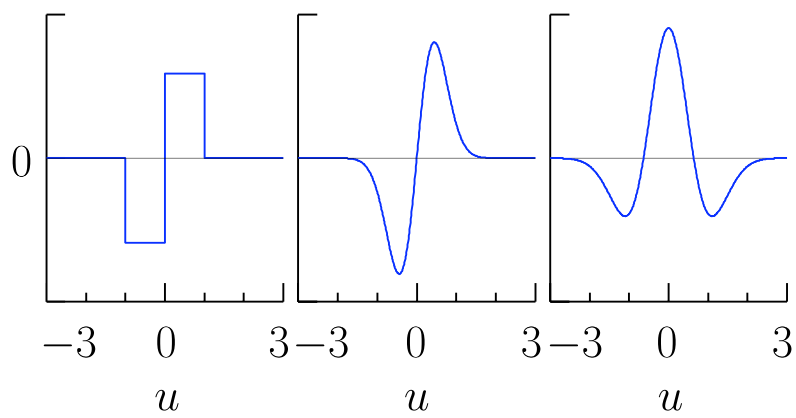


- wavelets are 'small waves' (left-hand is Haar wavelet $\psi^{(\mathrm{H})}(\cdot)$)



Technical Definition of a Wavelet

- real-valued function $\psi(\cdot)$ defined over real axis is a wavelet if
  1. integral of $\psi^2(\cdot)$ is unity: $\int_{-\infty}^{\infty} \psi^2(u)\, du = 1$
     (called 'unit energy' property, with apologies to physicists)
  2. integral of $\psi(\cdot)$ is zero: $\int_{-\infty}^{\infty} \psi(u)\, du = 0$
     (technically, need an 'admissibility condition,' but this is almost equivalent to integration to zero)



What is Wavelet Analysis?

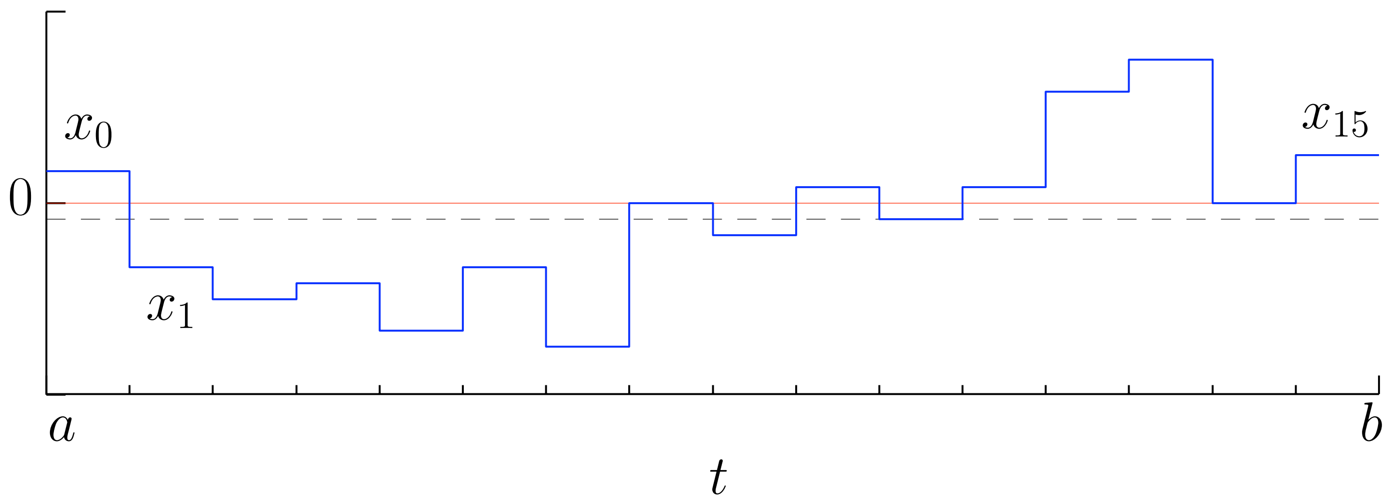
- wavelets tell us about variations in local averages
- to quantify this description, let $x(\cdot)$ be a time series
  - real-valued function of t defined over real axis
  - will refer to t as time (but it need not be such)
- consider 'average value' of $x(\cdot)$ over $[a, b]$:

$$\frac{1}{b-a}\int_a^b x(t)\,dt$$

(above notion discussed in elementary calculus books)

Example of Average Value of a Time Series

- let $x(\cdot)$ be step function taking on values $x_0, x_1, \ldots, x_{15}$ over 16 equal subintervals of $[a, b]$:



- here we have

$$\frac{1}{b-a}\int_a^b x(t)\,dt = \frac{1}{16}\sum_{j=0}^{15} x_j = \text{ height of dashed line}$$

Average Values at Different Scales and Times

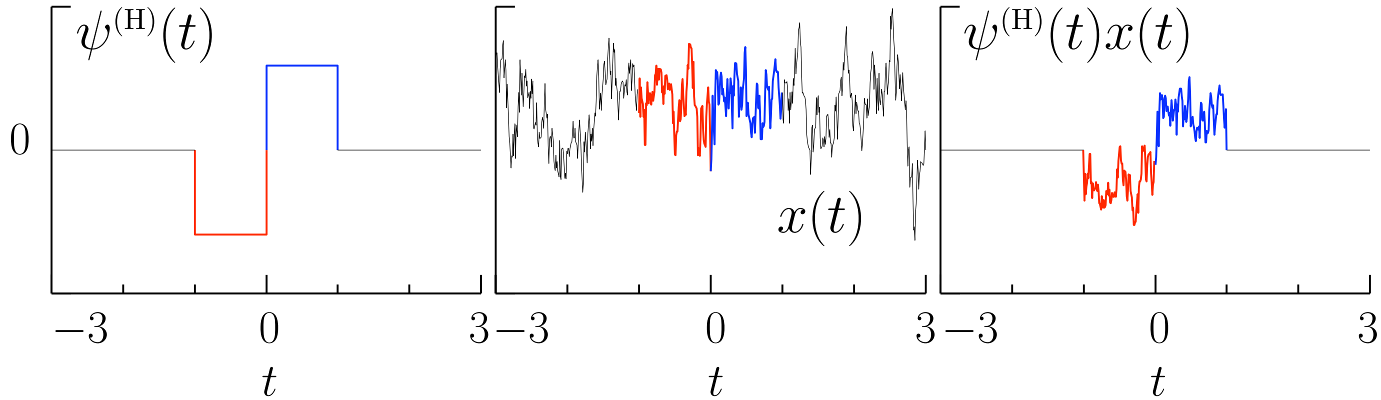
- define the following function of τ and t

$$A(\tau, t) \equiv \frac{1}{\tau} \int_{t-\frac{\tau}{2}}^{t+\frac{\tau}{2}} x(u)\, du$$

  – τ is width of interval – referred to as 'scale'
  – t is midpoint of interval
- $A(\tau, t)$ is average value of $x(\cdot)$ over scale τ centered at t
- average values of time series are of wide-spread interest
  – one second average temperatures over forest
  – ten minute rainfall rate during severe storm
  – yearly average temperatures over central England

Defining a Wavelet Coefficient W

- multiply Haar wavelet & time series $x(\cdot)$ together:



- integrate resulting function to get 'wavelet coefficient' $W(1,0)$:

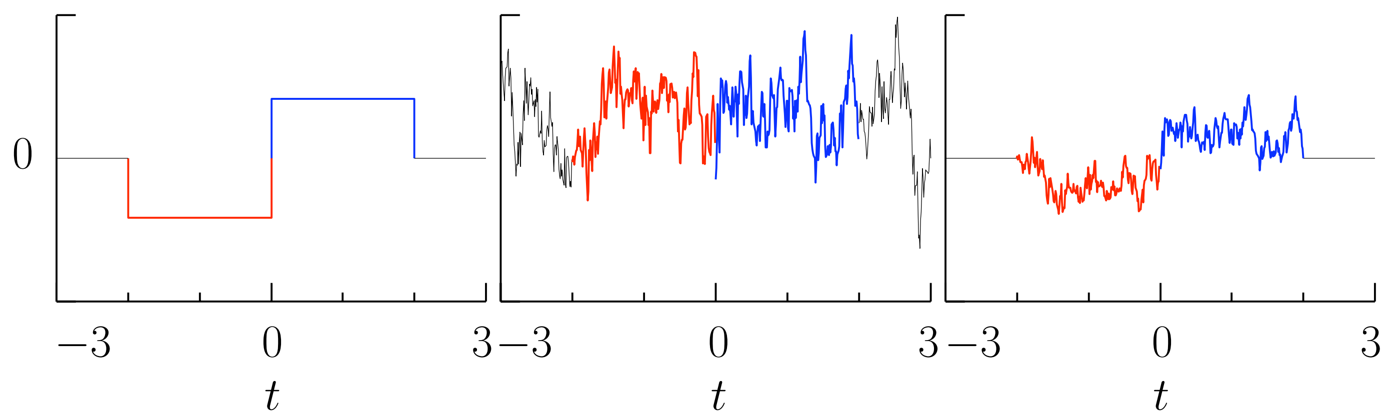
$$\int_{-\infty}^{\infty} \psi^{(\mathrm{H})}(t)x(t)\,dt = W(1,0)$$

- to see what $W(1,0)$ is telling us about $x(\cdot)$, note that

$$W(1,0) \propto \frac{1}{1}\int_{0}^{1} x(t)\,dt - \frac{1}{1}\int_{-1}^{0} x(t)\,dt = A(1,\tfrac{1}{2}) - A(1,-\tfrac{1}{2})$$

Defining Wavelet Coefficients for Other Scales

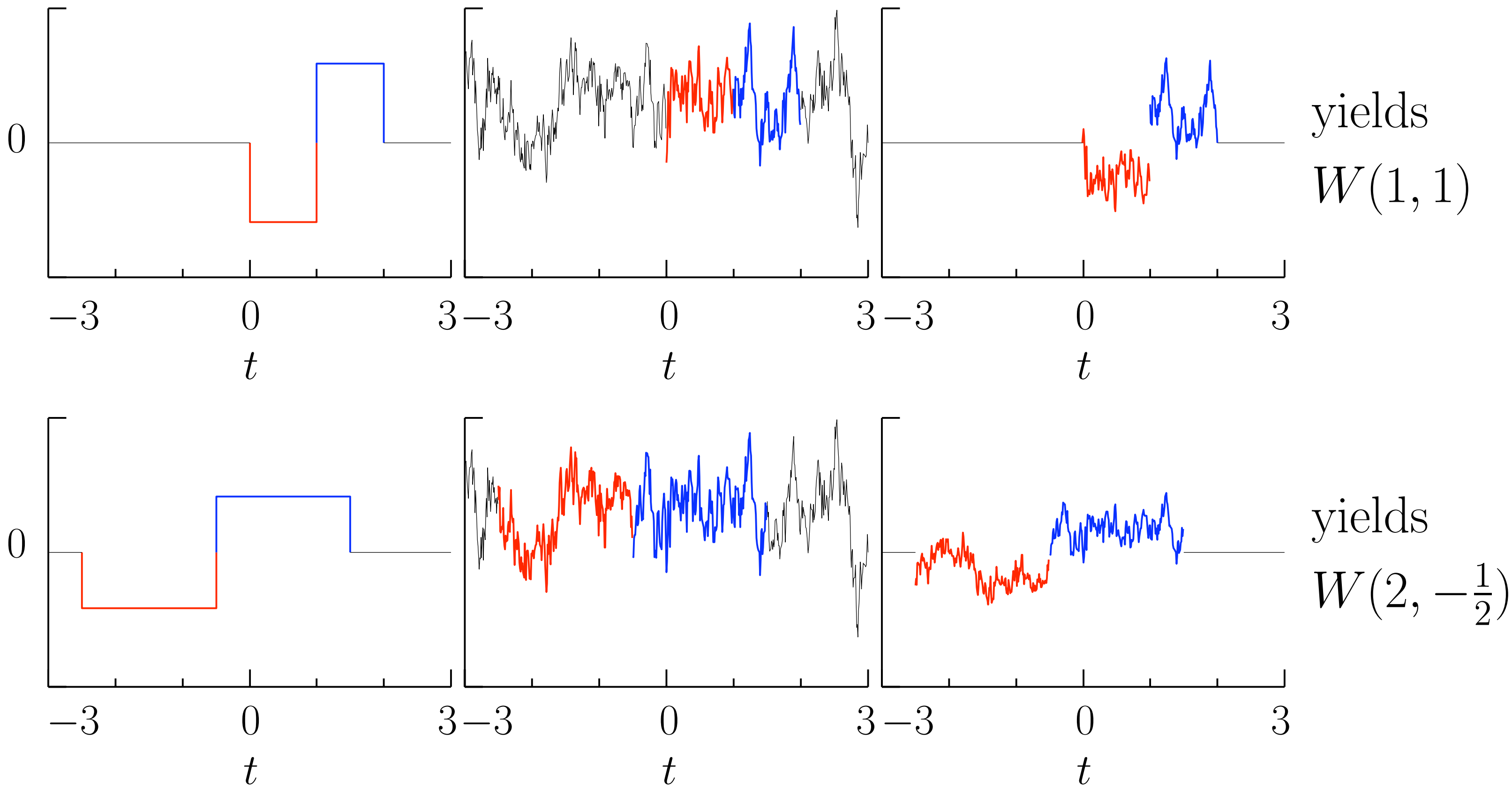
- $W(1,0)$ proportional to difference between averages of $x(\cdot)$ over $[-1,0]$ & $[0,1]$, i.e., two unit-scale averages before/after $t=0$
  - '1' in $W(1,0)$ denotes scale 1 (width of each interval)
  - '0' in $W(1,0)$ denotes time 0 (center of combined intervals)
- stretch or shrink wavelet to define $W(\tau,0)$ for other scales τ:



yields $W(2,0)$

Defining Wavelet Coefficients for Other Locations

- relocate to define $W(\tau, t)$ for other times t:



Haar Continuous Wavelet Transform (CWT)

- for all $\tau > 0$ and all $-\infty < t < \infty$, can write
$$W(\tau, t) = \frac{1}{\sqrt{\tau}} \int_{-\infty}^{\infty} x(u) \psi^{(\mathrm{H})}\left(\frac{u-t}{\tau}\right) du$$
  - $\frac{u-t}{\tau}$ does the stretching/shrinking and relocating
  - $\frac{1}{\sqrt{\tau}}$ needed so $\psi^{(\mathrm{H})}_{\tau,t}(u) \equiv \frac{1}{\sqrt{\tau}} \psi^{(\mathrm{H})}\left(\frac{u-t}{\tau}\right)$ has unit energy
  - since it also integrates to zero, $\psi^{(\mathrm{H})}_{\tau,t}(\cdot)$ is a wavelet
- $W(\tau, t)$ over all $\tau > 0$ and all t is Haar CWT for $x(\cdot)$
- analyzes/breaks up/decomposes $x(\cdot)$ into components
  - associated with a scale and a time
  - physically related to a difference of averages

Other Continuous Wavelet Transforms: I

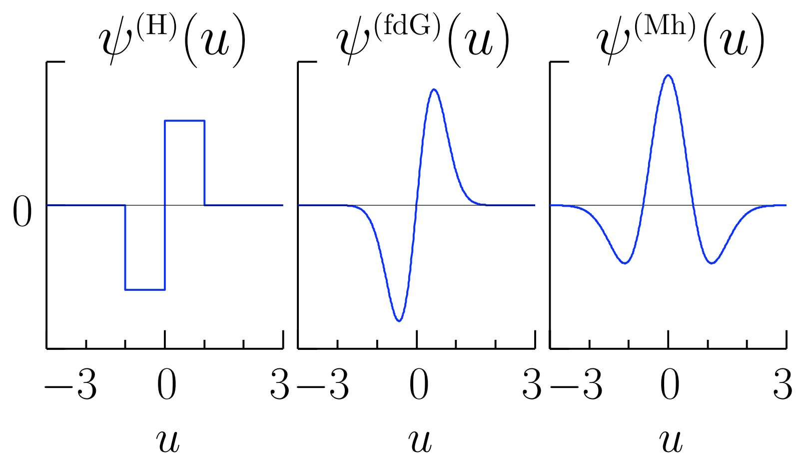
- can do the same for wavelets other than the Haar
- start with basic wavelet $\psi(\cdot)$
- use $\psi_{\tau,t}(u) = \frac{1}{\sqrt{\tau}}\psi\left(\frac{u-t}{\tau}\right)$ to stretch/shrink & relocate
- define CWT via

$$W(\tau, t) = \int_{-\infty}^{\infty} x(u)\psi_{\tau,t}(u)\, du = \frac{1}{\sqrt{\tau}} \int_{-\infty}^{\infty} x(u)\psi\left(\frac{u-t}{\tau}\right) du$$

- analyzes/breaks up/decomposes $x(\cdot)$ into components
  - associated with a scale and a time
  - physically related to a difference of *weighted* averages

Other Continuous Wavelet Transforms: II

- consider two companions of Haar wavelet:



- $\psi^{(\mathrm{fdG})}(\cdot)$ proportional to 1st derivative of Gaussian PDF
- 'Mexican hat' wavelet $\psi^{(\mathrm{Mh})}(\cdot)$ proportional to 2nd derivative
- $\psi^{(\mathrm{fdG})}(\cdot)$ looks at difference of adjacent *weighted* averages
- $\psi^{(\mathrm{Mh})}(\cdot)$ looks at difference between weighted average and sum of weighted averages occurring before & after

First Scary-Looking Equation

- CWT equivalent to $x(\cdot)$ because we can write
$$x(t) = \int_0^\infty \left[\frac{1}{C\tau^2} \int_{-\infty}^{\infty} W(\tau, u) \frac{1}{\sqrt{\tau}} \psi\left(\frac{t-u}{\tau}\right) du \right] d\tau,$$
where C is a constant depending on specific wavelet $\psi(\cdot)$
- can synthesize (put back together) $x(\cdot)$ given its CWT; i.e., nothing is lost in reexpressing time series $x(\cdot)$ via its CWT
- regard stuff in brackets as defining 'scale τ' time series at t
- says we can reexpress $x(\cdot)$ as integral (sum) of new time series, each associated with a particular scale
- similar additive decompositions are a central theme of wavelet analysis

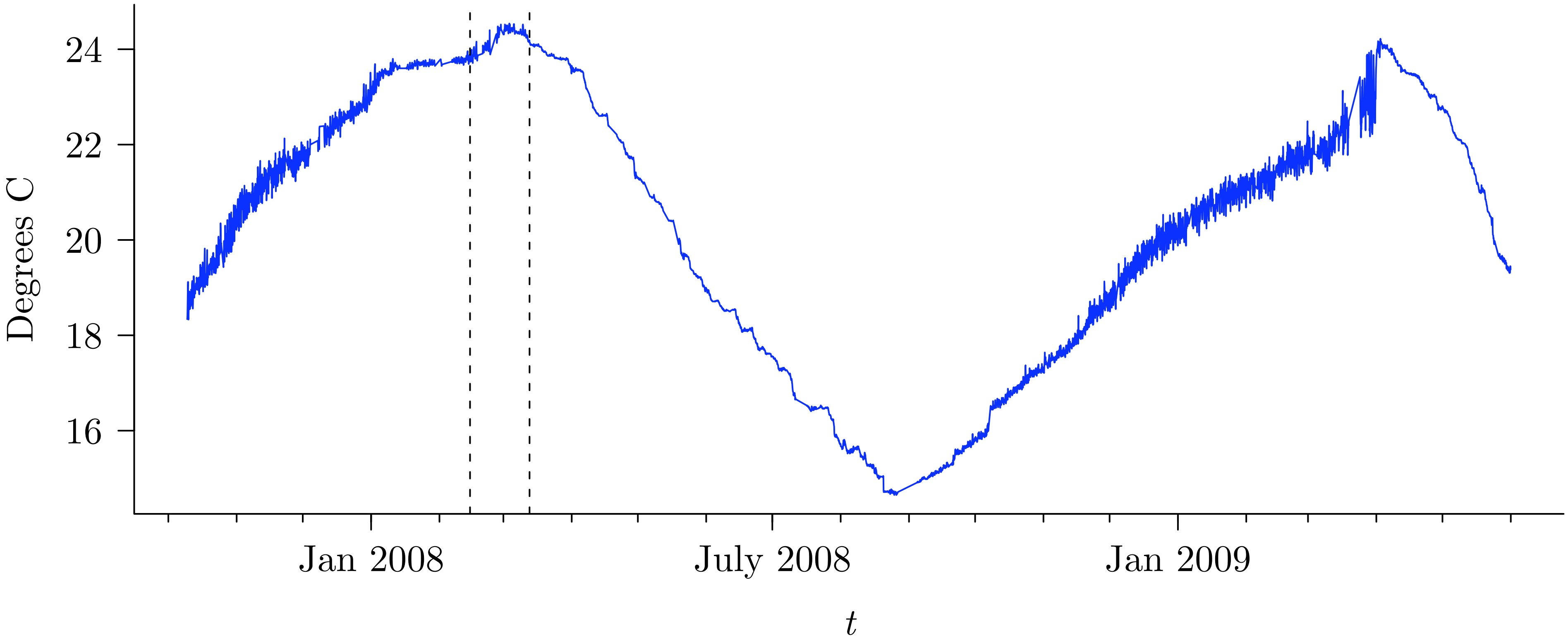
Second Scary-Looking Equation

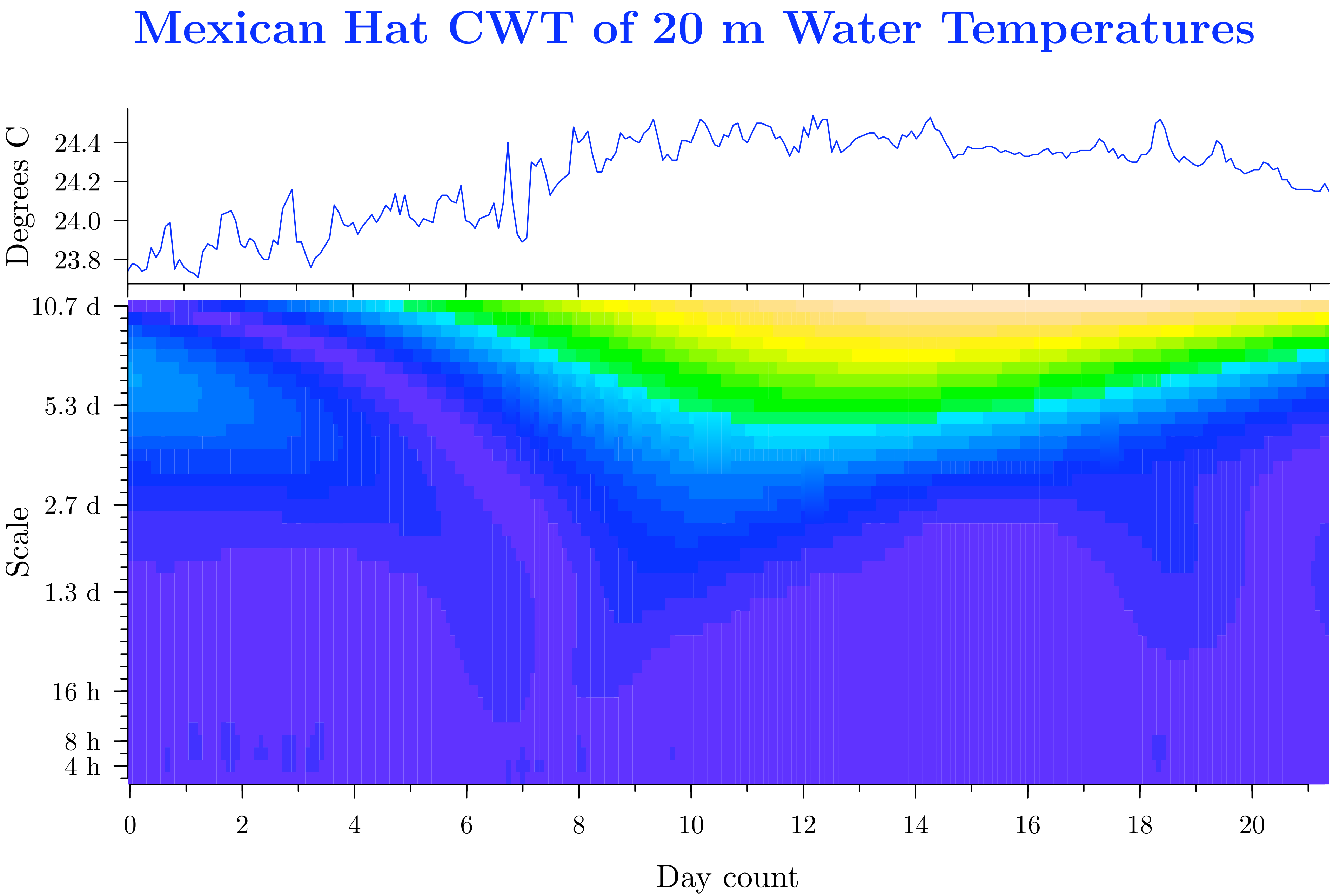
- energy in $x(\cdot)$ is reexpressed in CWT because

$$\text{energy} = \int_{-\infty}^{\infty} x^2(t)\,dt = \int_0^{\infty} \left[\frac{1}{C\tau^2}\int_{-\infty}^{\infty} W^2(\tau,t)\,dt\right] d\tau$$

