

Climate and Land Use Change Earth Resources Observation and Science (EROS) Center



Brian Davis¹, Song Guo¹, Kelcy Smith¹
January 5, 2017

¹ SGT, Inc., Contractor to the USGS EROS Center. 47914 252nd Street, Sioux Falls, SD 57198. Work performed under USGS contract G15PC00012.

Agenda

- Background
- ARD
- Basic Terminology
- Standard Modeling Components
 - Variogram (Madogram)
 - LASSO Regression
 - Robust Regression
 - Tmask
- Walk Through Change Detection Processing Steps
- Describe Output Results

Background, Context

- Provide understanding of Continuous Change Detection (CCD) algorithm as implemented to aid evaluation, analysis, recommendations for future updates
- Corresponds to Matlab Version 12.30 used for processing Chesapeake Bay
- Implemented in python
 - Open source, eliminating licensing restrictions and cost
 - Platform-independent
 - Information Warehouse + Data Store (IW + DS) interaction
 - Science Execution Environment (SEE) processing
 - python and associated tools, modules, common to HPC Systems
 - Increase modularity, decrease complexity
- Enable sending results to Information Warehouse for subsequent interaction
- Future ARD access through API, eliminating file-system data management:

“Answers, Not Data”

ARD

- Pixel Alignment
- Consistent Projection
- Standardized Extents (tiles)
- Standard Level 2 Products Suite
- Eliminates Swath Overlap
- Adds Side-lap, (Temporal Density)
- User-defined AOI
- ARD Named Future Standard Landsat Product

Landsat WRS-2 Side-lap*

Table 5.1 Image Sidelap of Adjacent Swaths

Latitude (degrees)	Image Sidelap (%)
0	7.3
10	8.7
20	12.9
30	19.7
40	29.0
50	40.4
60	53.6
70	68.3
80	83.9

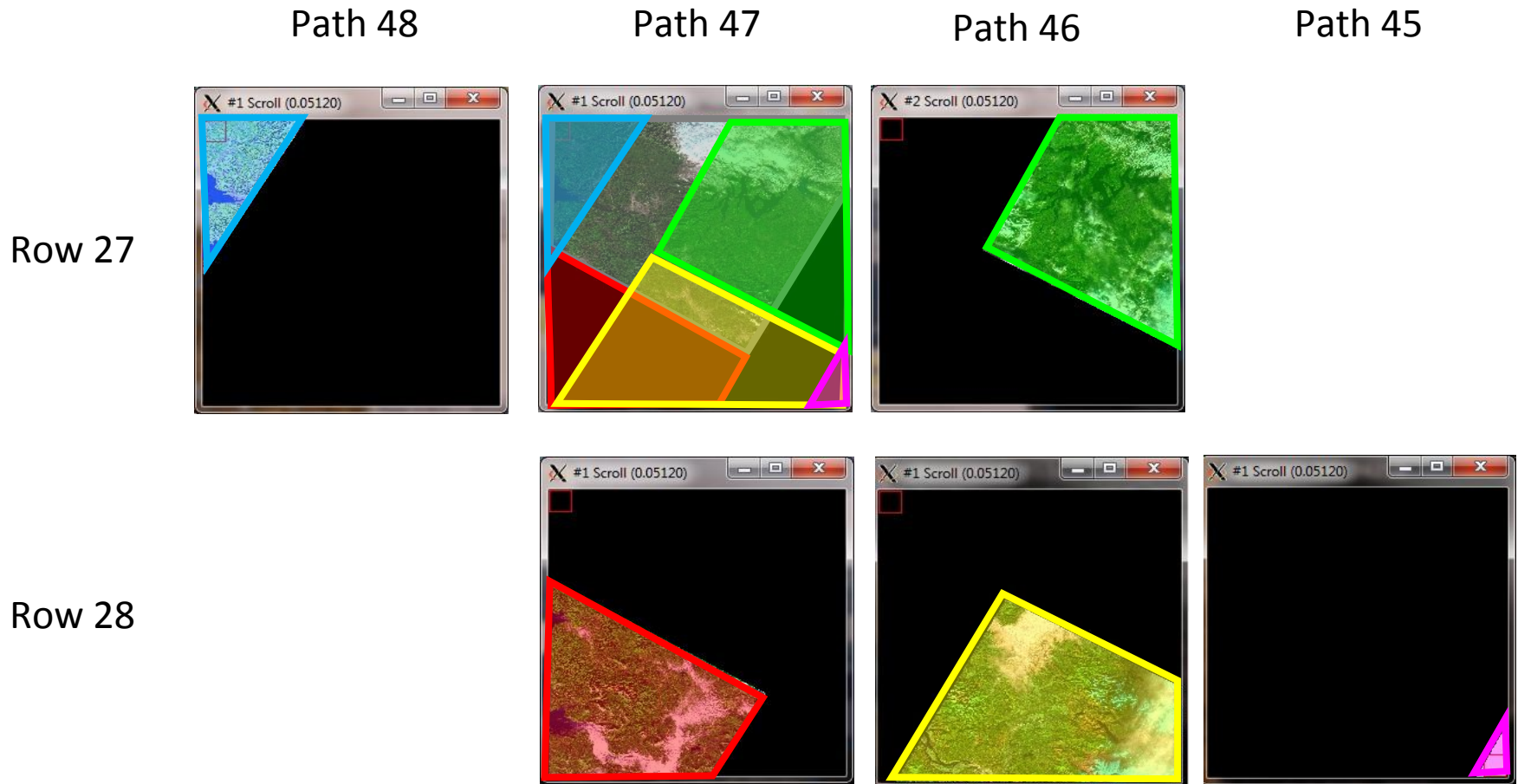
CONUS:

TX/FL $\approx 24^{\circ}$ N $\approx 15.62\%$
 WA/ME $\approx 36^{\circ}$ N $\approx 20.28\%$
 Avg. $\approx 30^{\circ}$ N $\approx 19.70\%$

* From the Landsat 7 Data Users Handbook

Input – Analysis Ready Data (ARD)

WA grid07 Example, 1982 – 2015 TM, ETM+, OLI/TIRS



Terminology

Model	regression fit of points to a curve
Variogram	Variogram is a description of the spatial continuity of the data, usually the squares of the differences. Madogram is instead the abs of the diffs, and the “distance” between points is ignored.
MEOW	Minimum Expected Observation Window (number of coefficients * 3)
Tmask	time-series mask, outliers determined by regression fit using green (SWIR1) band
Window	current sliding array of X # number of observations (model window, peek window, fit window)
Fit Window	observations that are used to fit a regression model to, subset of the model window
Model Window	observations that are currently represented by a fitted regression model
Peek Window	observations that are currently being analyzed / under statistical scrutiny
Stable Model	generalized fit 4 coefs (min) from LASSO Regression
Detection Bands	subset of spectral bands used to determine outliers and detect change: red, green, NIR, SWIR1, SWIR2
LASSO Regression	Initial first attempt at fitting points to curve, and subsequent start detection, change detection
Robust Regression	more general purpose curve fitting for identifying outliers
Residual	predicted value from a curve fit - actual value, used as a measure of error
Persistent Processing Mask	Application of Tmask, cfmask, etc., to maks, not eliminate values from input arrays

Current CCD Operational Parameters

Change Probability Threshold

Inverse of χ^2 (0.99, number of detection bands)

```
In [6]: from scipy.stats import chi2  
chi2.ppf(0.99, 5)
```

```
Out[6]: 15.086272469388987
```

Outlier Threshold

inverse of χ^2 (**$1-1e-6$** , number of detection bands)

```
In [5]: from scipy.stats import chi2  
chi2.ppf(1-1e-6, 5)
```

```
Out[5]: 35.888186879610423
```

Minimum Number of Consecutive Observations required to identify a change

6

Maximum Number of Model Coefficients Produced

8 (insert FFT graphic)

Clear count threshold percent = 0.25

Snow count threshold percent = 0.75

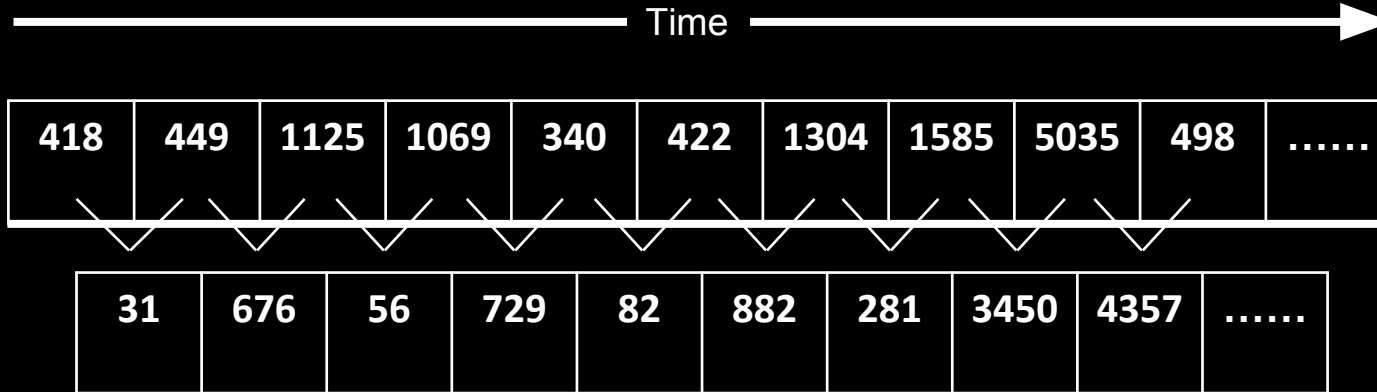
LASSO Regression Iteration limit = 20

CCD Input Landsat Data

Spectral Band	LT4, LT5, LE7	LC8
	Surface Reflectance (SR) Band #	
Red	1	2
Green	2	3
Blue	3	4
Visible Near InfraRed (NIR)	4	5
Short Wave InfraRed (SWIR) 1	5	6
Short Wave InfraRed (SWIR) 2	7	7
	Brightness Temperature Band # (Top Of Atmosphere (TOA) Band)	
Thermal	6	10
	CFMASK Band	
MASK	cfmask	cfmask

Variogram (Median Madogram) Example

Example Red Band DN Values



Absolute value of differences of consecutive pixels

Median of differences: 676

LASSO Regression

$$C + c_1 \text{date} + c_2 \sin\left(2\pi * \frac{\text{date}}{365.25}\right) + c_3 \cos\left(2\pi * \frac{\text{date}}{365.25}\right) + c_4 \sin\left(2 * 2\pi * \frac{\text{date}}{365.25}\right) + c_5 \cos\left(2 * 2\pi * \frac{\text{date}}{365.25}\right) + c_6 \sin\left(3 * 2\pi * \frac{\text{date}}{365.25}\right) + c_7 \cos\left(3 * 2\pi * \frac{\text{date}}{365.25}\right)$$

C constant or intercept

C₁ slope or trend line

C₂ thru C₇ sinusoid amplitudes

4 Coefficients defined as (referred to as a generalized fit):

$$C + C_1 + C_2 + C_3$$

6 Coefficients adds:

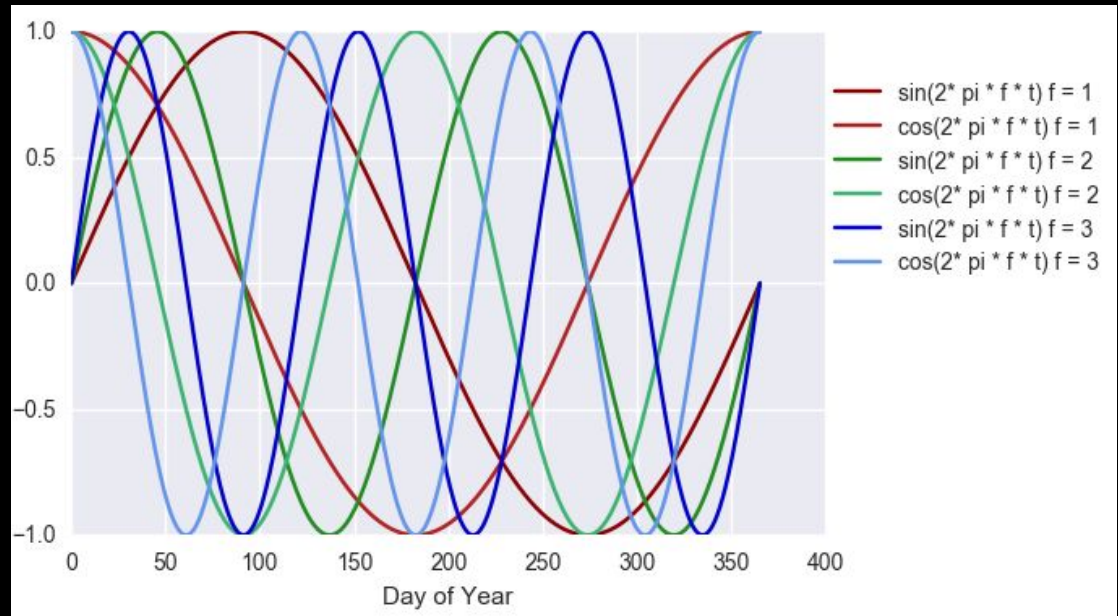
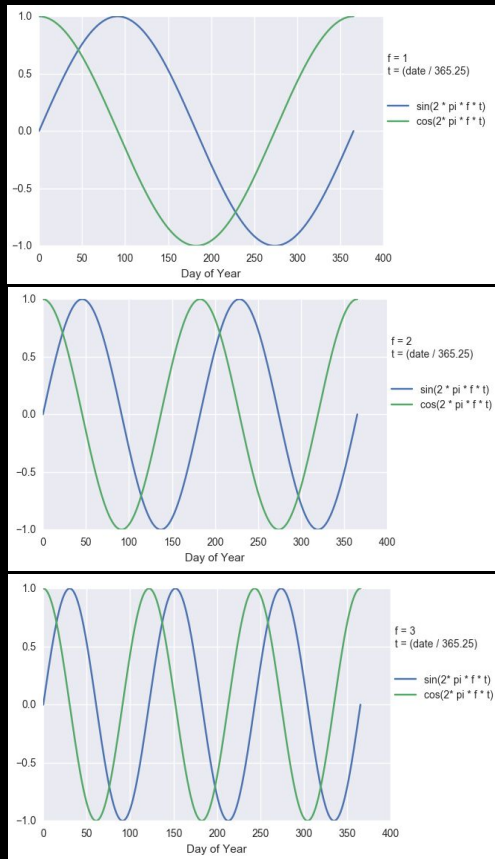
$$C_4 + C_5$$