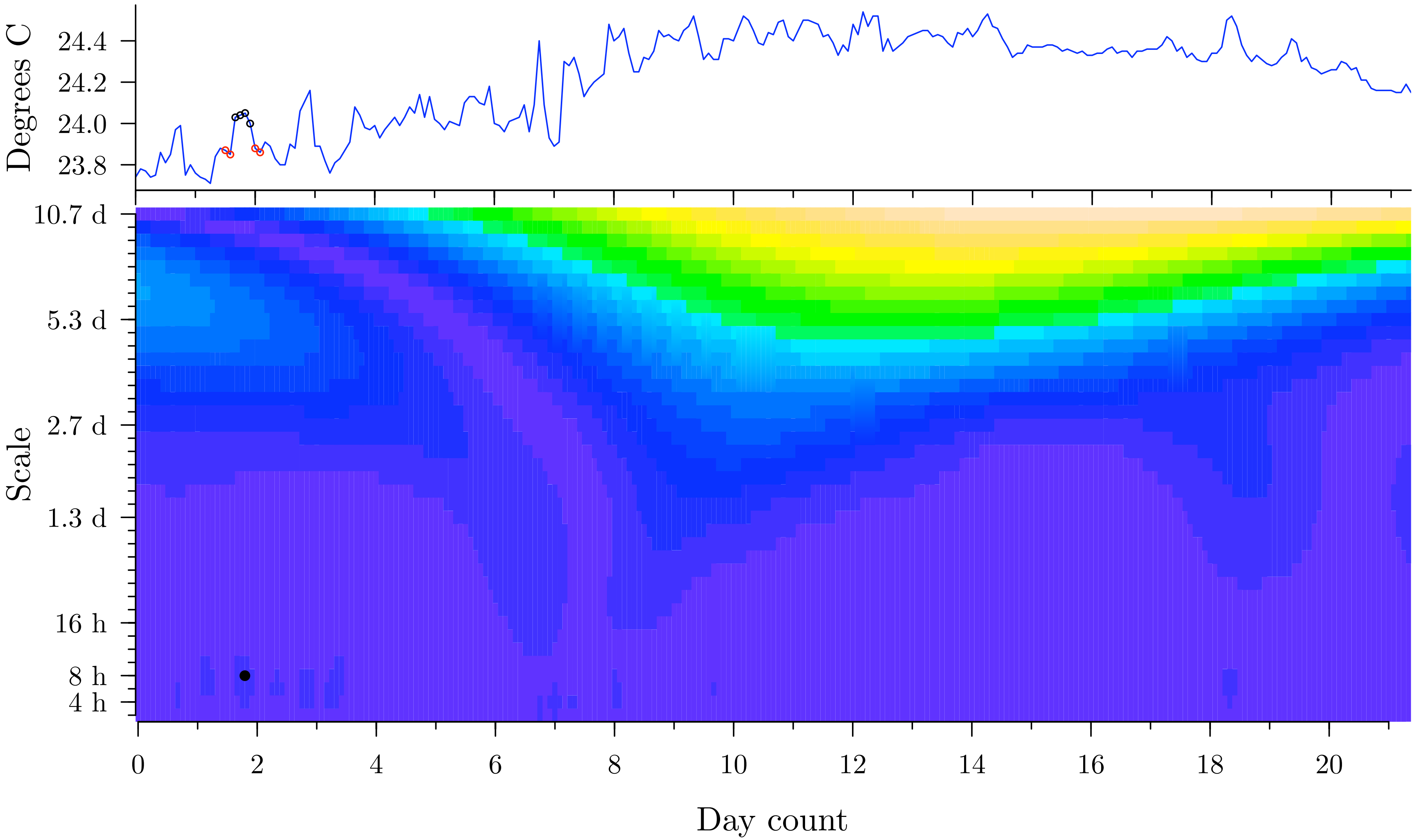
- can regard $x^2(t)$ versus t as breaking up the energy across time (i.e., an 'energy density' function)
- regard stuff in brackets as breaking up the energy across scales
- says we can reexpress energy as integral (sum) of components, each associated with a particular scale
- function defined by $W^2(\tau,t)/C\tau^2$ is an energy density across both time and scale
- similar energy decompositions are a second central theme

Example: Portion of Water Temperatures at 20 m

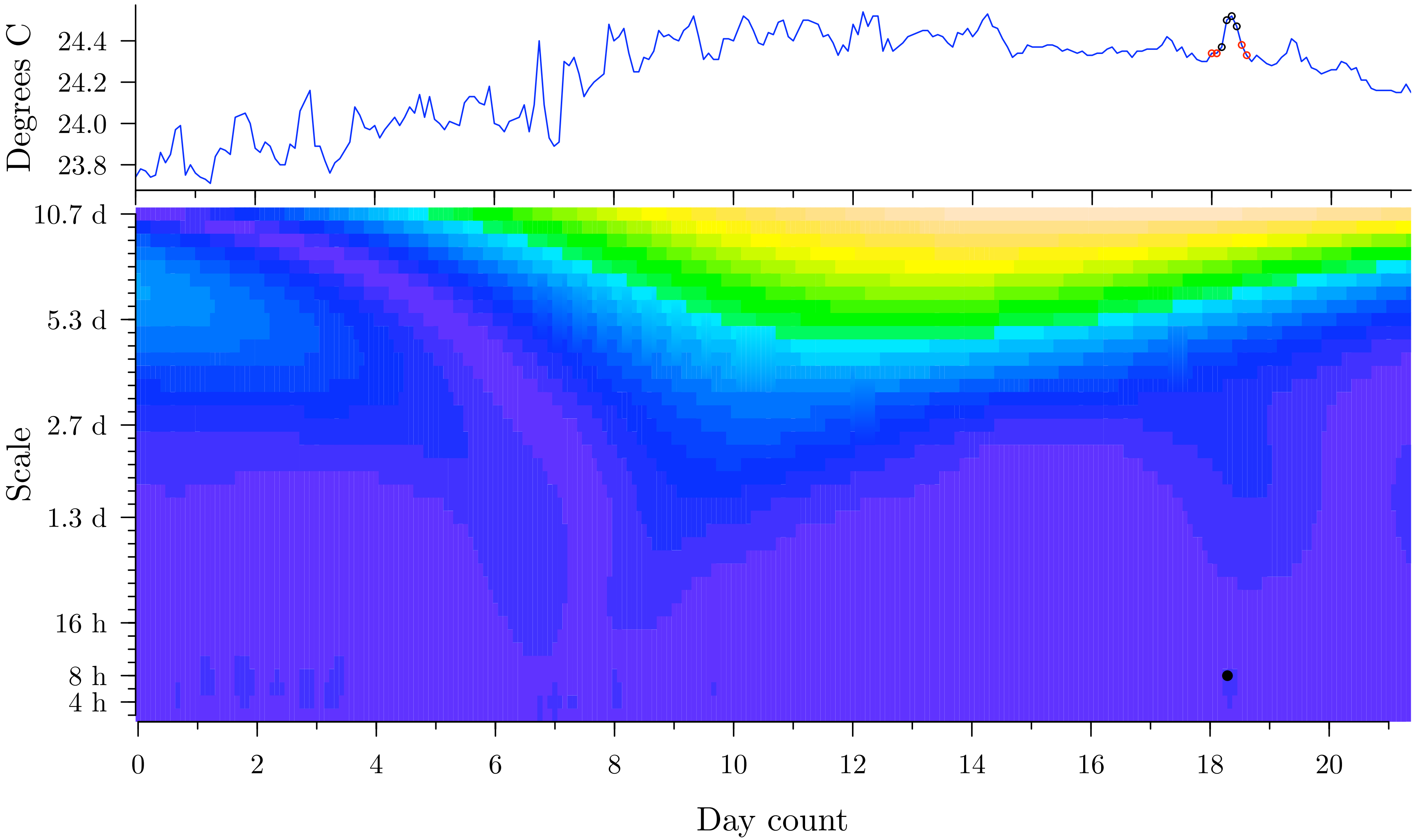



Mexican Hat CWT of 20 m Water Temperatures
Degrees C
24.4
24.2
24.0
23.8
Scale
10.7 d
5.3 d
2.7 d
1.3 d
16 h
8 h
4 h
0
2
4
6
8
10
12
14
16
18
20
Day count

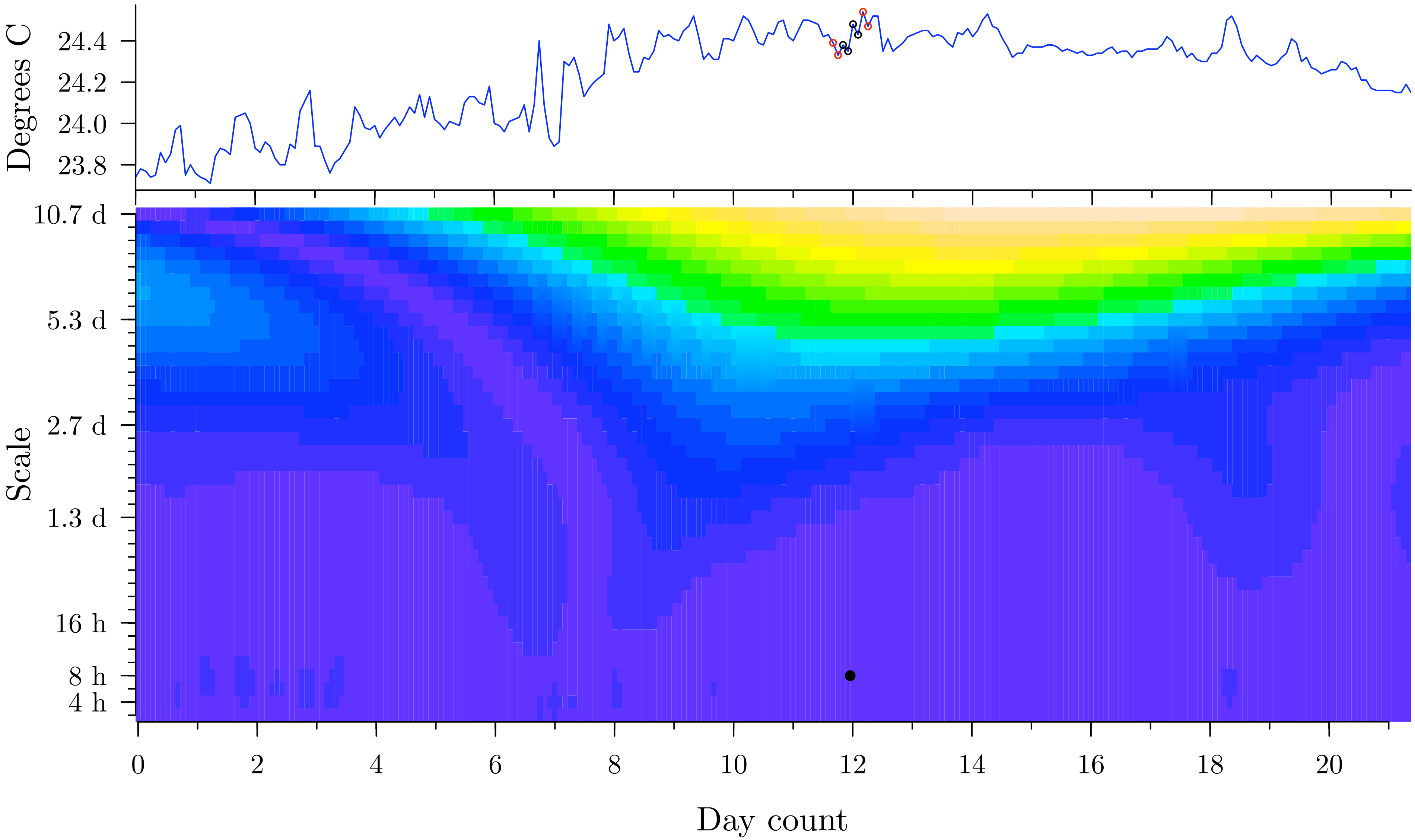
Mexican Hat CWT of 20 m Water Temperatures



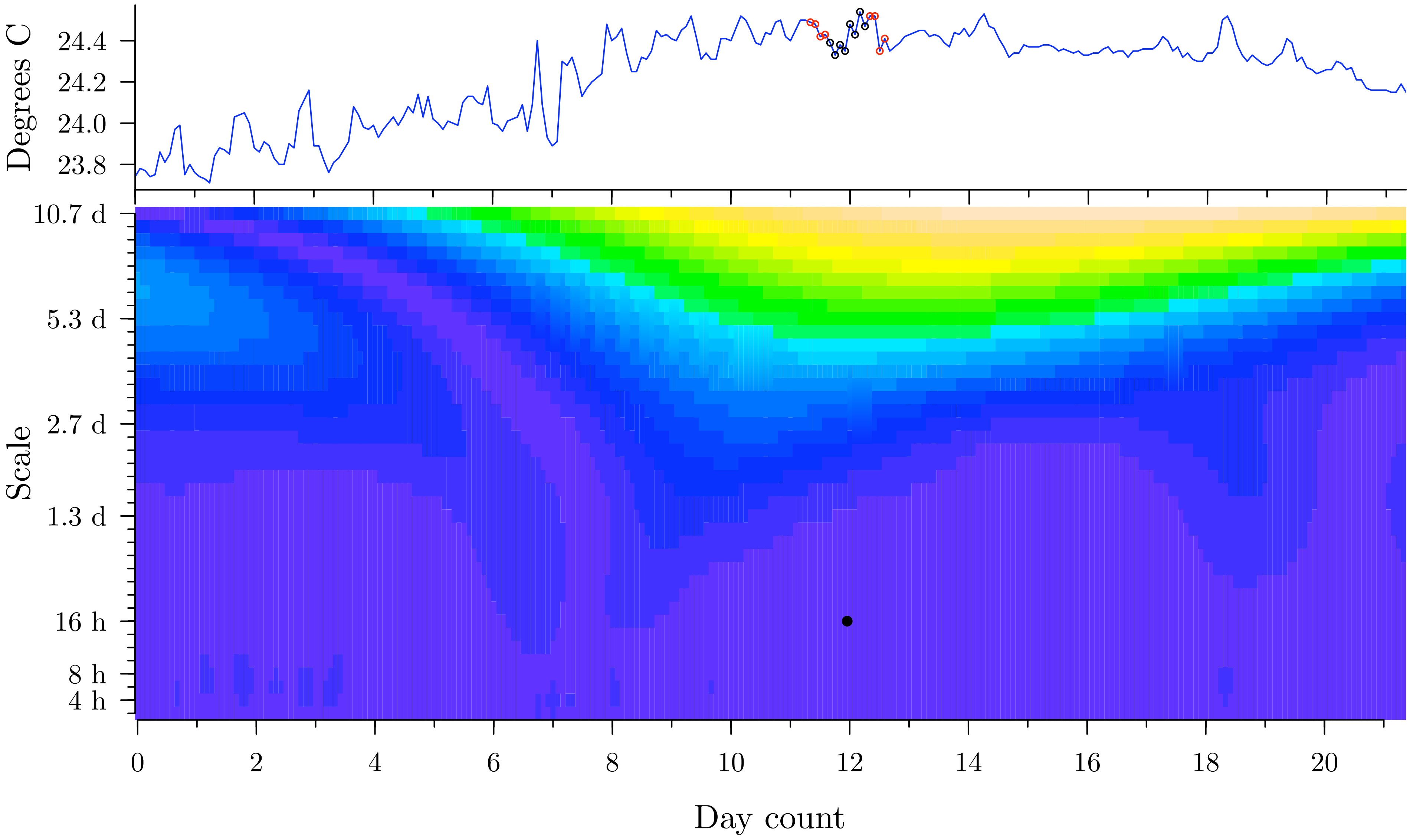
Mexican Hat CWT of 20 m Water Temperatures