8 further adds:

$$C_6 + C_7$$

Minimum number of observations required for a fit is defined as:
number of coefficients (4,6,8) * NUM_OBS_FACTOR (3)

Sine Wave Visualization



Robust Regression

(Ordinary Least Squares)

$$C + C_1 \text{date} + C_2 \cos\left(2\pi * \frac{\text{date}}{365.25}\right) + C_3 \sin\left(2\pi * \frac{\text{date}}{365.25}\right) + C_4 \cos\left(2\pi * \frac{\text{date}}{365.25} * \frac{1}{\text{num yrs}}\right) + C_5 \sin\left(2\pi * \frac{\text{date}}{365.25} * \frac{1}{\text{num yrs}}\right)$$

Number of years refers to the span of years represented by the model window, rounded up

Tmask

Use the Robust Regression method to fit curves for the green and SWIR1 bands, for a given observation window.

If any residual is $>$ the Tmask Threshold
($4.89 * \text{variogram value}$),
mark the observation as an outlier and update
the Persistent Processing Mask

main

If the percentage of clear pixels < clear threshold:

 If the percentage of snow > snow threshold:

Persistent Snow Procedure

 else:

Insufficient Clear Procedure

else:

Standard Procedure

Persistent Snow Procedure

Include observations flagged as snow by the QA

Total observations = snow observations + clear observations

Use only acceptable values (unsaturated)

if num (observations) > meow size threshold (default = 12):

for band in bands:

if thermal band:

do a generalized curve fit for entire thermal set

else:

if number of observations < meow size threshold (default=12)

set intercept value to constant value

set coefficients to zero

else:

do a generalized curve fit for observation set

Insufficient Clear Procedure

Use the insufficient clear observations

Using the green band, update Persistent Processing Mask to exclude observations exceeding:

$\text{green_band_median} + 400$

Do a generalized curve fit for observation set

Initial Change Detection Setup

Mask out duplicate observations using first of two swath overlap observations

Sort temporally

Convert thermal values from degrees Kelvin to Celsius:

$$\text{value} = \text{value} * 10 - 27315$$

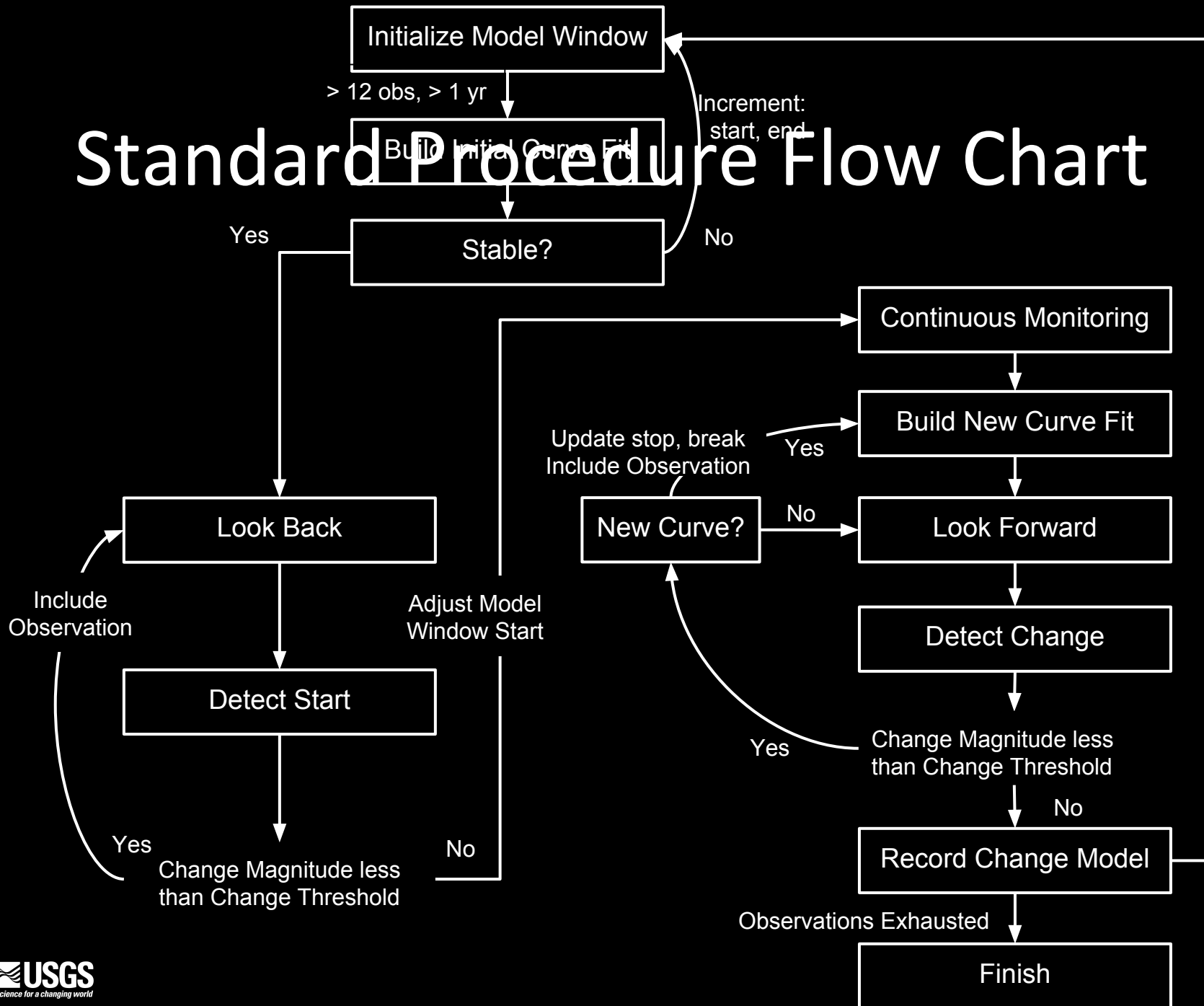
Create mask(s) of values that fall outside of acceptable ranges, considered saturated if > max:

Reflectance acceptable range: 0 -> 10,000

Thermal acceptable range: -9,320 -> 7,070

Create Variogram (madogram) for each band

Standard Procedure Flow Chart





Initialize Model Window

> 12 obs, > 1 yr

Build Initial Curve Fit

Stable?

Yes

Increment:
start, end

No

Continuous Monitoring

Build New Curve Fit

Update stop, break
Include Observation

Yes

New Curve?

No

Look Forward

Detect Change

Change Magnitude less
than Change Threshold

No

Record Change Model

Observations Exhausted

Finish

Look Back

Include
Observation

Detect Start

Adjust Model
Window Start

Yes

Change Magnitude less
than Change Threshold

No

Initialize Model Window

initialize first model window

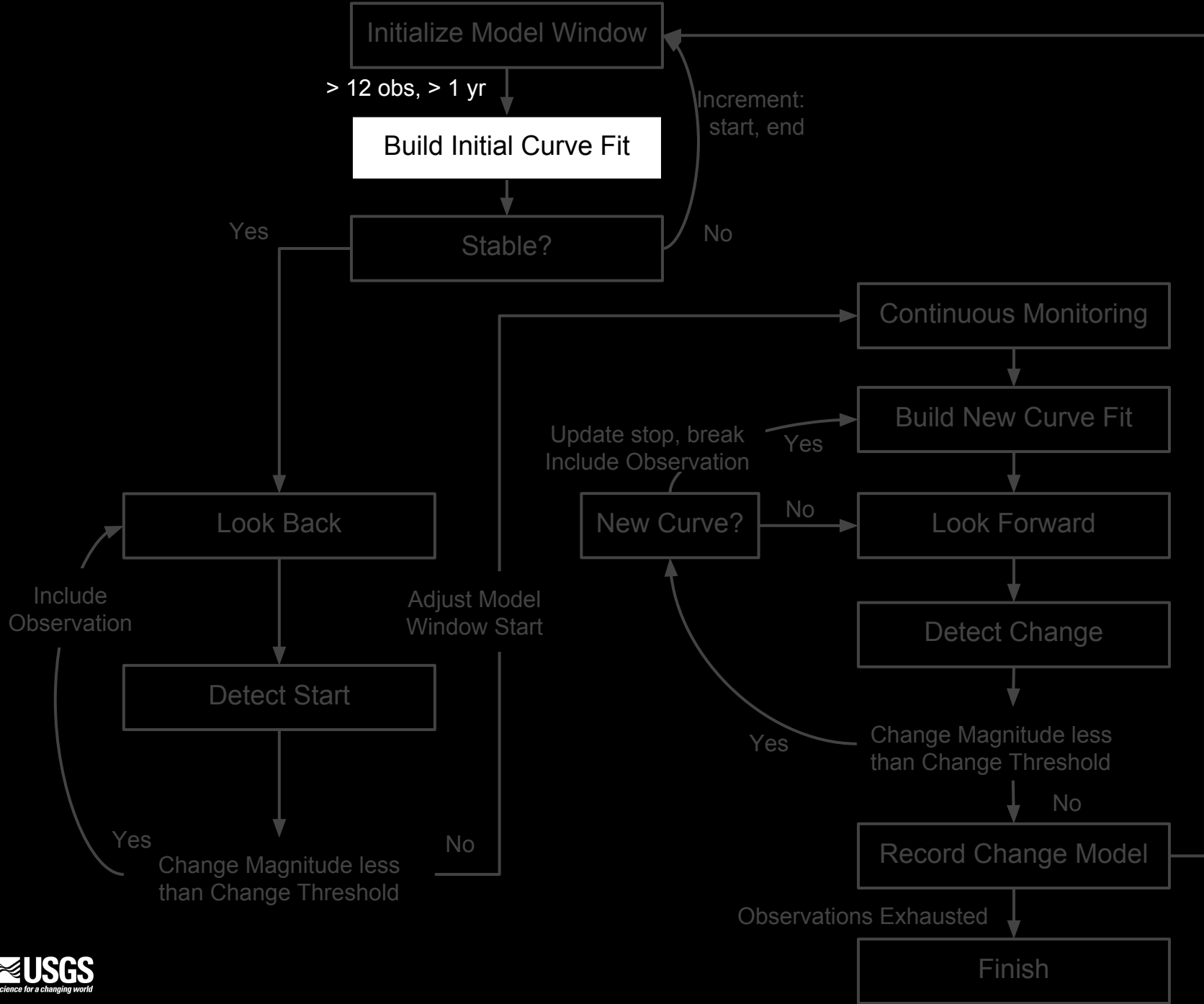
(start is set to previous break point; window size set to contain 12 observations)

while there are clear observations:

mask outliers in current window as defined by tmask

if not enough observations (12) and minimum temporal span (1 year) in the window:

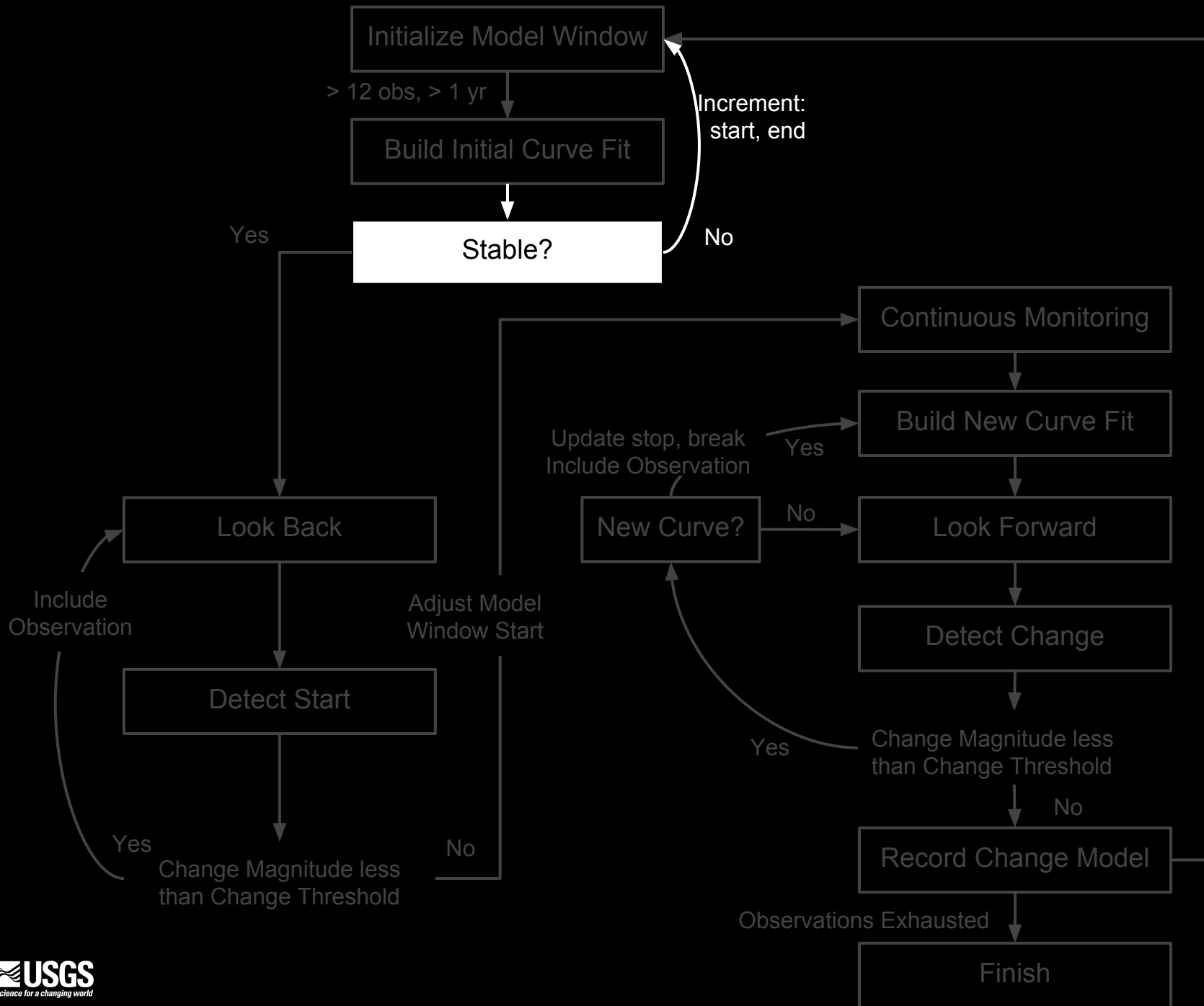
increase window size by 1 continue to the next observation



Build Initial Curve Fit

perform LASSO Regression using 4-coefficient model

Inc
Obse



Stable?

Stable?

```
for each of the detection bands ( r, g, n, s1, s2)
  model slope coef * diff ( end – start dates )
  check_vals = ( |slope| + |first model residual| + |last model residual| ) / MAX
    (where MAX = max (variogram, model RMSE) )
  if summation of ( check_vals )2 < change threshold:
    stable
  else:
    increment model window start and end by 1
```

Stable? (example)

Stable?

4614 514 602 484 549 2673 731 2662 954 1329 1827 5102
Observations (Green Band)

Fitted Model (Green Band)

$C = -44782.78$

$C_1 = 0.062$ (slope coefficient)

$C_2 = 194.19$

$C_3 = 0$

Model Window = start: 724785 end: 725777

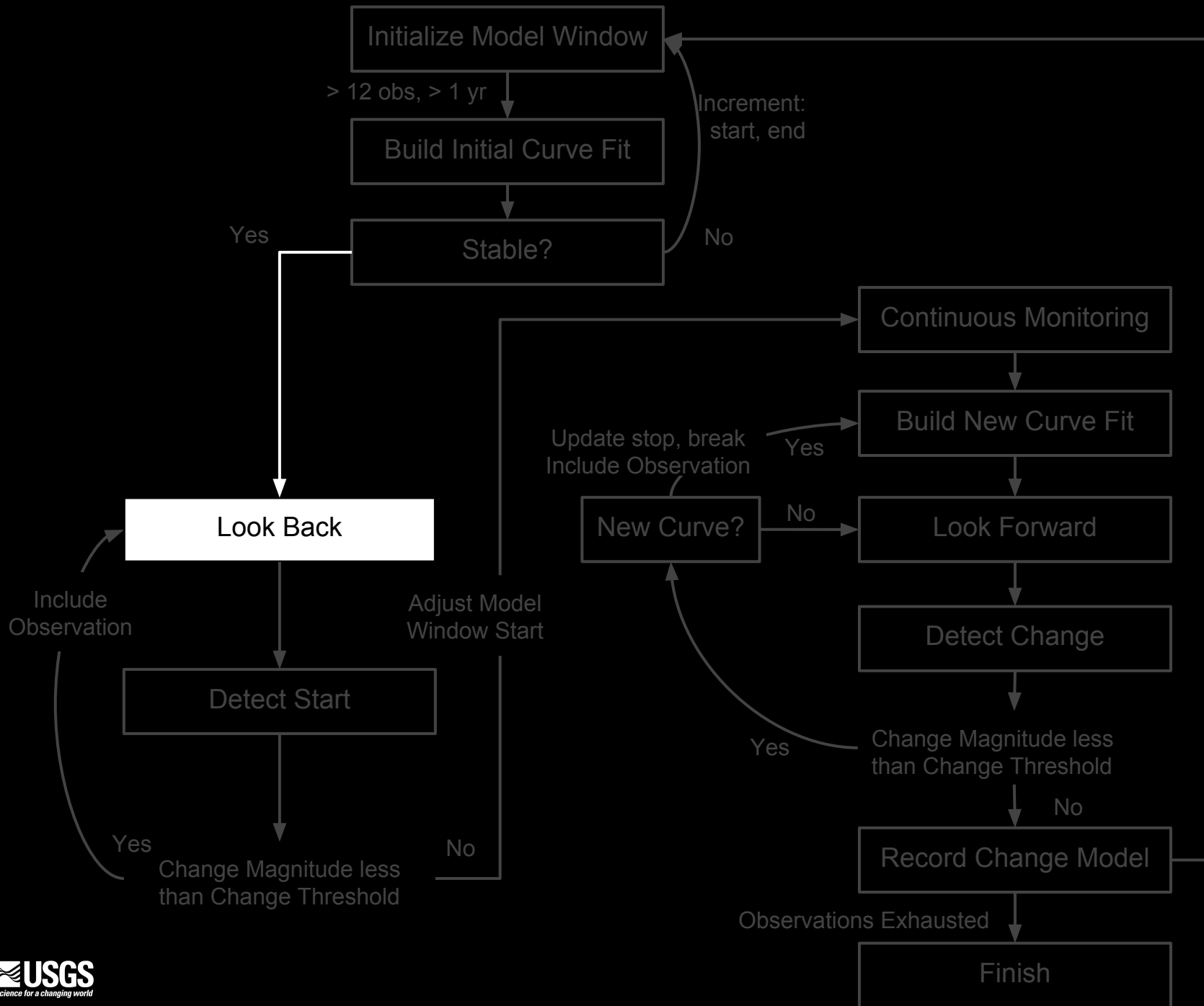
First Residual = 200 Last Residual = -56

Green check value = $(|0.062 * (725777 - 724785)| + |200| + |-56|) / 32.56 = 9.751351351351351$

Red = 10.56 NIR = 15.23 SWIR1 = 18.78 SWIR2 = 37.96

Check Value = $9.75^2 + 10.56^2 + 15.23^2 + 18.78^2 + 37.96^2 = 1440.96$

Finish



Look Back

for each observation preceding current model window start
and end of previous model window:

 If change magnitude $<$ change threshold:

 add observation to current model window

 else:

 proceed with continuous monitoring

note: when look-back is complete, if this is the very first model window, and
there 6 or more unused observations, do a general curve fit on them

Look Back

(adjusting the model start)

Time (Julian days)

day (i_{start})

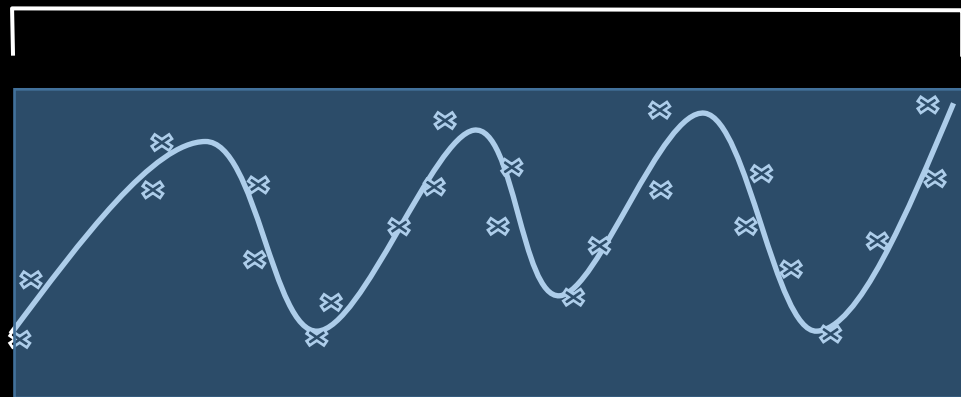
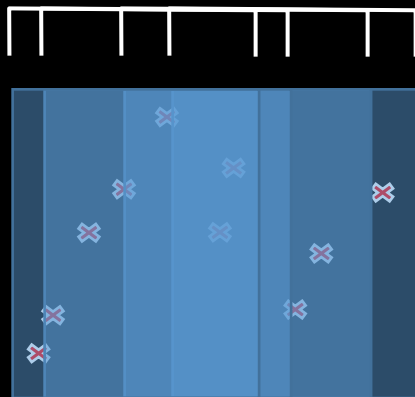
day (j)

day ($j + T$)

Peek Window / Look Back

Model Window

Inc
Obs



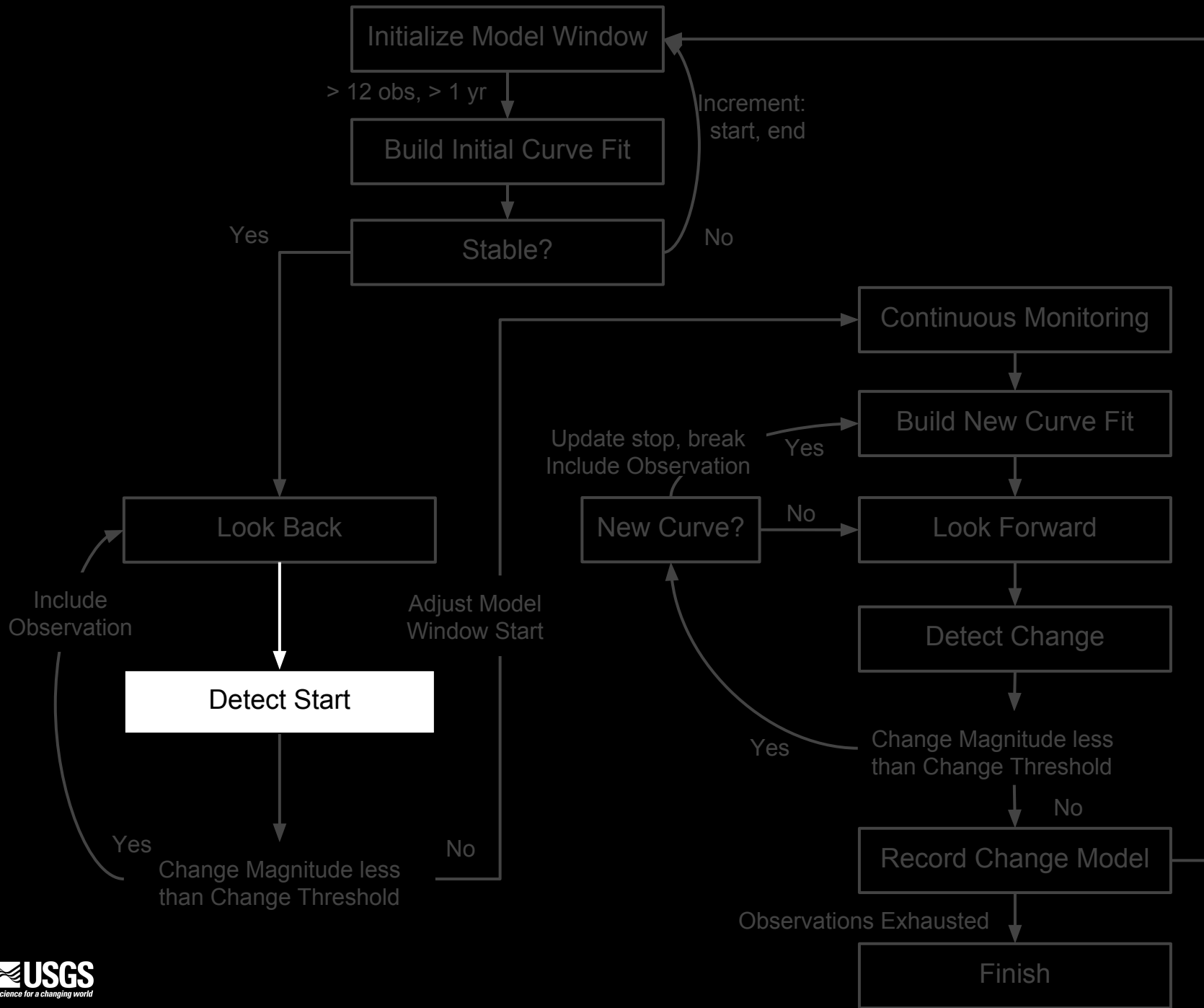
clear_x (i_{start})

clear_x (i)

clear_x ($i + N$)

Observations

Finish



Detect Start

using the observations in the Peek Window, calculate a change magnitude per observation:

for each of the detection bands (r, g, n, s1, s2):

 difference magnitude = residuals / MAX

 (where MAX = max (variogram, model RMSE))

change magnitude = sum (difference magnitude ²)

if the minimum (change magnitude) > change threshold:

 change detected

if the first change magnitude > outlier threshold:

 update Persistent Processing Mask

else:

 include the observation in the Model Window

Detect Start (Calculate Change Magnitudes)

Peek Window

Observations

Green	793	652	803	968	1514	840	361
Red	853	745	1034	1305	1662	1052	383
NIR	2389	2168	2217	2322	2240	1602	610
SWIR1	2779	2622	2974	3195	1786	2740	545
SWIR2	1708	1595	1770	2134	1157	1994	350

Predictions

Green	812	662	823	979	1530	857	373
Red	868	760	1044	1312	1681	1058	398
NIR	2396	2173	2222	2336	2250	1612	622
SWIR1	2796	2629	2981	3209	1805	2752	554
SWIR2	1714	1600	1780	2148	1162	2014	362

Residuals

Green	19	10	20	11	16	17	12
Red	15	15	10	7	19	6	15
NIR	7	5	5	14	10	10	12
SWIR1	17	7	7	14	19	12	9
SWIR2	6	5	10	14	5	20	12

Difference Magnitudes

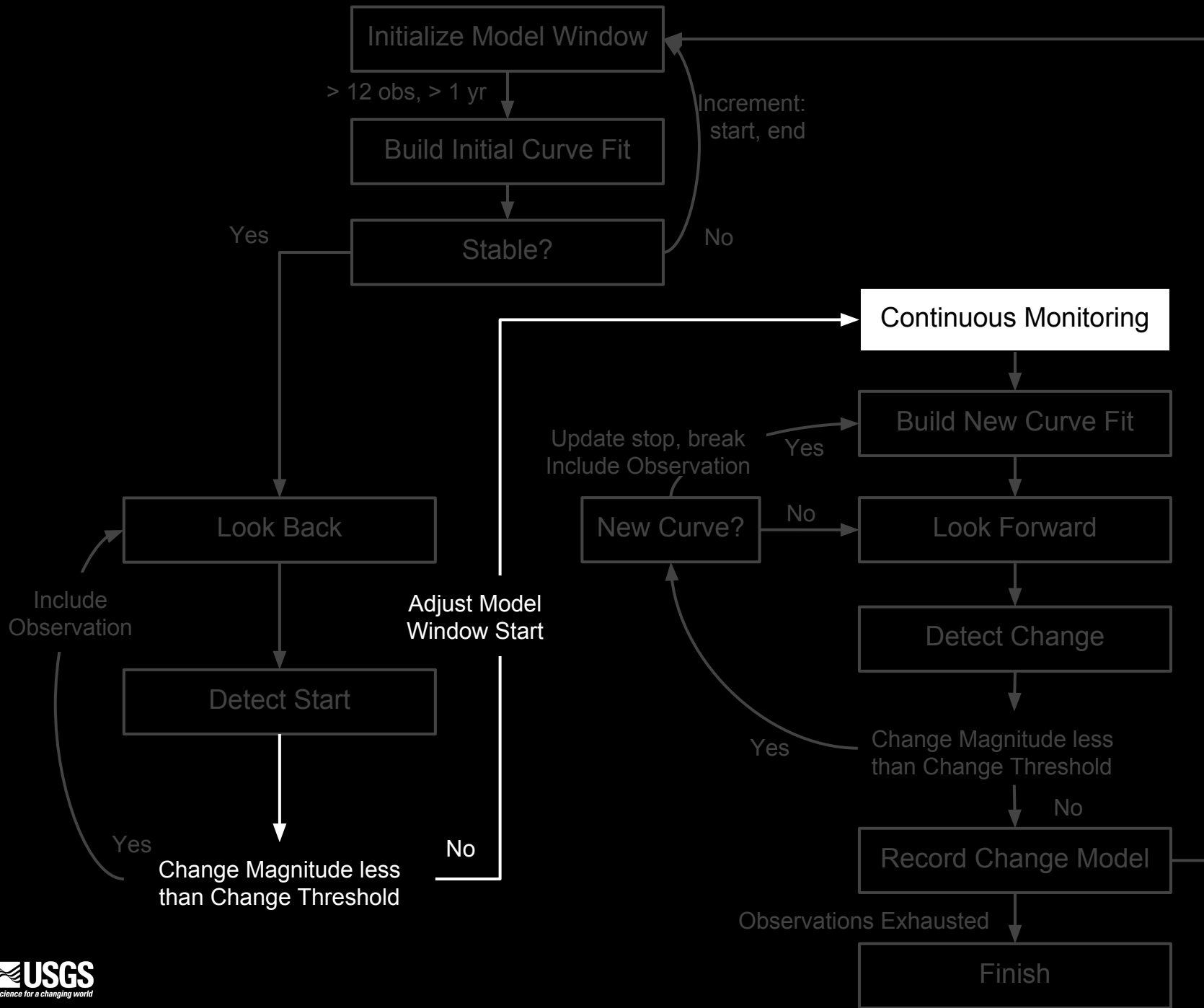
Green	0.339286	0.178571	0.357143	0.196429	0.285714	0.303571	0.214286
Red	0.267857	0.267857	0.178571	0.125	0.339286	0.107143	0.267857
NIR	0.125	0.089286	0.089286	0.25	0.178571	0.178571	0.214286
SWIR1	0.303571	0.125	0.125	0.25	0.339286	0.214286	0.160714
SWIR2	0.107143	0.089286	0.178571	0.25	0.089286	0.357143	0.214286

Change Magnitudes

0.306122	0.135204	0.214923	0.241709	0.351722	0.308992	0.235332
----------	----------	----------	----------	----------	----------	----------

Inc
Obse

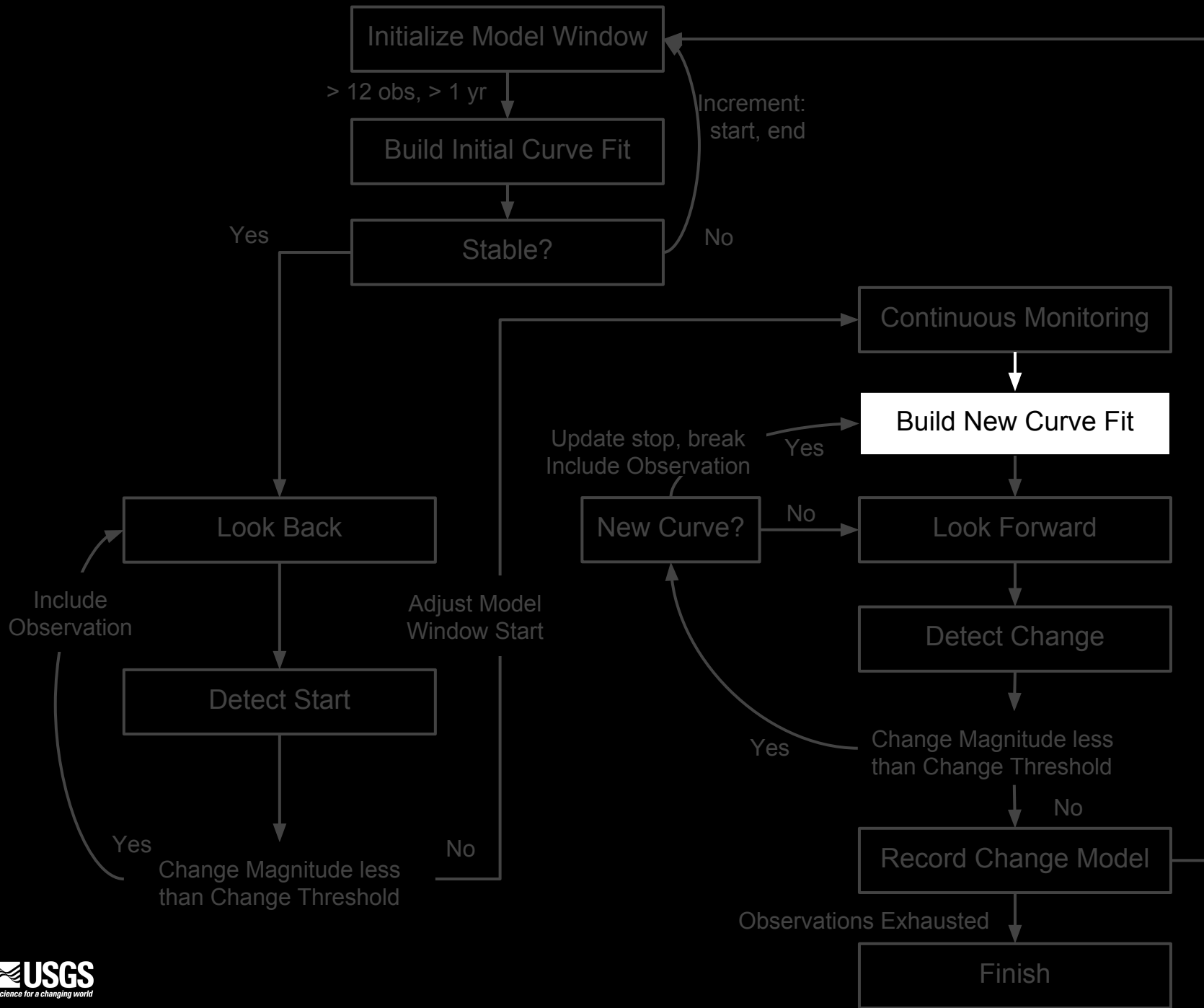




Continuous Monitoring

loop through remaining observations until change is detected or observations are exhausted

Inc
Obse



Build New Curve Fit

utilize LASSO Regression with a set number of coefficients based on the number of observations encompassed by the model window

Number of Coefficients per Model Window size:

<u># coefficients</u>	<u>observation count*</u>
4	12
6	18
8	24

*number of coefficient times the NUM_OBS_FACTOR (3)

Look Forward

(extending the model in time)

Time (Julian days)

day (0)

day (i)

day (i * 1.33)

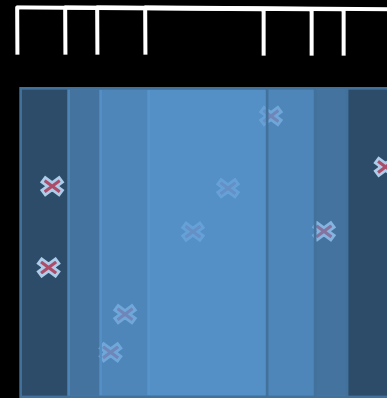
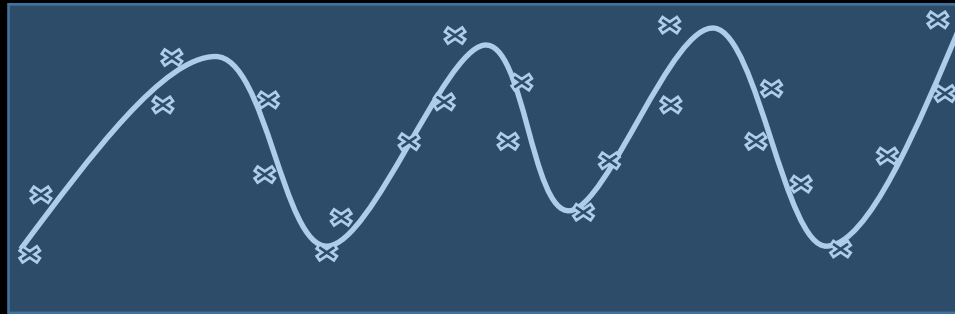
Fit Window