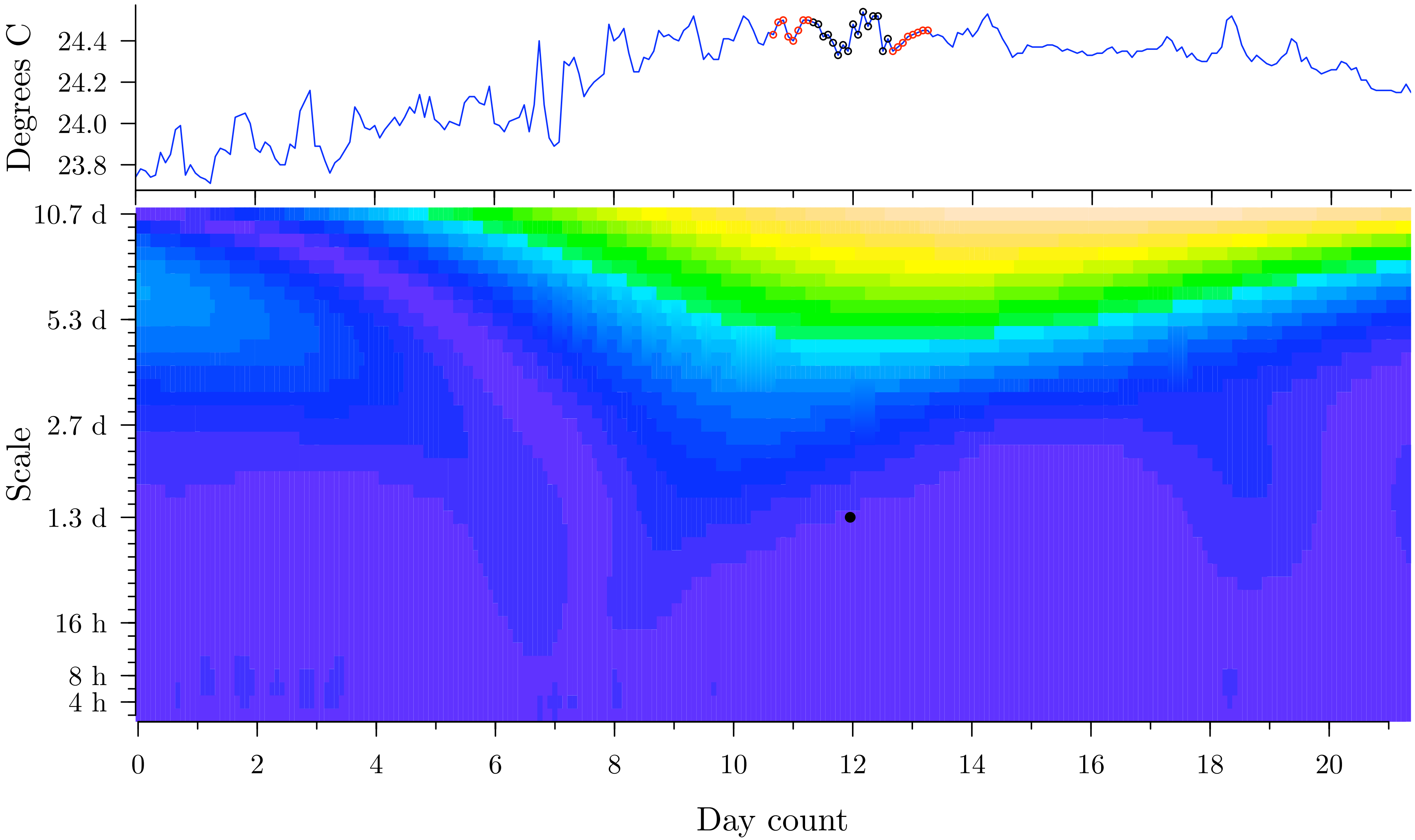
Mexican Hat CWT of 20 m Water Temperatures



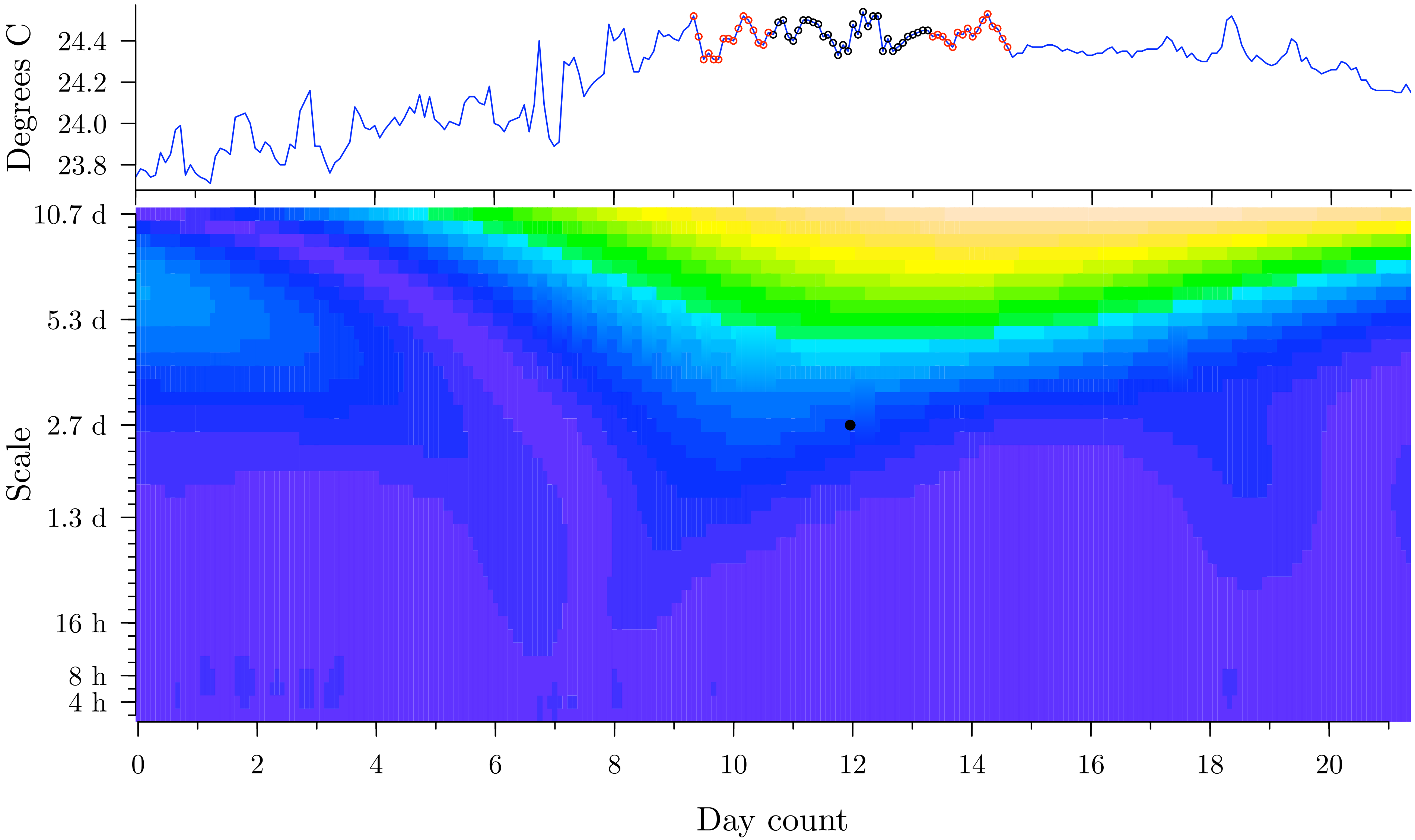
Mexican Hat CWT of 20 m Water Temperatures



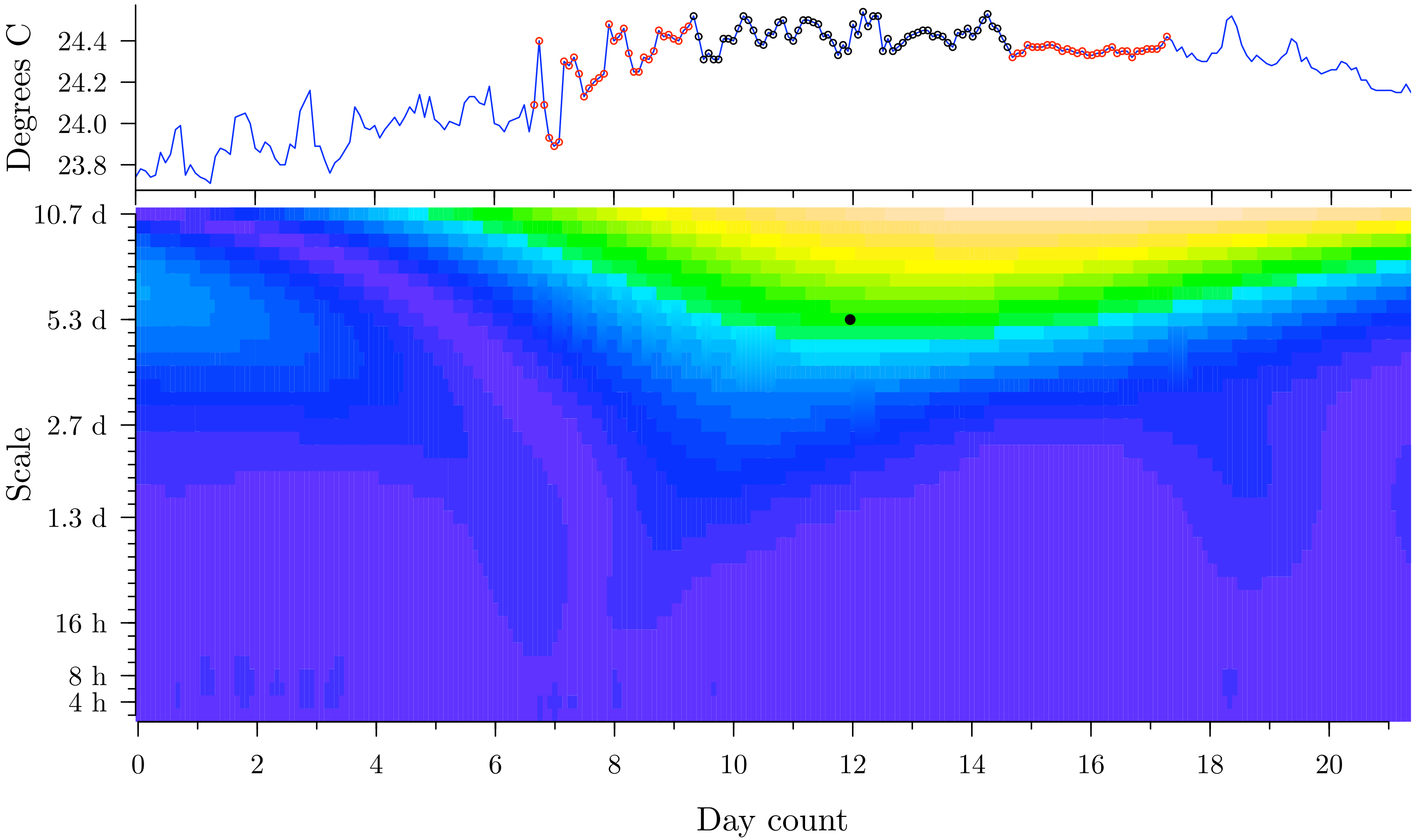
Mexican Hat CWT of 20 m Water Temperatures