Fit window

Model Window

Peek Window N / N

Inc
Obse



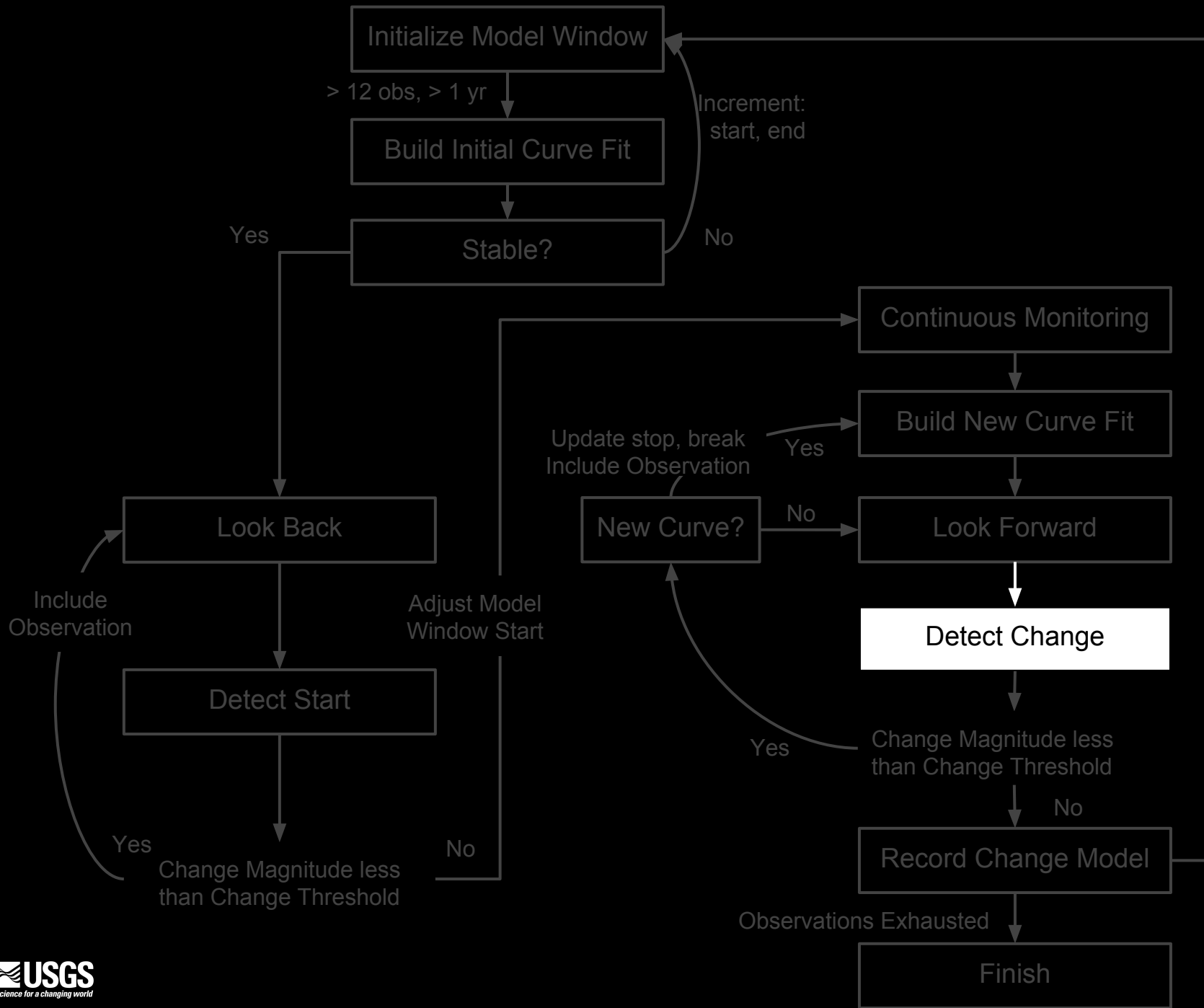
clear_x (0)

clear_x (24)

clear_x (i_start)

Observations

Finish



Detect Change

using the observations in the Peek Window

if the Model Window > 24 observations:

 model RMSE used is recalculated using the closest 24 observations temporally to the Peek Window

calculate a change magnitude per observation

for each of the detection bands (r ,g, n, s1, s2):

 difference magnitude = residuals / MAX

 (where MAX = max (variogram, model RMSE))

Inc
Obse
change magnitude = sum (difference magnitude ^2)

if the minimum (change magnitude) > change threshold:

 change detected

if the first change magnitude > outlier threshold:

 update persistent processing mask

else:

 include the observation in the model window



New Curve?

If < 24 observations:

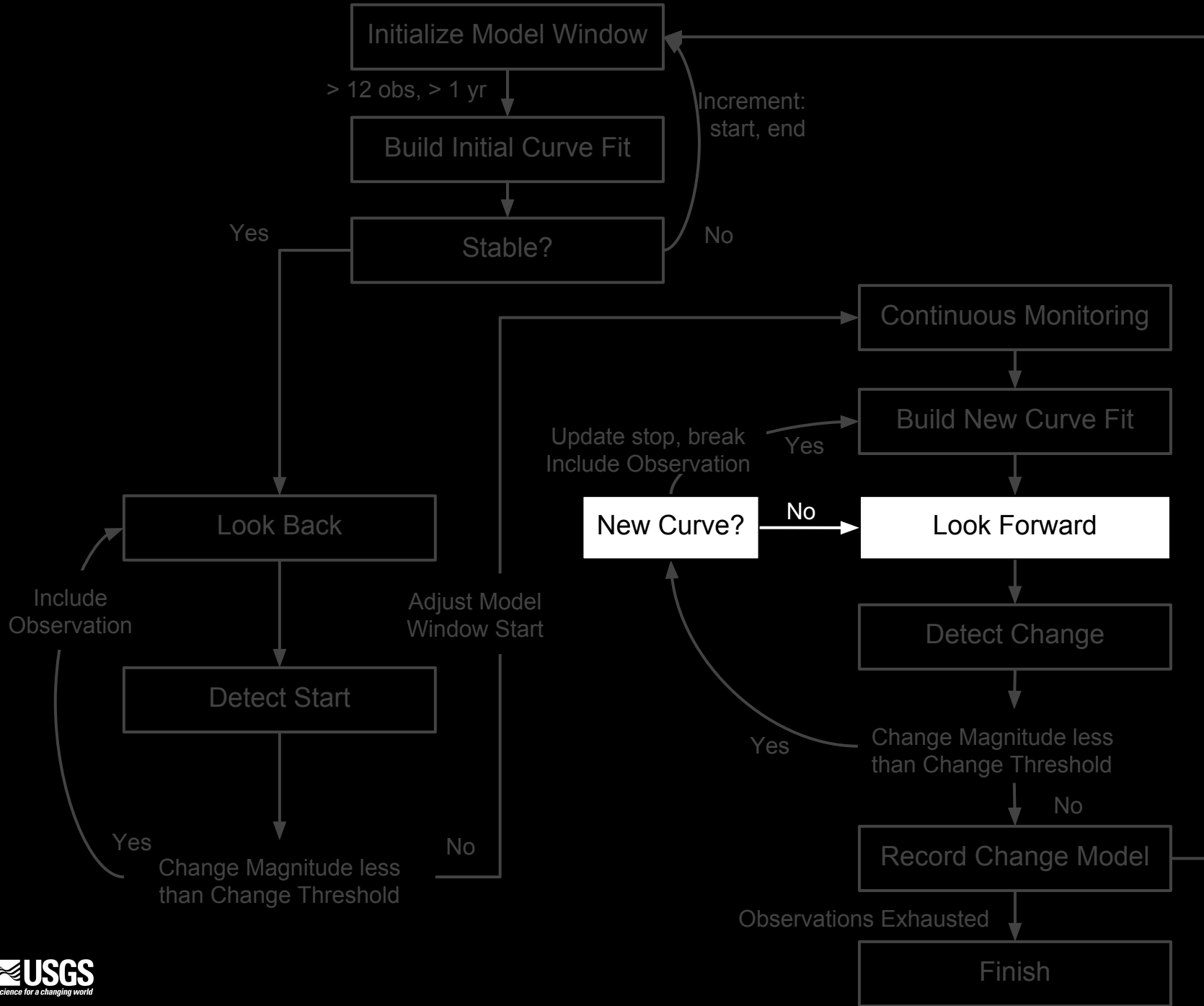
model new curve with LASSO Regression

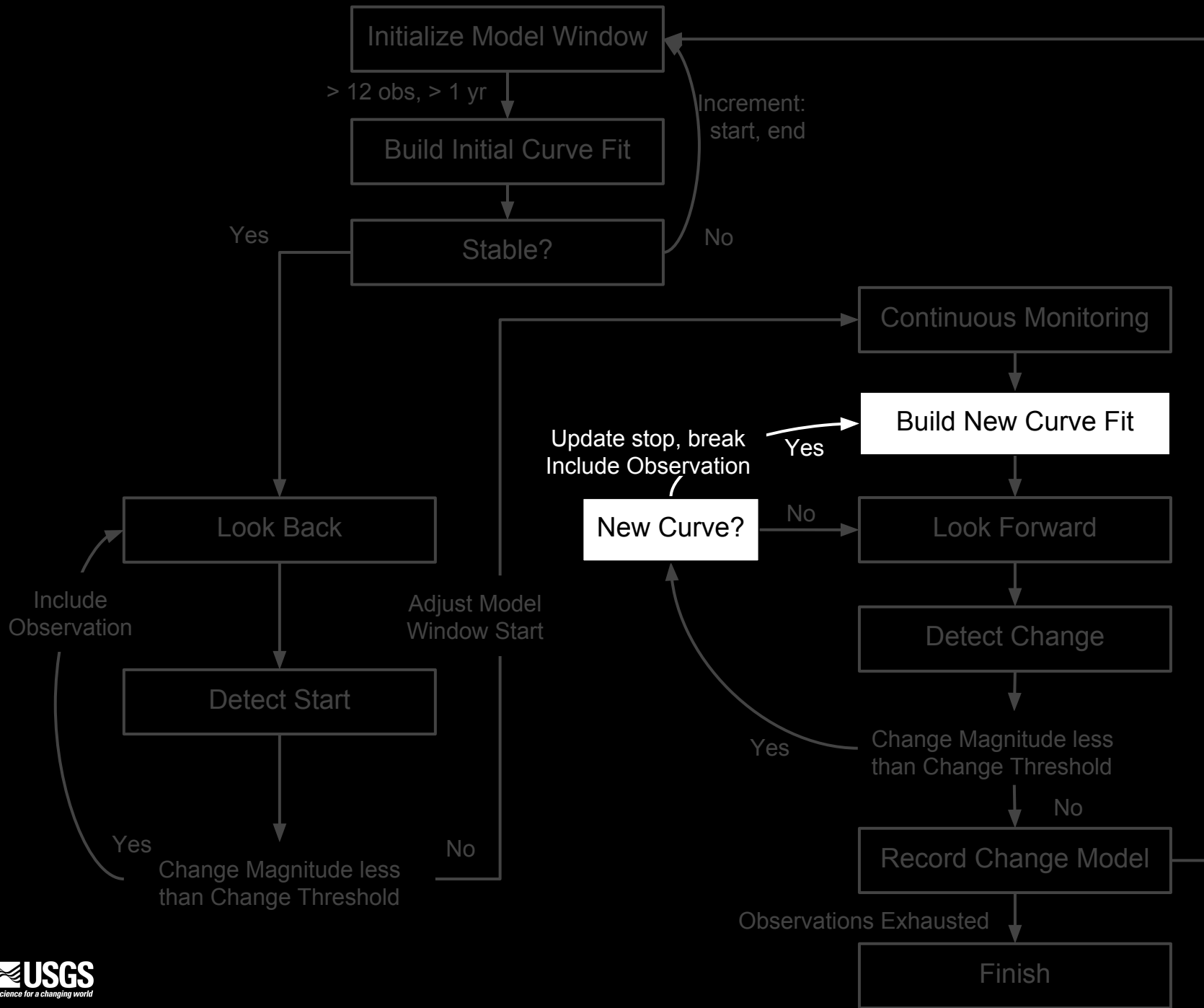
if > 24 observations and the time span of observations has increased by a factor of 1.33 since the last curve fit:

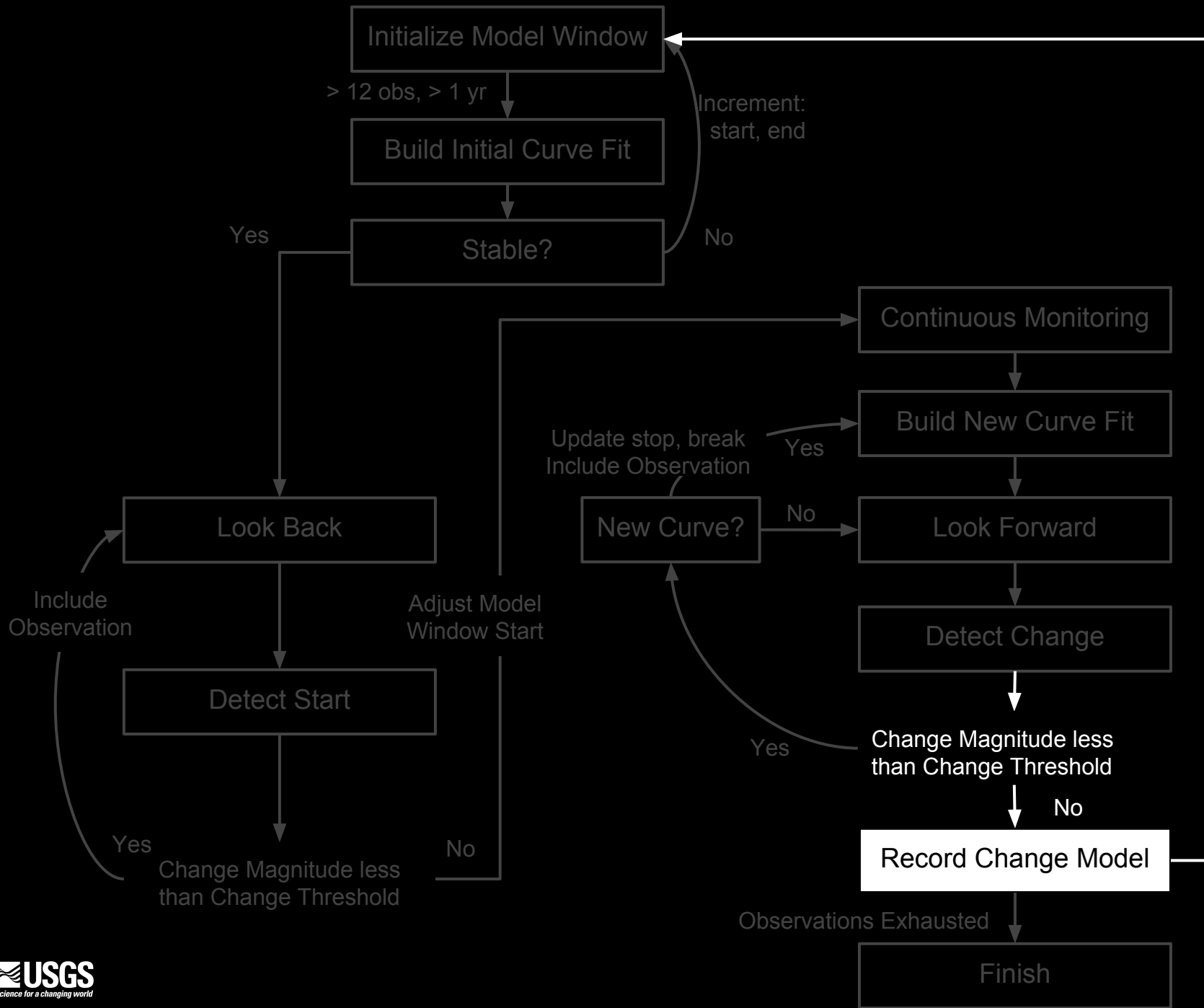
model new curve LASSO Regression

else:

continue to use existing model curve







Record Change Model

start

end

break

per band coefficients

per band RMSE's

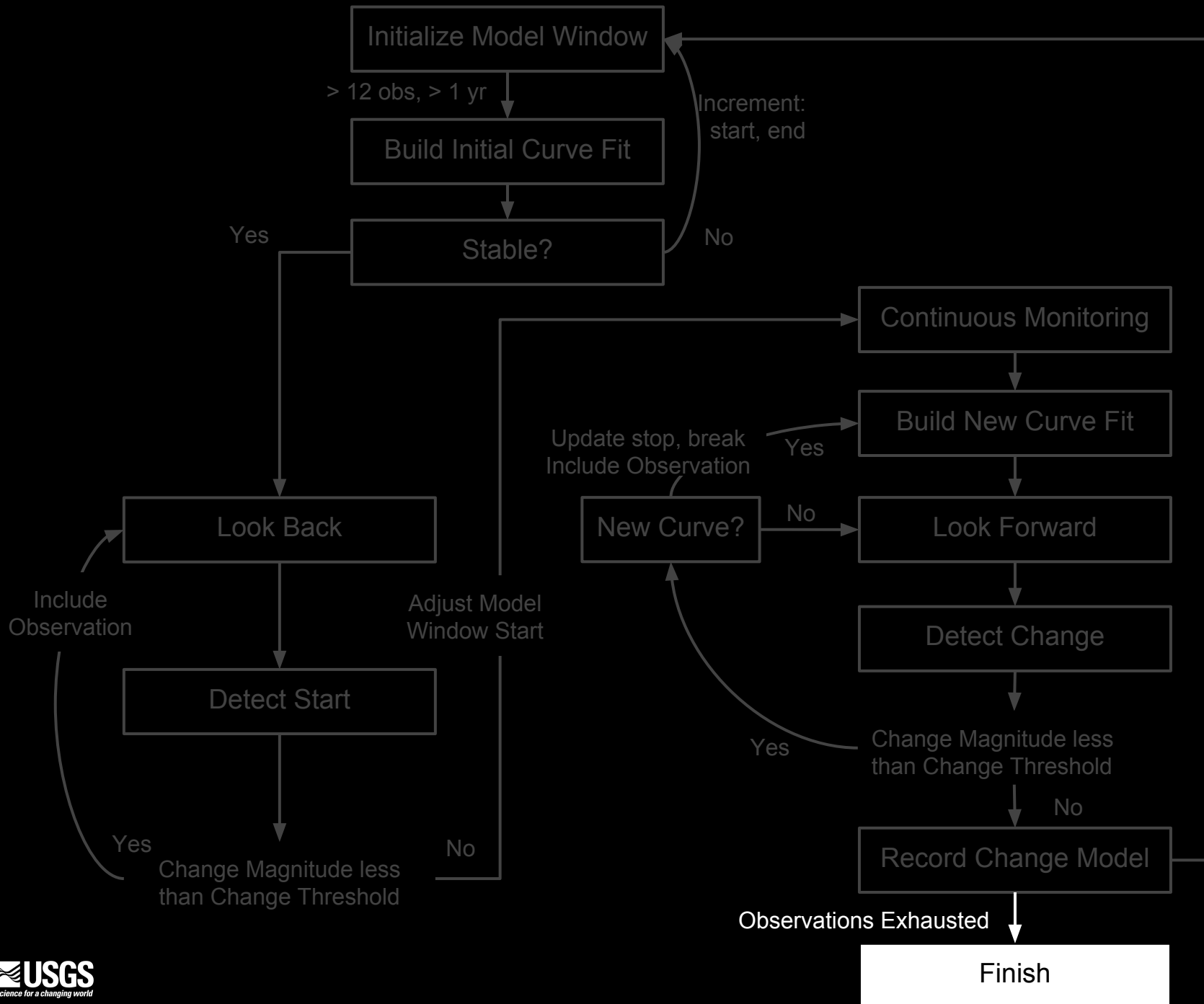
per band median of the last used Peek Window residuals

number of observations

number of fitted coefficients

change probability

Finish



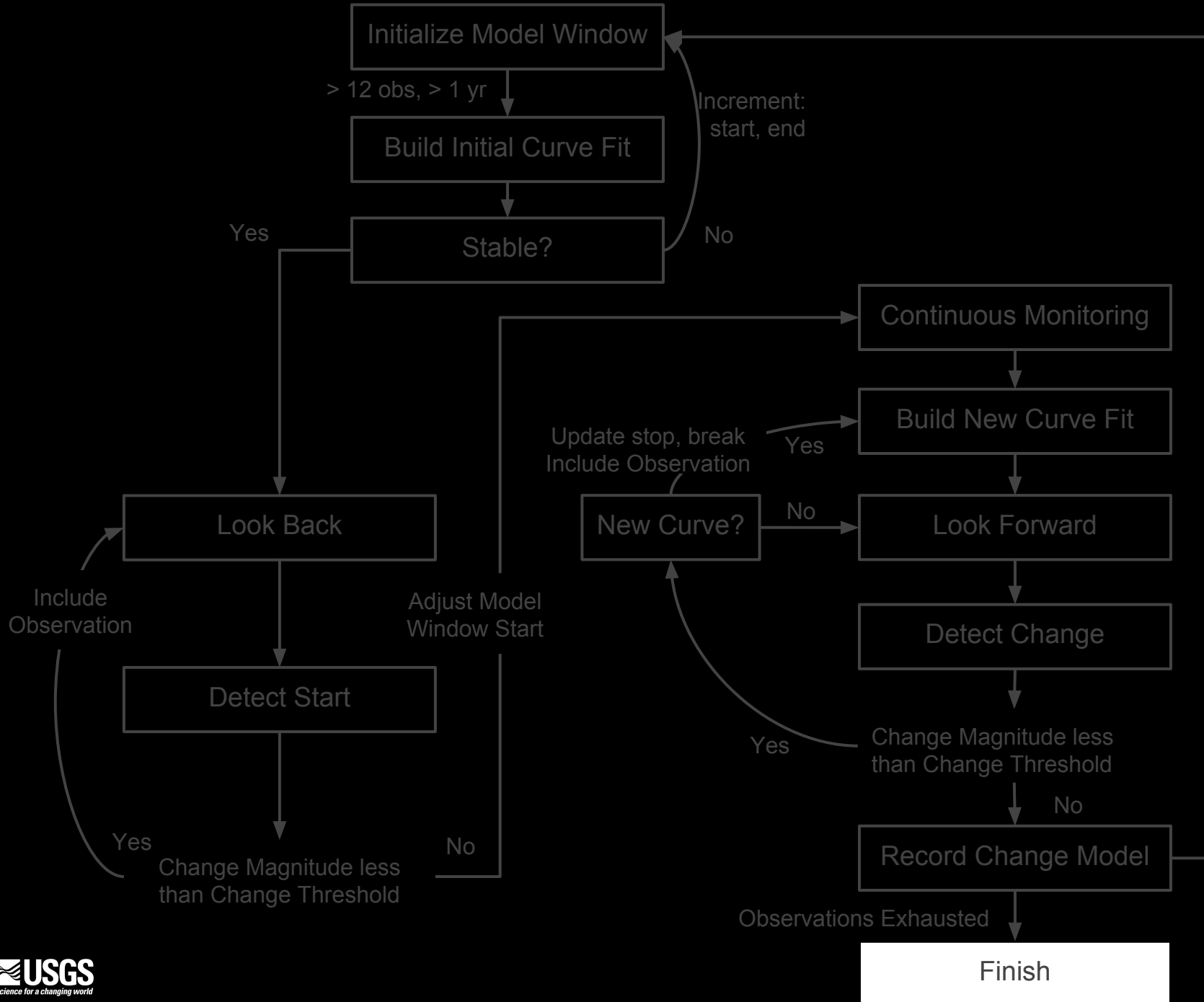
Finish

if the number of remaining observations is $< \text{MEOW size}$ but $\geq \text{Peek Size}$:

fit a generalized model is for remaining observations

send/save output results

exit



Results Output

start
end
break
per band coefficients
per band RMSE's
per band median of the last used Peek Window residuals
number of observations
number of fitted coefficients
change probability
list of recorded change models
procedure used
Persistent Processing Mask
algorithm version

Finish

Questions / Discussion

Acronym List

AOI	Area Of Interest
API	Application Programming Interface
ARD	Analysis Ready Data
CCD	Continuous Change Detection
CFMASK	C version of Function of MASK
DN	Digital Number
EROS	Earth Resources Observation and Science
IW+DS	Information Warehouse + Data Store
LASSO	Least Absolute Shrinkage and Selection Operator
LCMAP	Land Change Monitoring and Projection
MEOW	Minimum Expected Observation Window
NIR	Near InfraRed
SR	Surface Reflectance
SWIR	Short Wave InfraRed
TOA	Top Of Atmosphere

Bibliography

CCD:

<http://www.sciencedirect.com/science/article/pii/S0034425714000248>

<http://www.sciencedirect.com/science/article/pii/S0034425715000590>

<http://www.sciencedirect.com/science/article/pii/S0034425712000387>

ARD:

<https://drive.google.com/drive/folders/0B5www78kdogOfmZIWGI1S2dNRE5LeDVNeGxDU1phUndyTHpMbUVuNFdFNTZBV2RsbVhqMjQ>

<https://drive.google.com/a/doi.gov/file/d/0B5www78kdogONnN5N1BGTjFvdGM/view?usp=sharing>

ADD, these slides:

<https://drive.google.com/drive/folders/0BzELHvbrg1pDOFlwbGZiT29IMU0>

git Repository:

<https://github.com/USGS-EROS/lcmap-pyccd/tree/develop>

Slide Morgue

Initialize Model Window

initialize first model window

(start is set to previous break point; window size set to contain 12 observations)

while there are clear observations:

mask outliers in current window as defined by tmask

if not enough observations (12) and minimum temporal span (1 year) in the window:

increase window size by 1 continue to the next observation

Build Initial Curve Fit

Perform Lasso Regression using 4-coefficient model

Stable?