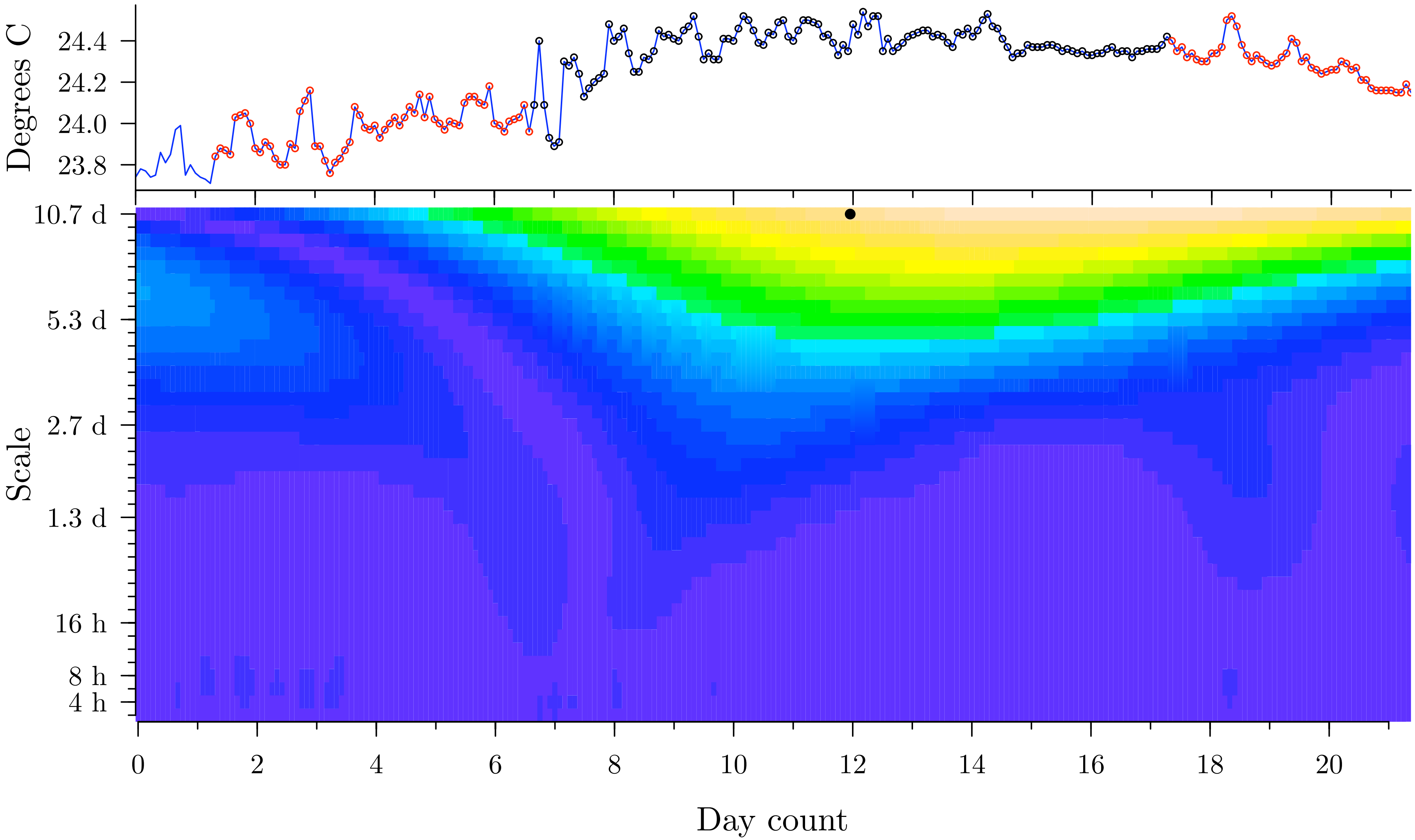
Mexican Hat CWT of 20 m Water Temperatures



Mexican Hat CWT of 20 m Water Temperatures



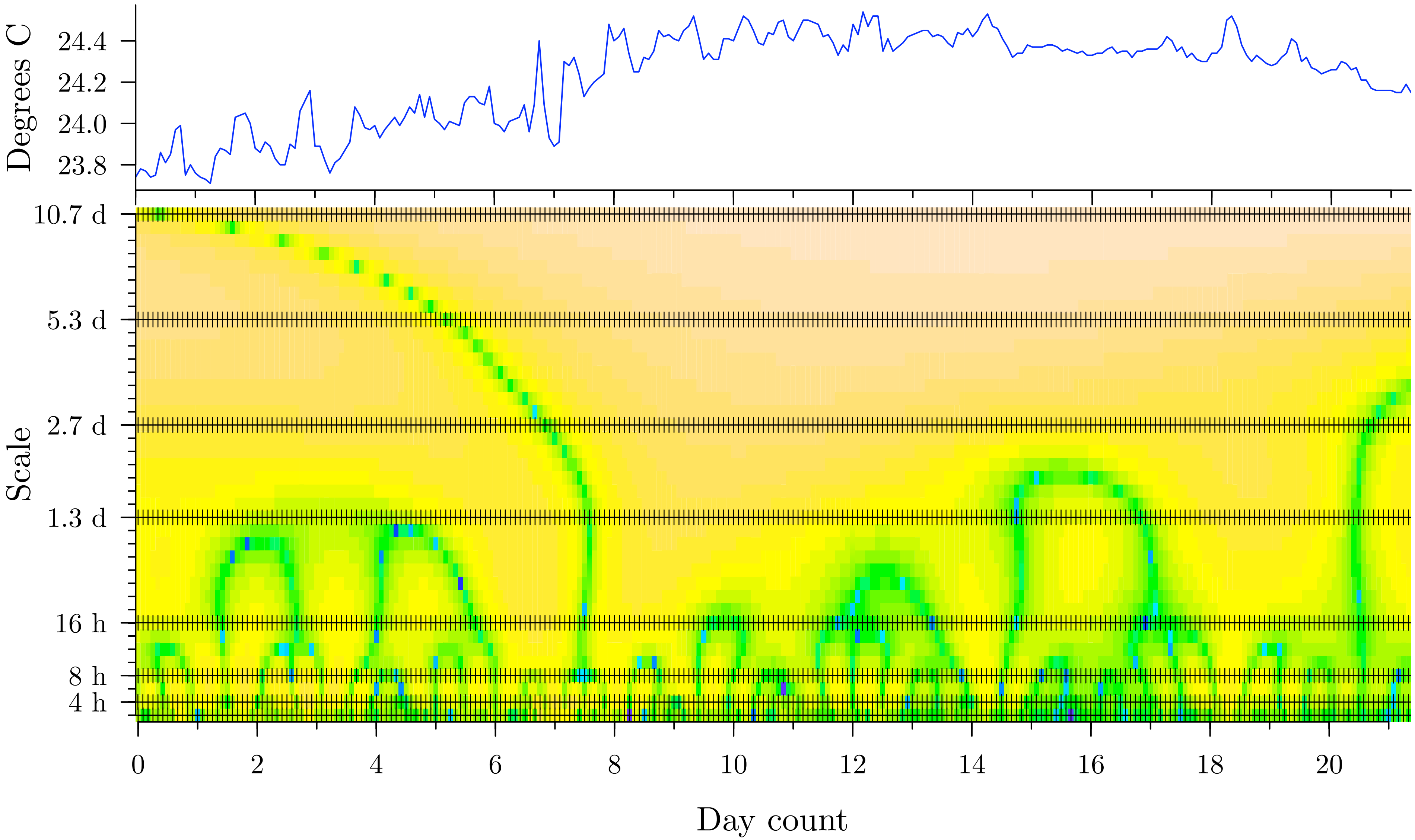
Mexican Hat CWT of 20 m Water Temperatures

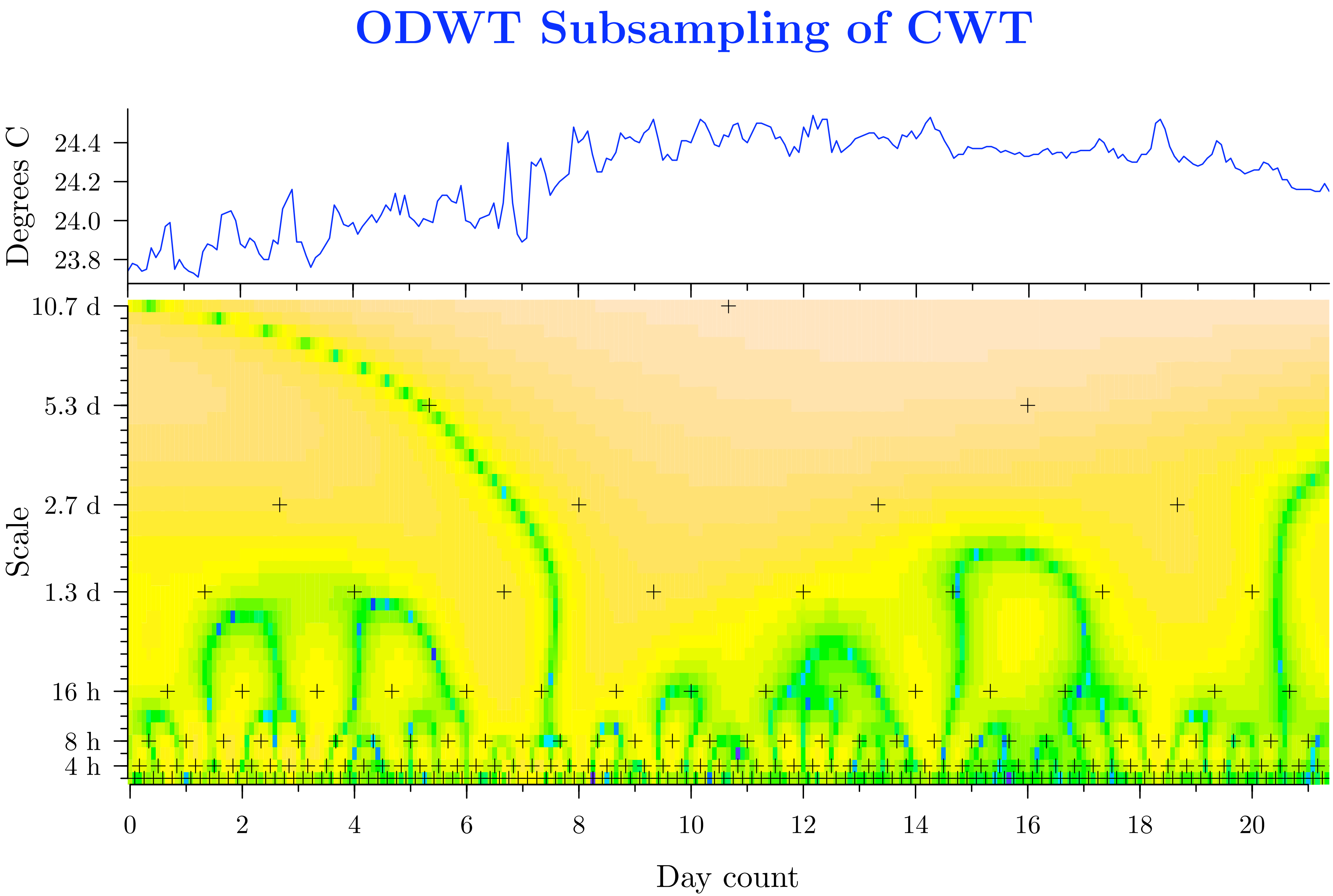


Beyond the CWT: the DWT

- critique: have transformed signal into an image (anti-statistics!)
- can often – but not always – get by with subsamples of $W(\tau, t)$
- leads to notion of discrete wavelet transform (DWT) – more convenient for use with samples x_0, x_1, $\ldots x_{N-1}$ from $x(\cdot)$,
- can regard DWT as 'slices' through CWT
  - restrict τ to 'dyadic' scales $\tau_j \equiv 2^{j-1}\Delta_t$, $j = 1, 2, \ldots, J$, where Δ_t is sampling interval (2 hours for water temperature data), and J is a maximum level chosen by user
  - restrict times to an offset $+\, t\,\Delta_t$, $t = 0,\ 1,\ \ldots,\ N-1$
  - this yields 'maximal overlap' DWT (MODWT) – can restrict times even further to get orthonormal DWT (ODWT)
- yields wavelet coefficients $W_{j,t} \propto W(\tau_j, t)$

MODWT Subsampling of CWT




ODWT Subsampling of CWT
Degrees C
24.4
24.2
24.0
23.8
Scale
10.7 d
5.3 d
2.7 d
1.3 d
16 h
8 h
4 h
0
2
4
6
8
10
12
14
16
18
20
Day count

The Discrete Wavelet Transform

- collect $W_{j,t}$ into vector $\mathbf{W}_j$ for levels $j = 1, 2, \ldots, J$
- also get scaling coefficients $V_{J,t}$
  - related to averages over a scale of $2\tau_J$
  - summary of information in $W(\tau, t)$ at $\tau \geq 2\tau_J = 2^J\,\Delta_t$
- collect $V_{J,t}$ into vector $\mathbf{V}_J$
- $\mathbf{W}_1, \ldots, \mathbf{W}_J$ and $\mathbf{V}_J$ form the DWT of $\mathbf{X} \equiv [x_0, \ldots, x_{N-1}]^T$

Multiresolution Analysis (MRA)