For each of the detection bands (r,g,n,s1,s2)

model slope coef * diff (end – start dates)

check_vals = (|slope| + |first model residual| + |last model residual|) / MAX

(where MAX = max (variogram, model RMSE))

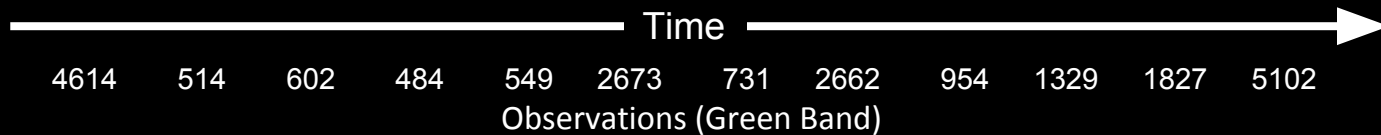
if summation of (checkvals)² < change threshold

stable

else

increment model window start and end by 1

Stable? (example)



Fitted Model (Green Band)

$$C = -44782.78$$

$$C_1 = 0.062 \quad (\text{slope coefficient})$$

$$C_2 = 194.19$$

$$C_3 = 0$$

Model Window = start: 724785 end: 725777

First Residual = 200 Last Residual = -56

$$\text{Green check value} = (|0.062 * (725777 - 724785)| + |200| + |-56|) / 32.56 = 9.751351351351351$$

Red = 10.56 NIR = 15.23 SWIR1 = 18.78 SWIR2 = 37.96

$$\text{Check Value} = 9.75^2 + 10.56^2 + 15.23^2 + 18.78^2 + 37.96^2 = 1440.96$$

Look Back

for each observation preceding current model window start
and end of previous model window

 If change magnitude $<$ change threshold:

 add observation to current model window

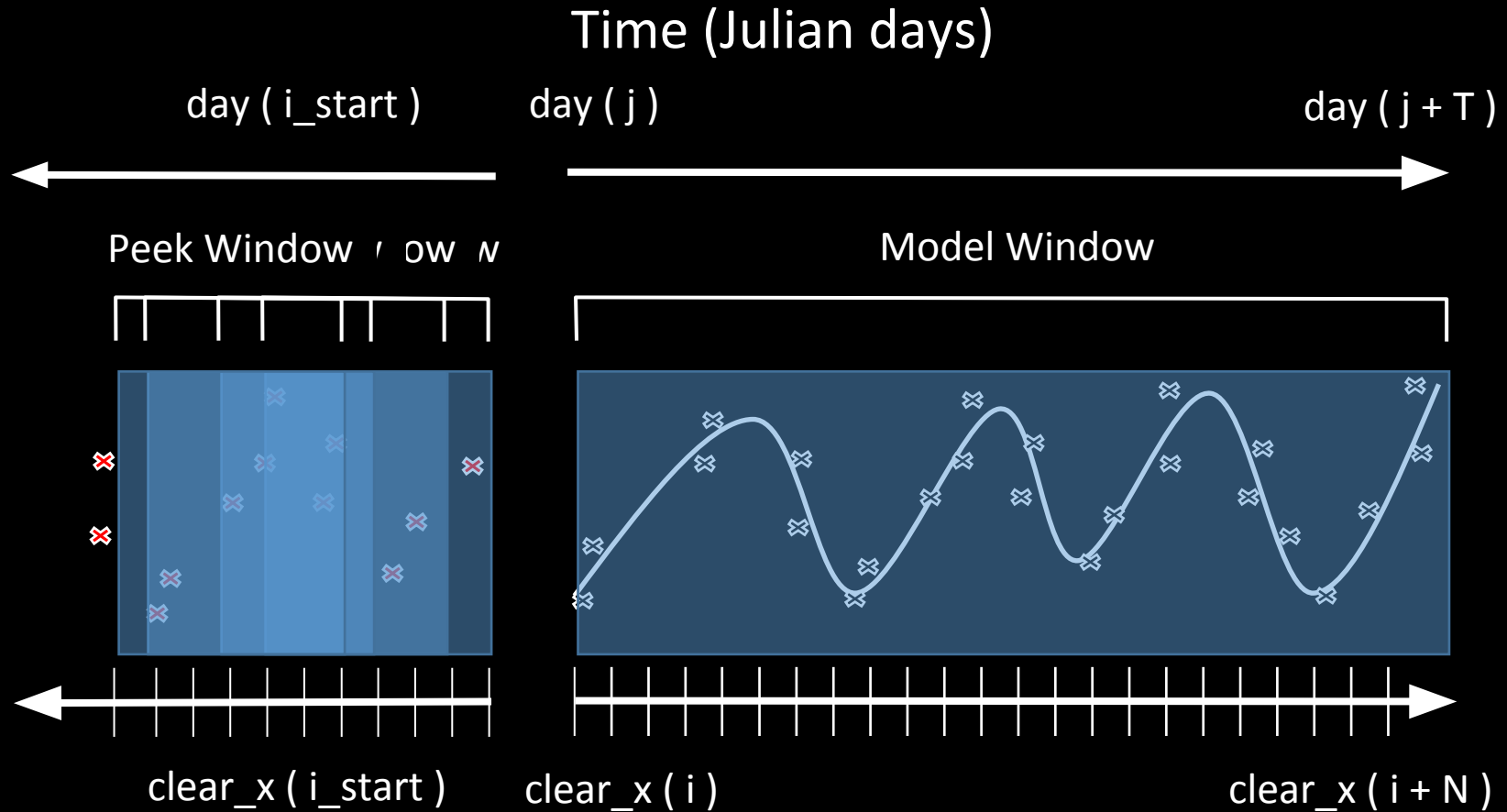
 else:

 proceed with continuous monitoring

note: when look-back is complete, if this is the very first model window, and
there 6 or more unused observations, do a general curve fit on them

Look Back

(adjusting the model start)



Detect Start

calculate change magnitudes for 5 detection bands
in current peek window

if smallest change magnitude $<$ change threshold:
include observation

else:

update current model window start, proceed
with Continuous Monitoring

note: if the change magnitudes of the first observation in the peek window is
greater than the outlier threshold, update persistent processing mask to exclude
obserservation.

Continuous Monitoring

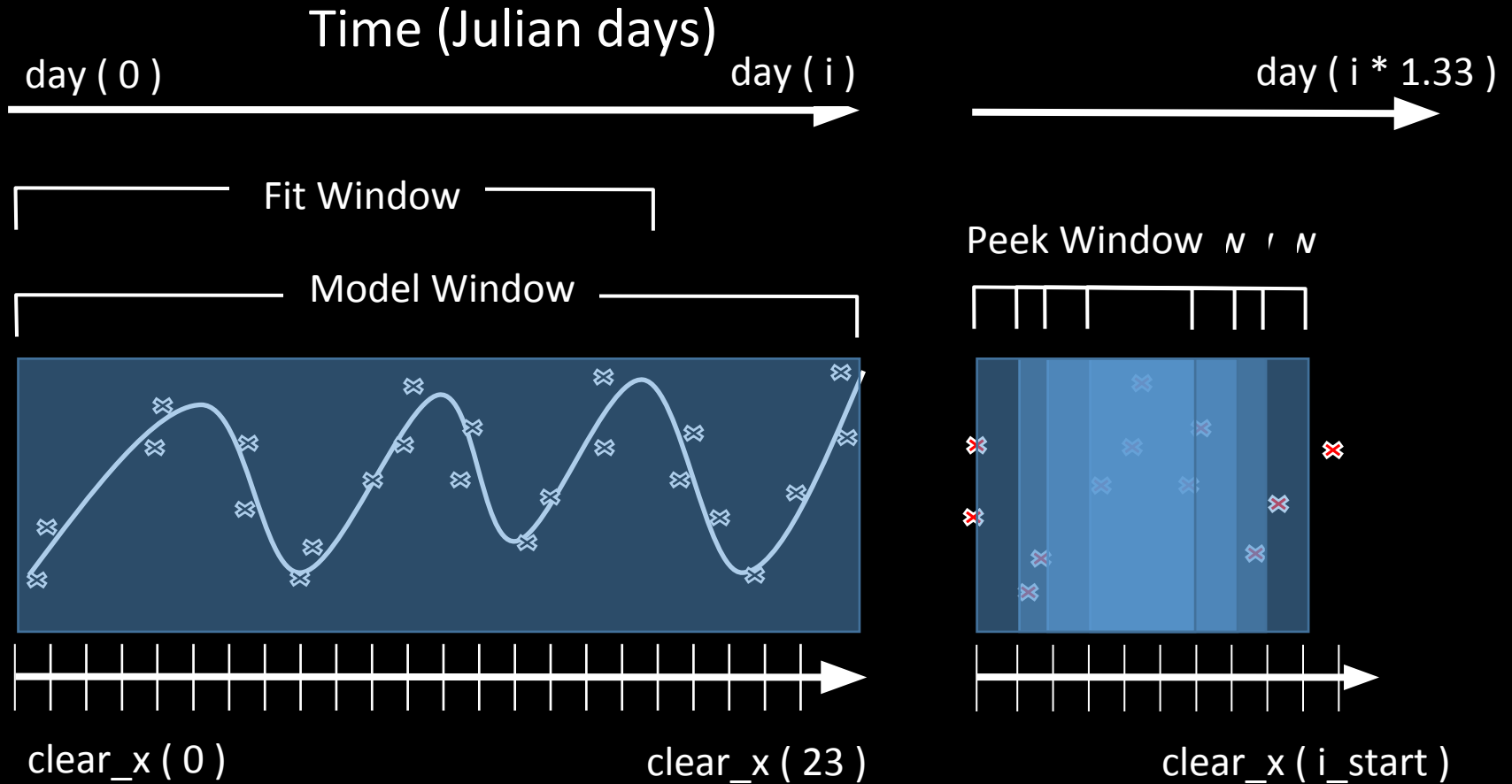
(description)

Build New Curve Fit

(description)

Look Forward

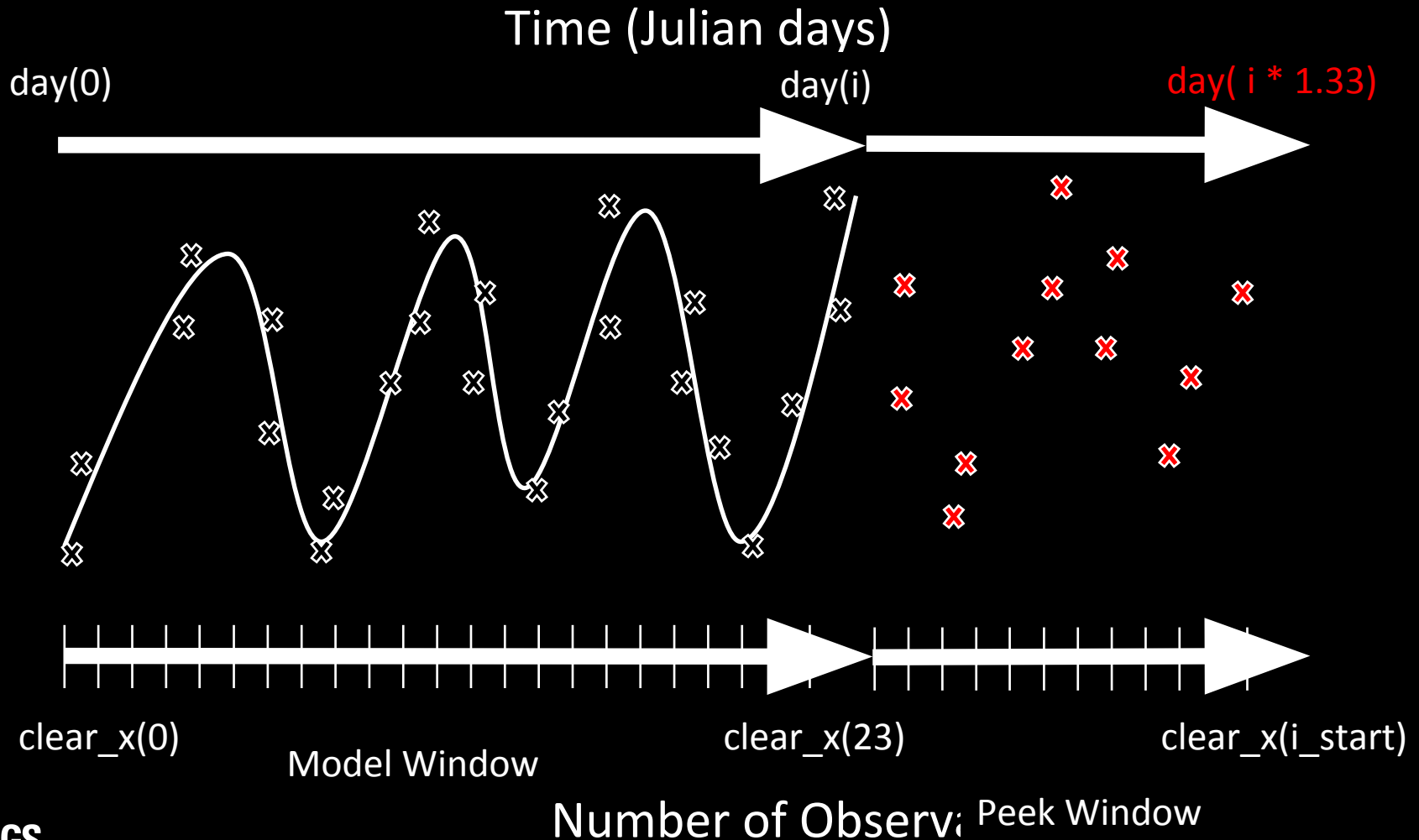
(extending the model in time)