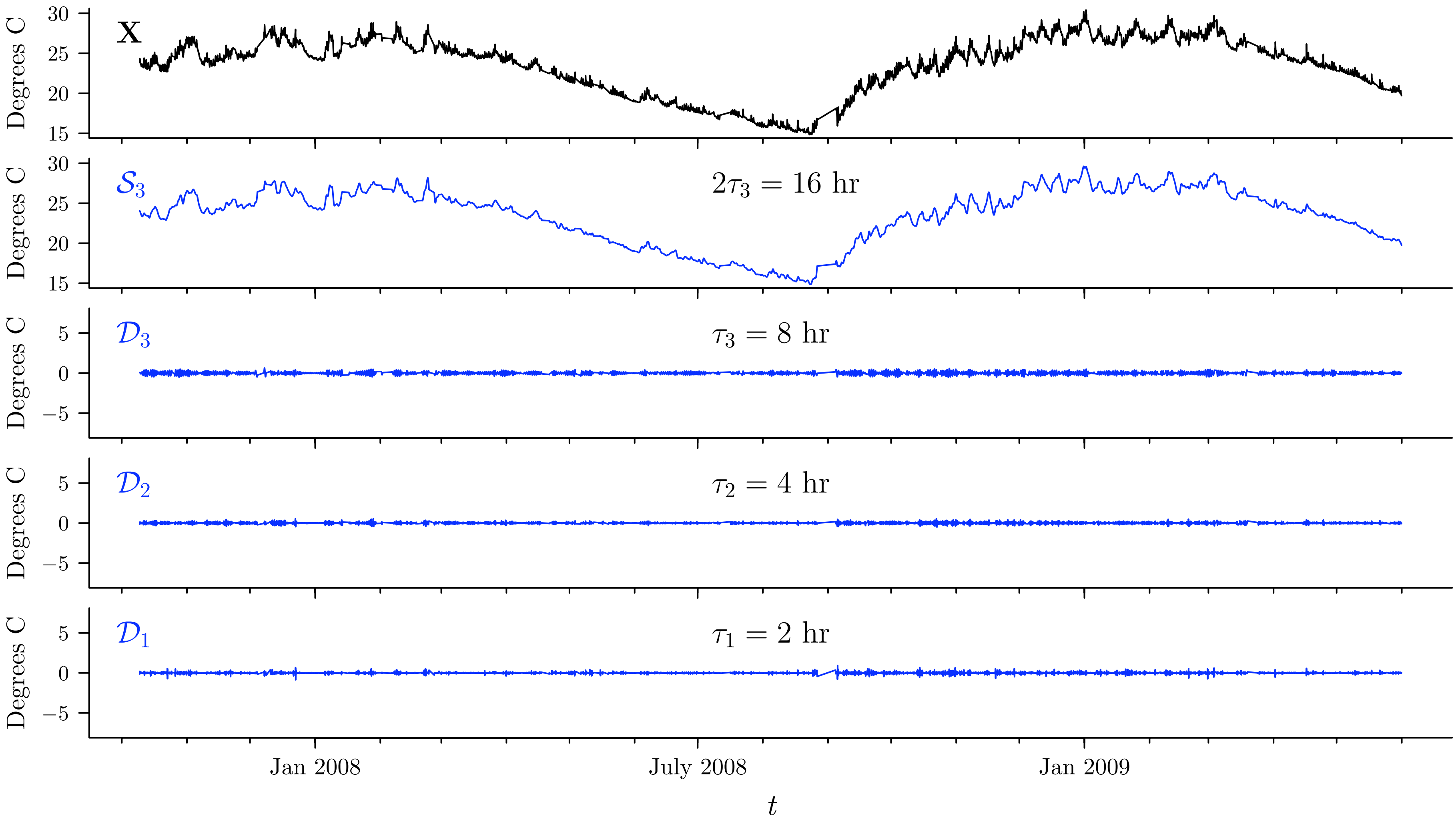
- DWT has equivalents of two 'scary-looking' equations
- analog of first equation is called a multiresolution analysis (MRA):

$$\mathbf{X} = \sum_{j=1}^{J} \mathcal{D}_j + \mathcal{S}_J,$$

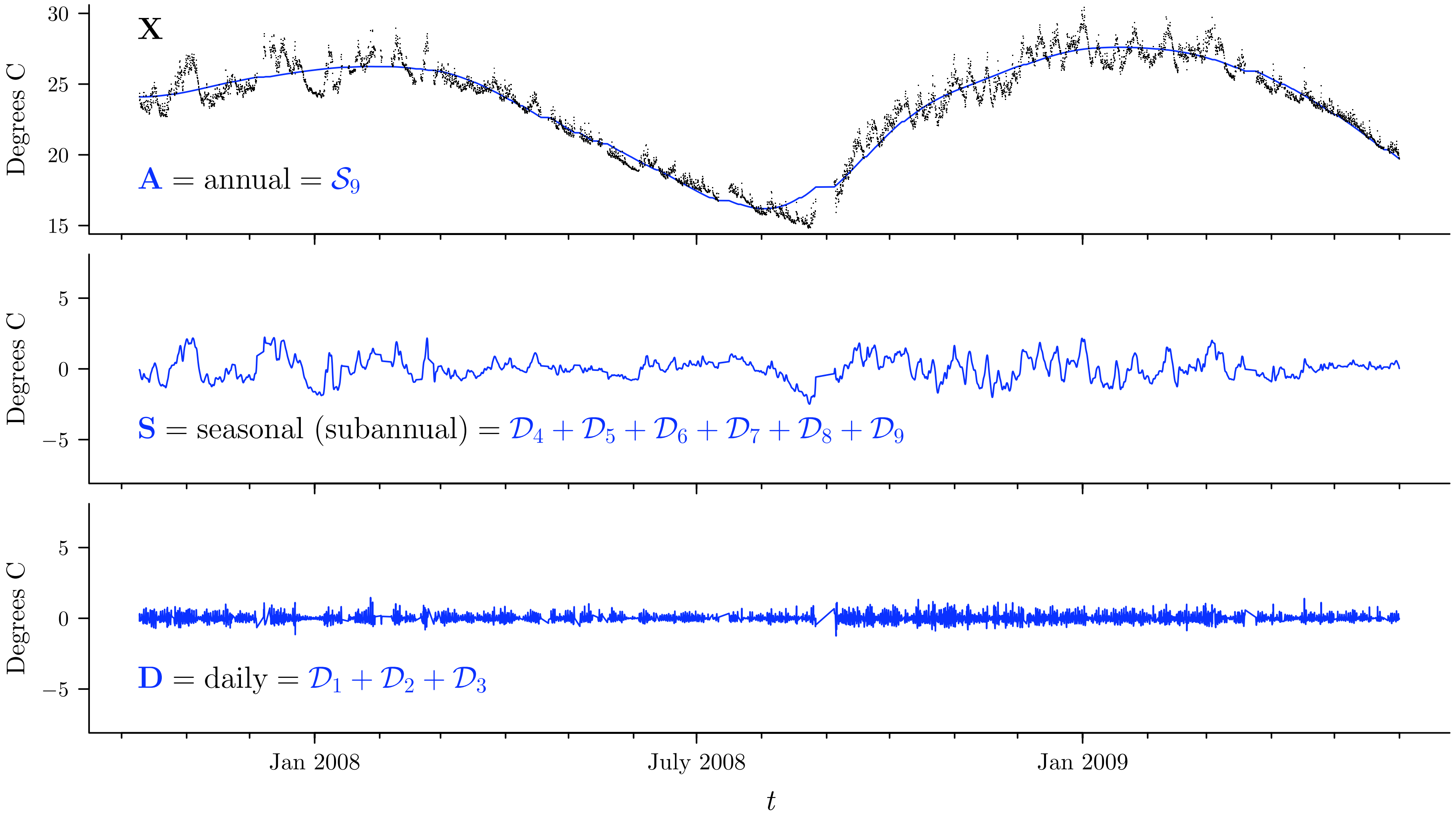
  where

  - $\mathcal{D}_j$ is a 'detail' series depending just on $\mathbf{W}_j$ and capturing part of $\mathbf{X}$ attributable to changes on a scale of τ_j
  - $\mathcal{S}_J$ is a 'smooth' series depending just on $\mathbf{V}_J$ and capturing part of $\mathbf{X}$ attributable to averages on a scale of $2\tau_J$
- since we can recover $\mathbf{X}$ perfectly from its DWT, $\mathbf{W}_1, \ldots, \mathbf{W}_J$ & $\mathbf{V}_J$ are fully equivalent to $\mathbf{X}$

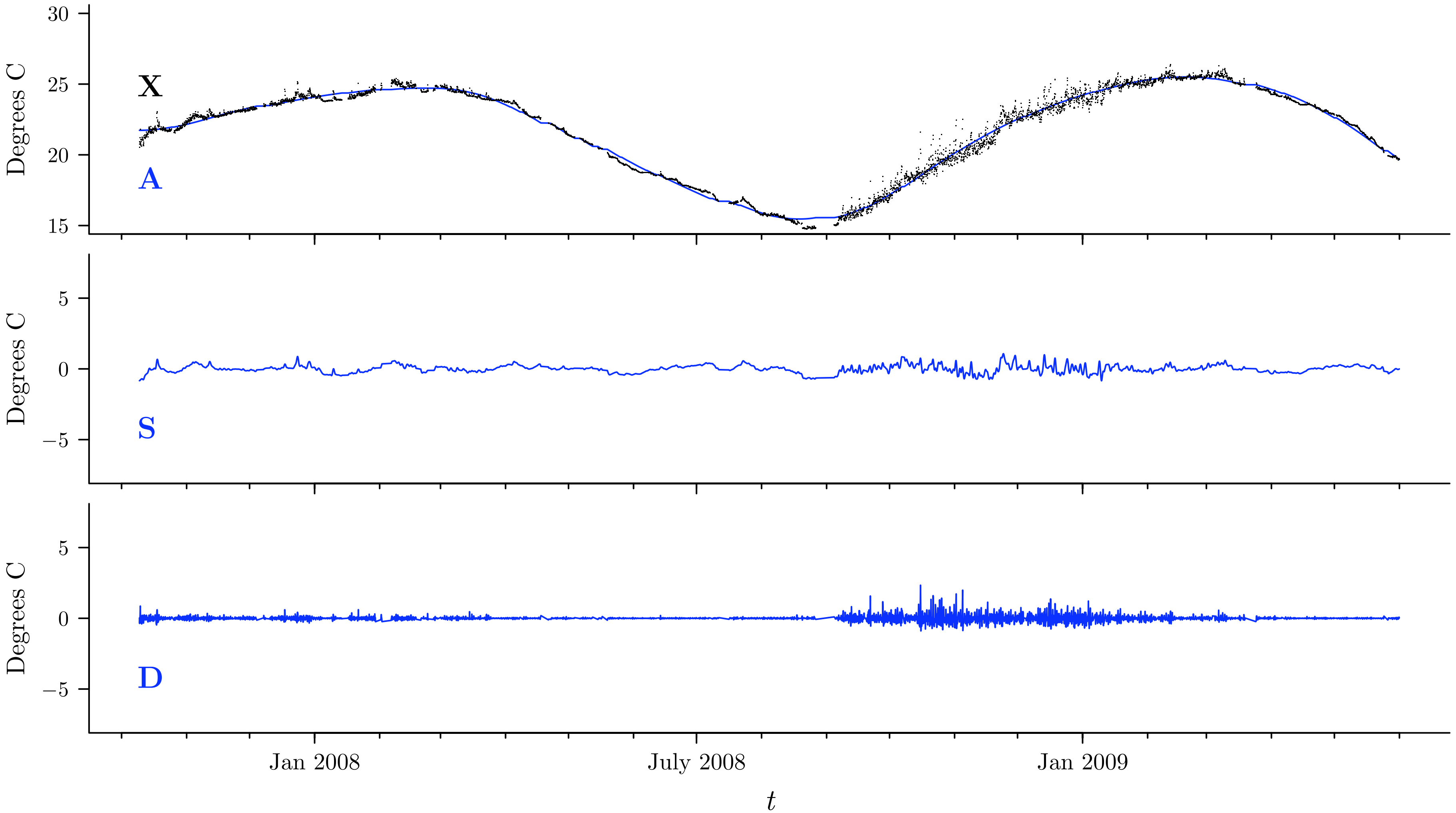
$J = 3$ MRA for 1 m Water Temperatures



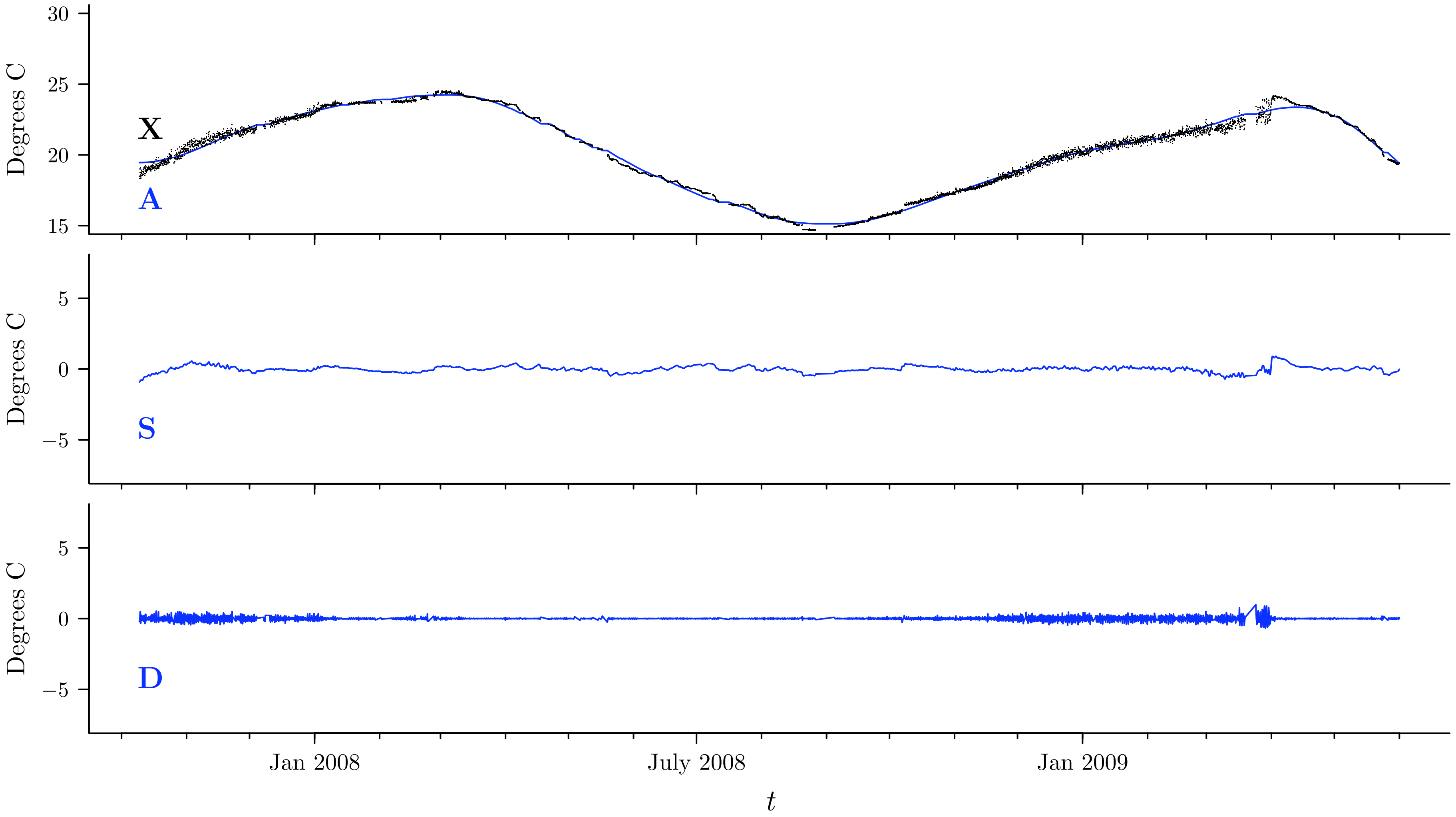
$J = 9$ Modified MRA for 1 m Water Temperatures