Look Forward

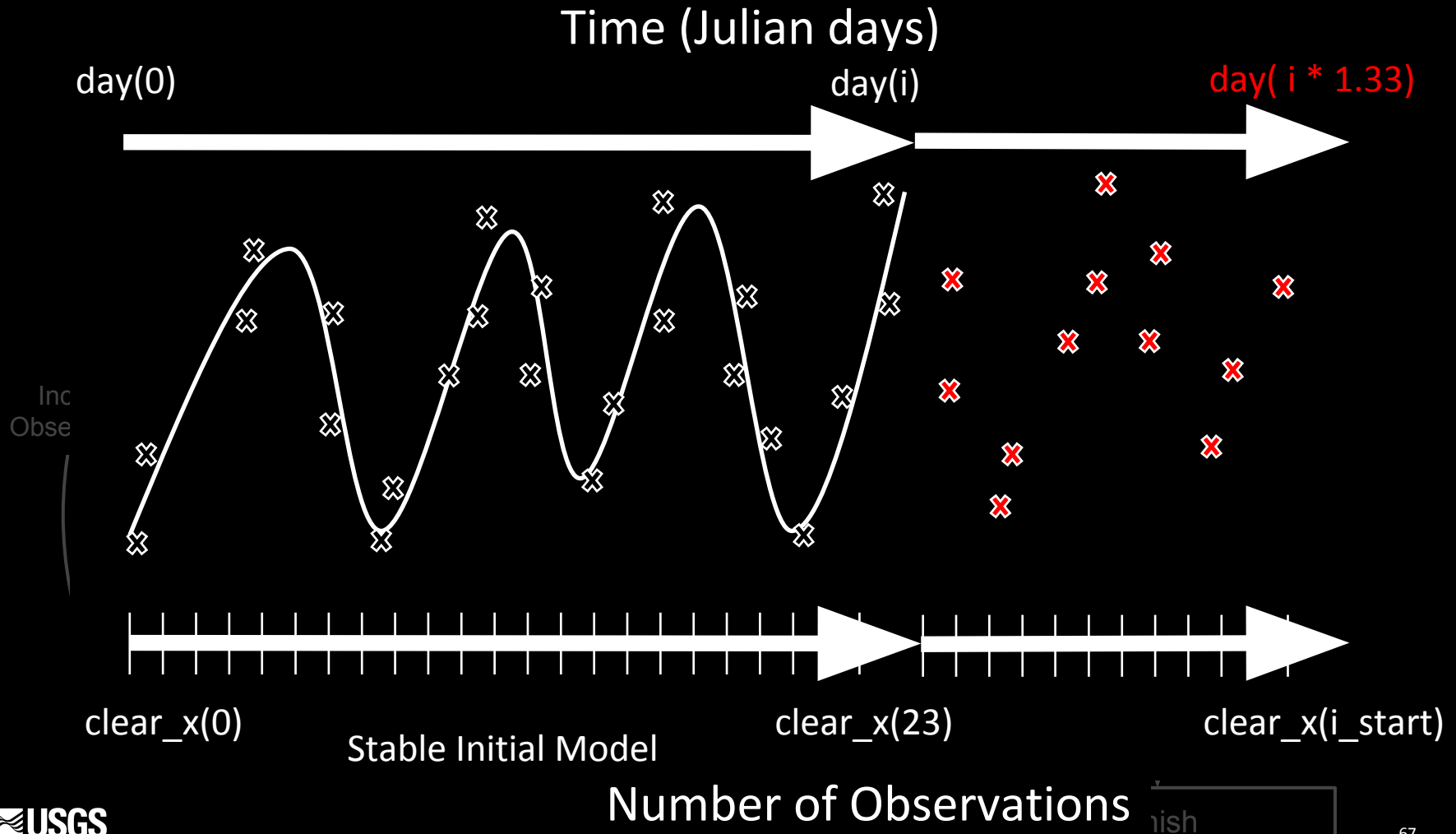
Fit Window

(extending the model in time)



Look Forward

(extending the model in time)



Look Forward

(description)

Detect Change

(description)

Record Change Model

(description)

New Curve

(description)

Finish

(description)

corresponding artifacts

- Create/update corresponding ADD and workflow diagram
- update .md to reflect pseudo-code of latest pyccd version
- make slides and other docs available

Steps for building a time segment (continuous monitoring)

1. Fit a new model to initial stable window (include tmask-filtered values excluded from initial model)
2. Calculate median residual value for each band (for the observations in the “peek window”)
3. Calculate a change magnitude
4. Make a determination if change or outlier based on threshold values
 1. Change: (Look-back, extend are previous to this)
 2. Outlier: skip?
 3. Else increment window/observation dates?

Standard Procedure (cont.)

Detect Change in Select Bands

~~nputs: per band observations, current models, start date, end date, median band variogram values, T_Cg value~~

~~for each detection band:~~

~~—minimum rmse = max(median band variogram, model rmse)~~

~~—slope = model slope coefficient * (end date - start date)~~

~~—band check val = (abs(slope) + abs(first residual) + abs(last residual)) / minimum rmse~~

~~if the norm(band check vals)**2 > T_cg:~~

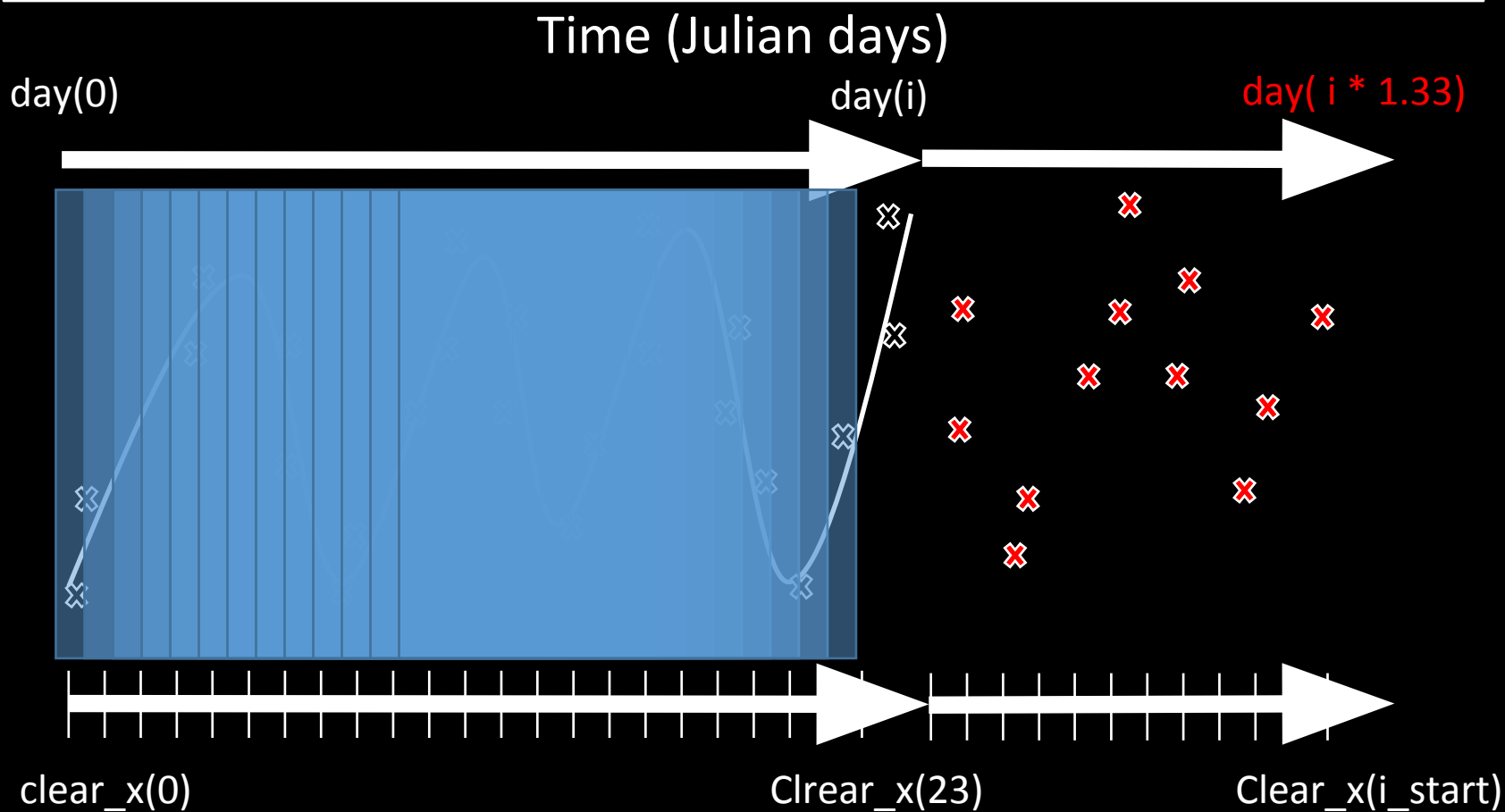
~~—change detected~~

Input ARD Example

WA grid07, 1982 – 2015, TM, ETM+, OLI/TIRS

Previous		Current	
Path/Row “scenes”	647	3,922	All ARD Observations
20% clear or better “scenes” (54%)	354		2,147 20% clear or better ARD Observations (N/A)
(what is the “single image has less than 50% data is screened out”?)			
Non-fill pixels, Single, Specific X/Y location (65%)	354	1,514	
Clear pixels, Single X/Y location	252	985	
Avg. observations per year (34 years)	7.4	28.97	
Area Of Interest (AOI) limited to Path/Row boundaries			Flexible AOI

Window



Initial Change Detection Setup

1. Mask out duplicate observations using first of two swath overlap observations, sort temporally *
2. Convert thermal values from degrees Kelvin to Celsius
 $\text{value} = \text{value} * 10 - 27315$
3. Create mask(s) of values that fall outside of acceptable ranges, considered saturated if $> \text{max}$
 - Reflectance acceptable range: 0 - 10,000
 - Thermal acceptable range: -9,320 - 7,070
4. Derive some QA (cfmask) information for reference during processing
 - Non-Fill mask
 $\text{value} < 255$
 - Clear pixel mask
 $\text{value} < 2$ (clear=0, water=1)
 - Percent Clear
 $\text{sum}(\text{clear}) / \text{sum}(\text{non-fill})$
 - Snow Mask
 $\text{value} == 3$
 - Percent Snow
 $\text{sum}(\text{snow}) / (\text{sum}(\text{clear}) + \text{sum}(\text{snow}) + 0.01) - - - \text{sum}(\text{snow}) / \text{sum}(\text{non-fill})$

Preliminary data prep for Standard Procedure (combine with previous)

Mask out any repeated date values, and their associated spectral values, keep first of 2 (northern)

Filter values outside acceptable min/max

Sort chronologically

Calculate a modified first-order
variogram/madogram across each spectral band

```
>>> import numpy as np  
>>> vario[band_2_idx] = np.median(numpy.abs(numpy.diff(band_2)))
```

Coefficients

1.

2.

3.

4.

5.

6.

7.

8.

Change Model

(describe what this is)
obsolete?

Curve Fitting Attributes

start point

end point

coeffs for each band

RMSE, magnitude for each band