$J = 9$ Modified MRA for 10 m Water Temperatures



$J = 9$ Modified MRA for 20 m Water Temperatures



Correlations Between ASD Components at 3 Depths

	A_t 1 m	A_t 10 m	A_t 20 m	S_t 1 m	S_t 10 m	S_t 20 m	D_t 1 m	D_t 10 m	D_t 20 m
A_t, 1 m	1.00								
A_t, 10 m	0.91	1.00							
A_t, 20 m	0.65	0.88	1.00						
S_t, 1 m	0.01	0.00	0.00	1.00					
S_t, 10 m	0.01	0.01	0.01	0.33	1.00				
S_t, 20 m	0.01	0.02	0.04	0.12	0.26	1.00			
D_t, 1 m	0.00	0.00	0.00	0.03	0.00	0.00	1.00		
D_t, 10 m	0.00	0.00	0.00	0.00	0.05	0.00	0.04	1.00	
D_t, 20 m	0.00	0.00	0.00	0.00	0.00	0.03	0.05	−0.07	1.00

Correlations Between ASD Components at 3 Depths

	A_t 1 m	A_t 10 m	A_t 20 m	S_t 1 m	S_t 10 m	S_t 20 m	D_t 1 m	D_t 10 m	D_t 20 m
A_t, 1 m	1.00								
A_t, 10 m	0.91	1.00							
A_t, 20 m	0.65	0.88	1.00						
S_t, 1 m	0.01	0.00	0.00	1.00					
S_t, 10 m	0.01	0.01	0.01	0.33	1.00				
S_t, 20 m	0.01	0.02	0.04	0.12	0.26	1.00			
D_t, 1 m	0.00	0.00	0.00	0.03	0.00	0.00	1.00		
D_t, 10 m	0.00	0.00	0.00	0.00	0.05	0.00	0.04	1.00	
D_t, 20 m	0.00	0.00	0.00	0.00	0.00	0.03	0.05	−0.07	1.00

Correlations Between ASD Components at 3 Depths

	A_t 1 m	A_t 10 m	A_t 20 m	S_t 1 m	S_t 10 m	S_t 20 m	D_t 1 m	D_t 10 m	D_t 20 m
A_t, 1 m	1.00								
A_t, 10 m	0.91	1.00							
A_t, 20 m	0.65	0.88	1.00						
S_t, 1 m	0.01	0.00	0.00	1.00					
S_t, 10 m	0.01	0.01	0.01	0.33	1.00				
S_t, 20 m	0.01	0.02	0.04	0.12	0.26	1.00			
D_t, 1 m	0.00	0.00	0.00	0.03	0.00	0.00	1.00		
D_t, 10 m	0.00	0.00	0.00	0.00	0.05	0.00	0.04	1.00	
D_t, 20 m	0.00	0.00	0.00	0.00	0.00	0.03	0.05	−0.07	1.00

Correlations Between ASD Components at 3 Depths

	A_t 1 m	A_t 10 m	A_t 20 m	S_t 1 m	S_t 10 m	S_t 20 m	D_t 1 m	D_t 10 m	D_t 20 m
A_t, 1 m	1.00								
A_t, 10 m	0.91	1.00							
A_t, 20 m	0.65	0.88	1.00						
S_t, 1 m	0.01	0.00	0.00	1.00					
S_t, 10 m	0.01	0.01	0.01	0.33	1.00				
S_t, 20 m	0.01	0.02	0.04	0.12	0.26	1.00			
D_t, 1 m	0.00	0.00	0.00	0.03	0.00	0.00	1.00		
D_t, 10 m	0.00	0.00	0.00	0.00	0.05	0.00	0.04	1.00	
D_t, 20 m	0.00	0.00	0.00	0.00	0.00	0.03	0.05	−0.07	1.00

Correlations Between ASD Components at 3 Depths

	A_t 1 m	A_t 10 m	A_t 20 m	S_t 1 m	S_t 10 m	S_t 20 m	D_t 1 m	D_t 10 m	D_t 20 m
A_t, 1 m	1.00								
A_t, 10 m	0.91	1.00							
A_t, 20 m	0.65	0.88	1.00						
S_t, 1 m	0.01	0.00	0.00	1.00					
S_t, 10 m	0.01	0.01	0.01	0.33	1.00				
S_t, 20 m	0.01	0.02	0.04	0.12	0.26	1.00			
D_t, 1 m	0.00	0.00	0.00	0.03	0.00	0.00	1.00		
D_t, 10 m	0.00	0.00	0.00	0.00	0.05	0.00	0.04	1.00	
D_t, 20 m	0.00	0.00	0.00	0.00	0.00	0.03	0.05	−0.07	1.00

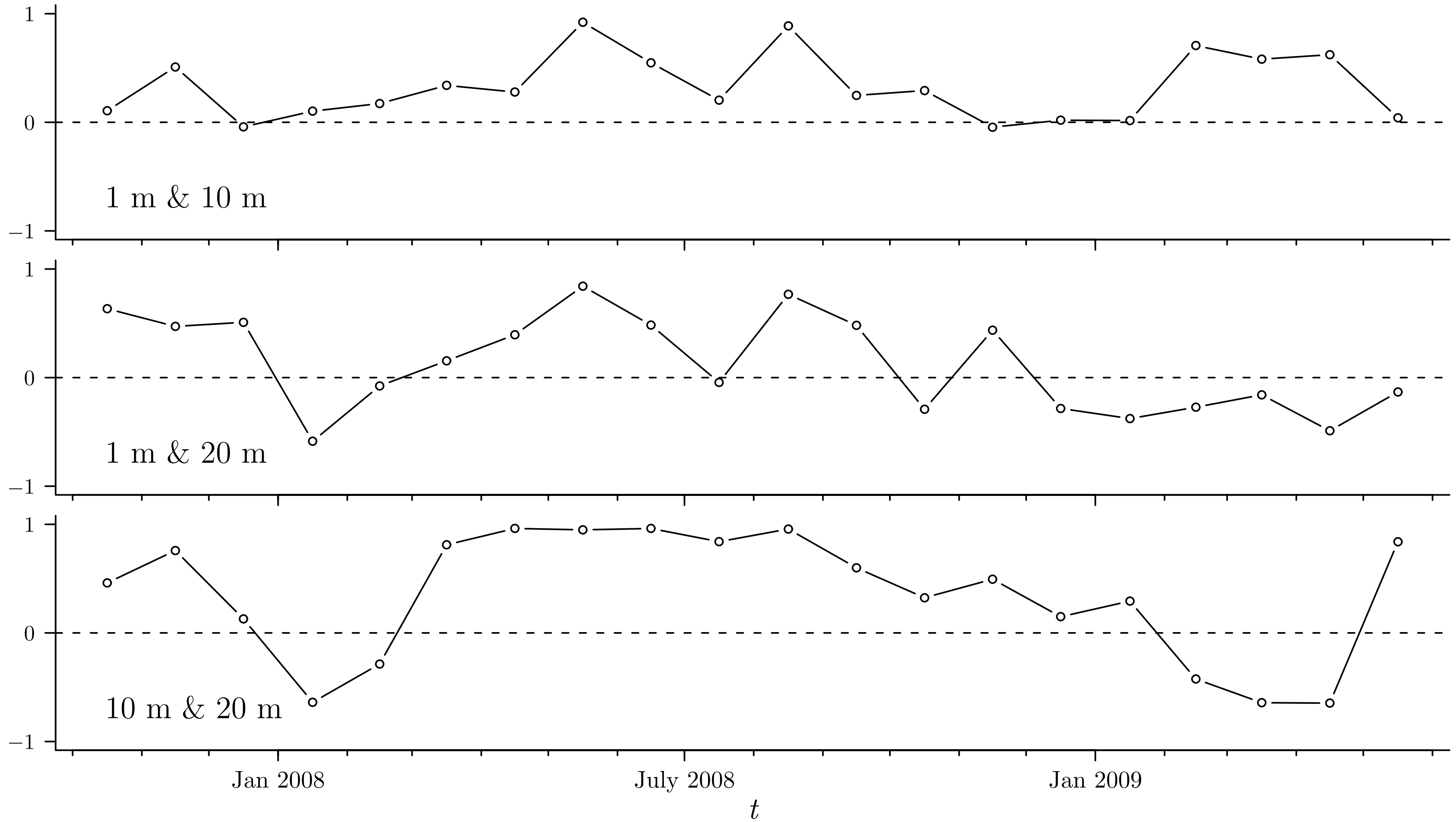
Correlations Between ASD Components at 3 Depths

	A_t 1 m	A_t 10 m	A_t 20 m	S_t 1 m	S_t 10 m	S_t 20 m	D_t 1 m	D_t 10 m	D_t 20 m
A_t, 1 m	1.00								
A_t, 10 m	0.91	1.00							
A_t, 20 m	0.65	0.88	1.00						
S_t, 1 m	0.01	0.00	0.00	1.00					
S_t, 10 m	0.01	0.01	0.01	0.33	1.00				
S_t, 20 m	0.01	0.02	0.04	0.12	0.26	1.00			
D_t, 1 m	0.00	0.00	0.00	0.03	0.00	0.00	1.00		
D_t, 10 m	0.00	0.00	0.00	0.00	0.05	0.00	0.04	1.00	
D_t, 20 m	0.00	0.00	0.00	0.00	0.00	0.03	0.05	−0.07	1.00

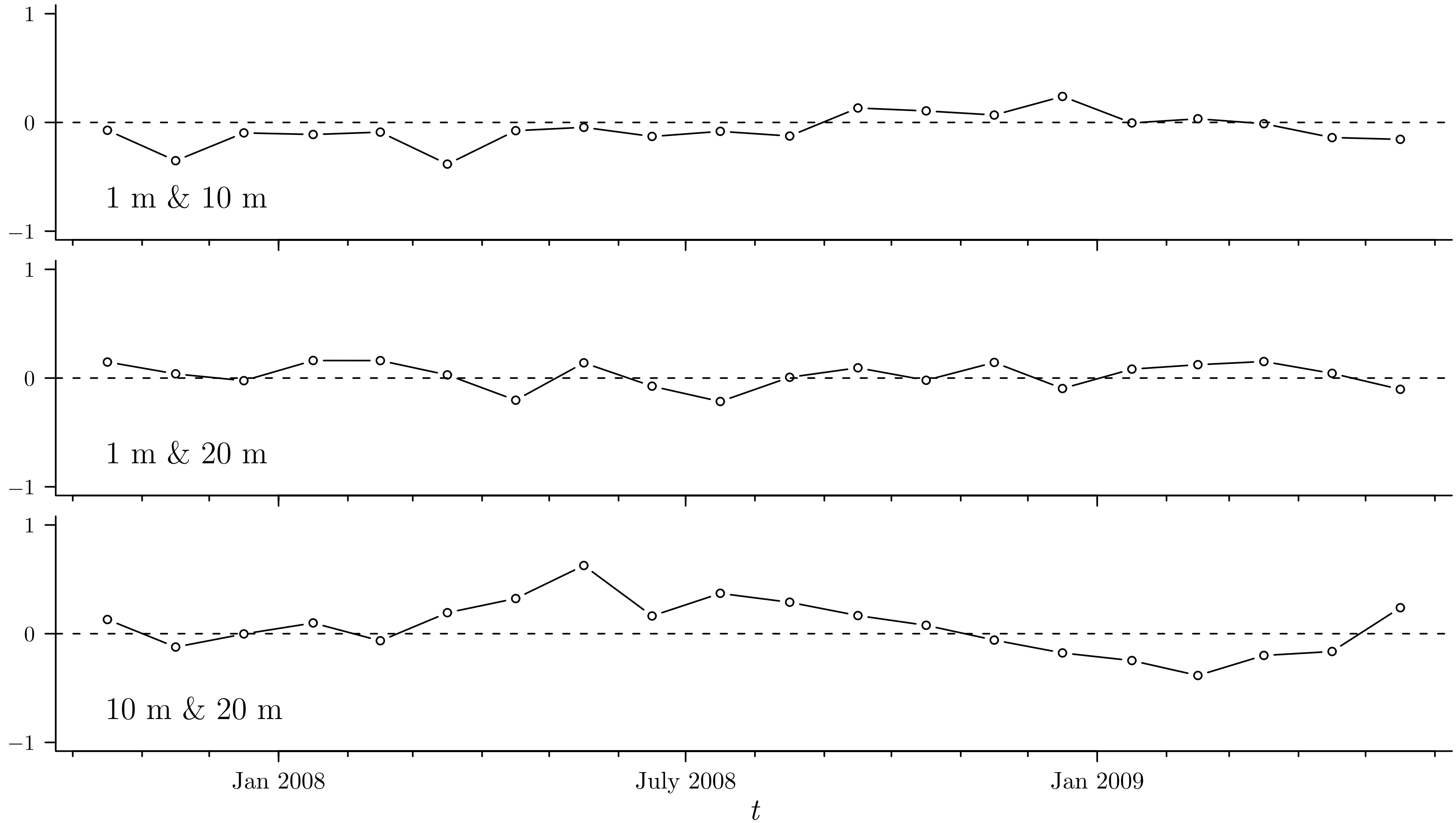
Correlations Between ASD Components at 3 Depths

	A_t 1 m	A_t 10 m	A_t 20 m	S_t 1 m	S_t 10 m	S_t 20 m	D_t 1 m	D_t 10 m	D_t 20 m
A_t, 1 m	1.00								
A_t, 10 m	0.91	1.00							
A_t, 20 m	0.65	0.88	1.00						
S_t, 1 m	0.01	0.00	0.00	1.00					
S_t, 10 m	0.01	0.01	0.01	0.33	1.00				
S_t, 20 m	0.01	0.02	0.04	0.12	0.26	1.00			
D_t, 1 m	0.00	0.00	0.00	0.03	0.00	0.00	1.00		
D_t, 10 m	0.00	0.00	0.00	0.00	0.05	0.00	0.04	1.00	
D_t, 20 m	0.00	0.00	0.00	0.00	0.00	0.03	0.05	−0.07	1.00

Time-Varying Seasonal Correlations at 3 Depths



Time-Varying Daily Correlations at 3 Depths

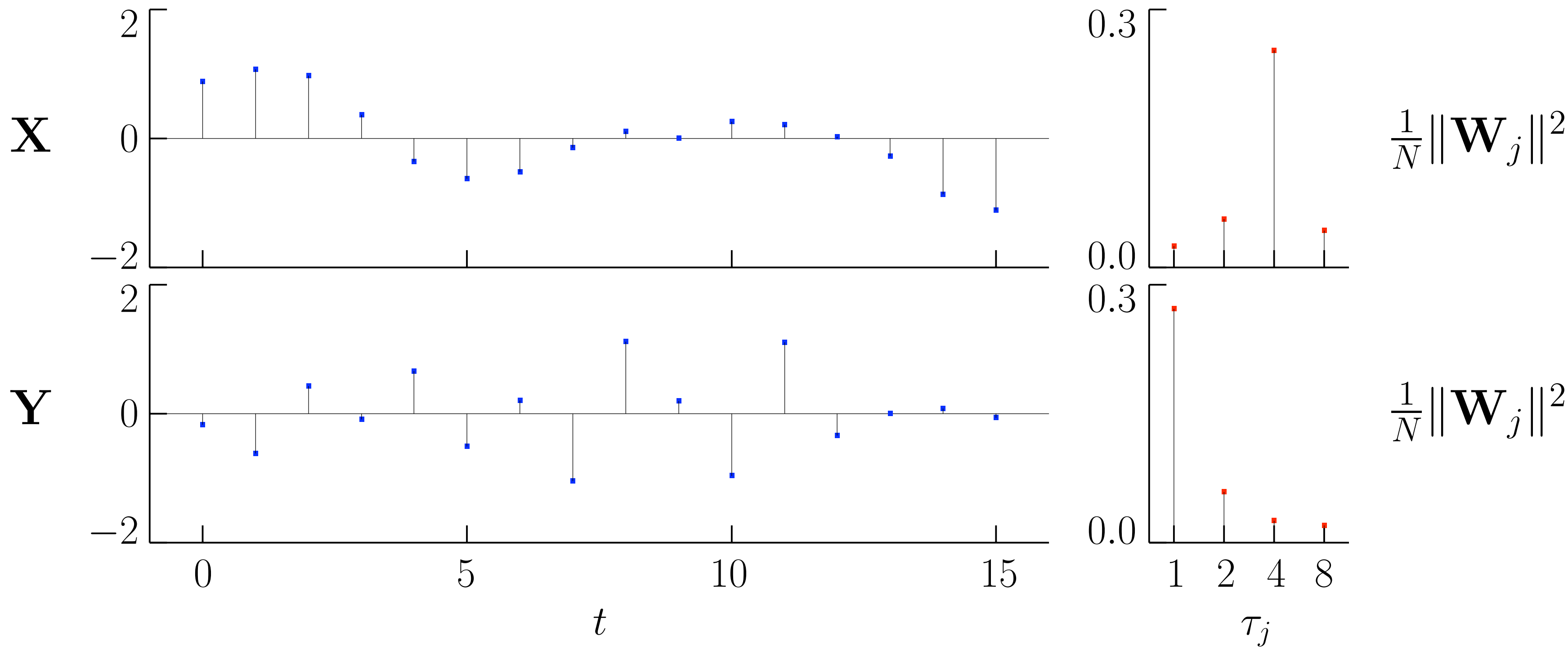


Wavelet Variance: I

- analog of second 'scary-looking' equation is
$$\|\mathbf{X}\|^2 \equiv \sum_{t=0}^{N-1} x_t^2 = \sum_{j=1}^{J} \|\mathbf{W}_j\|^2 + \|\mathbf{V}_J\|^2;$$
where $\|\mathbf{X}\|$ is the Euclidean norm of $\mathbf{X}$
- 'energy' preservation leads to analysis of sample variance:
$$\hat{\sigma}_x^2 \equiv \frac{1}{N} \sum_{t=0}^{N-1} (x_t - \bar{x})^2 = \frac{1}{N} \left(\sum_{j=1}^{J} \|\mathbf{W}_j\|^2 + \|\mathbf{V}_J\|^2 \right) - \bar{x}^2,$$
where $\bar{x} \equiv \sum_t x_t / N$
- $\frac{1}{N}\|\mathbf{W}_j\|^2$ called wavelet variance (or spectrum) and is portion of $\hat{\sigma}_x^2$ due to changes in averages over scale τ_j, thus providing a 'scale-by-scale' analysis of variance

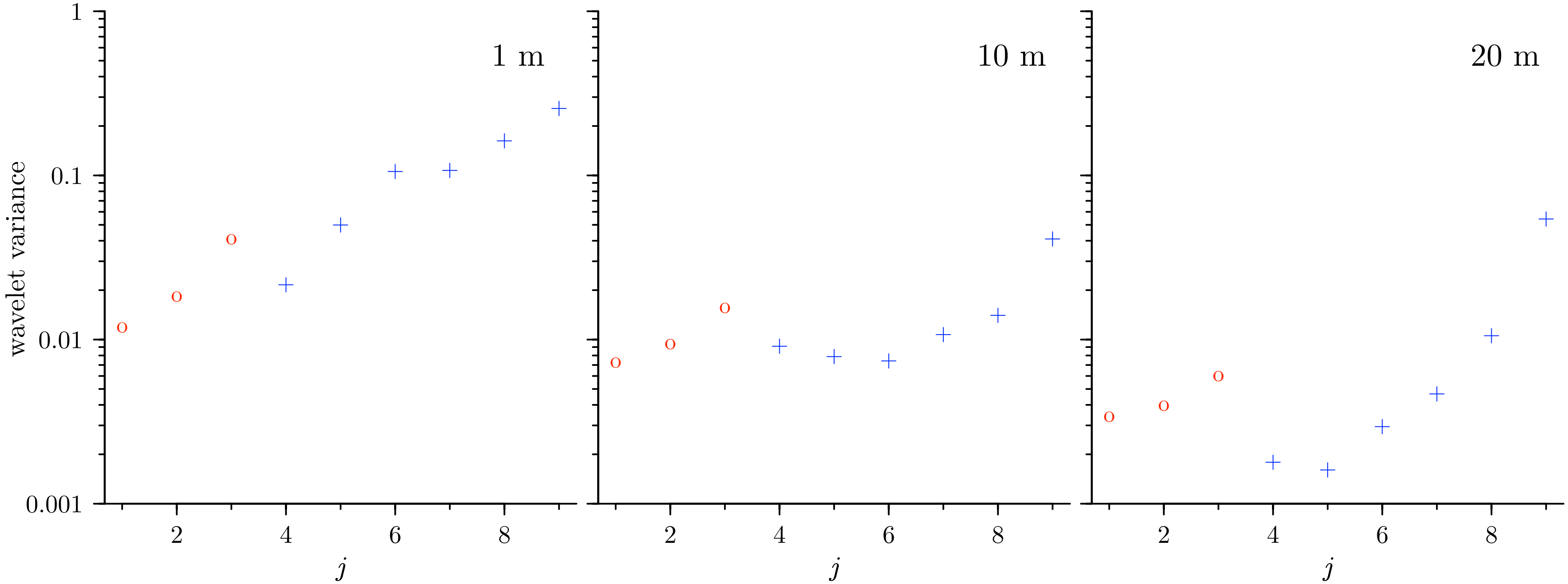
Wavelet Variance: II

- wavelet variances for time series $\mathbf{X}$ and $\mathbf{Y}$ of length $N = 16$, each with zero sample mean and same sample variance

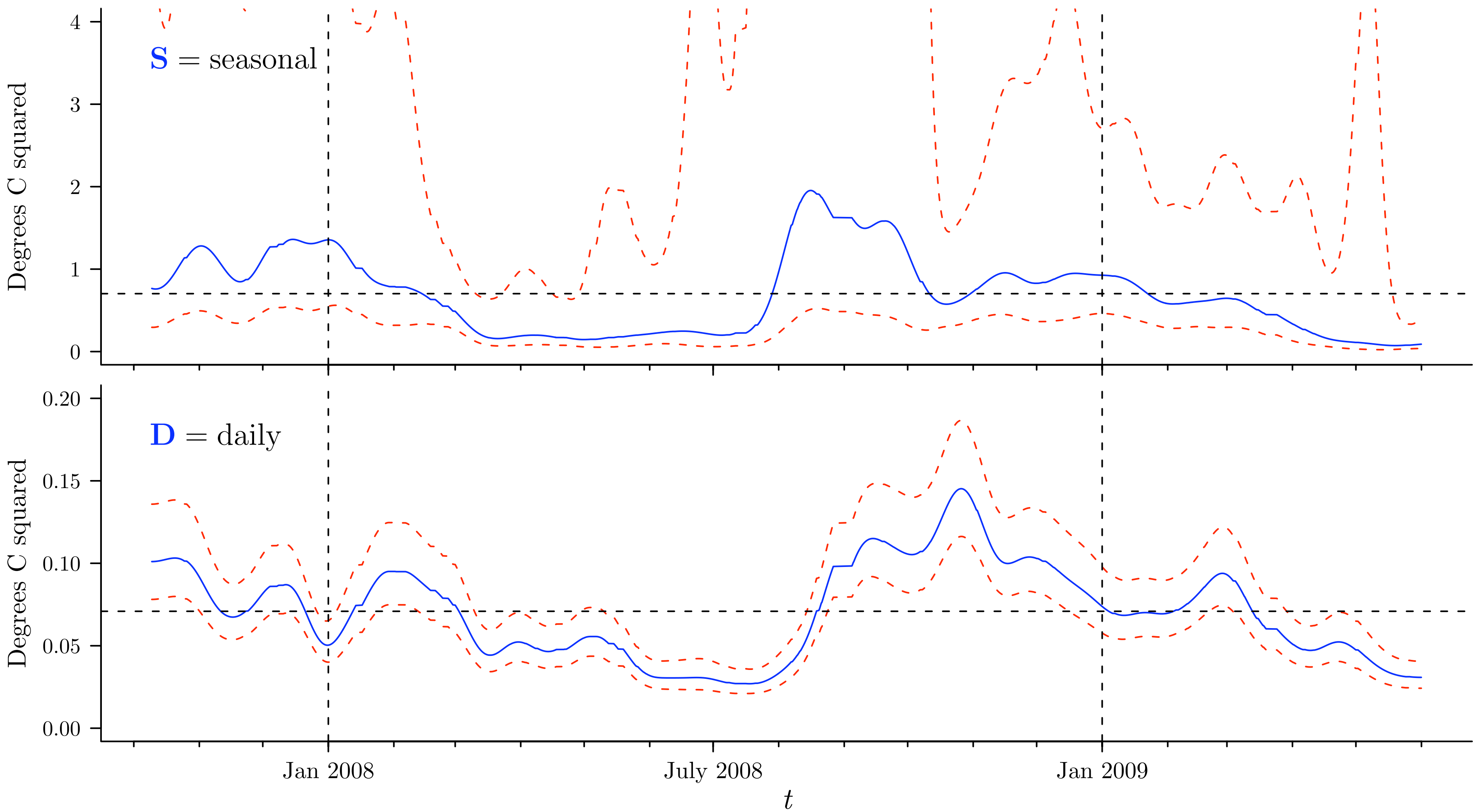


Wavelet Variance Estimates for Water Tempertures

- variance associated with daily component **D** is sum of circles
- variance associated with seasonal component **S** is sum of pluses



Time-Varying Wavelet Variance Estimates, 1 m Data



Concluding Remarks: I

- wavelets decompose time series with respect to two variables:
  - time (location)
  - scale (extent)
- CWT & DWT have two fundamental properties:
  1. fully equivalent to original time series
  2. energy (variance) of time series is preserved
- wavelet variance gives scale-based analysis of variance (natural match for many geophysical processes)
- techniques extend naturally to images

Concluding Remarks: II

- *many* other uses for wavelets (barely scratched the surface!)
  - approximately decorrelate certain time series (ODWT needed)
  - assessing sampling properties of statistics (ODWT or MODWT)
  - signal extraction ('wavelet shrinkage'; ODWT or MODWT)
  - edge identification in images (CWT best)
  - compression of time series/images (ODWT needed)
  - fast simulation of time series/images (ODWT needed)

Thanks to . . .

- conference organizers for opportunity to talk
- numerous folks at CSIRO who made my visit possible (and pleasureable!)
- Seqwater for opportunity to analyze their data