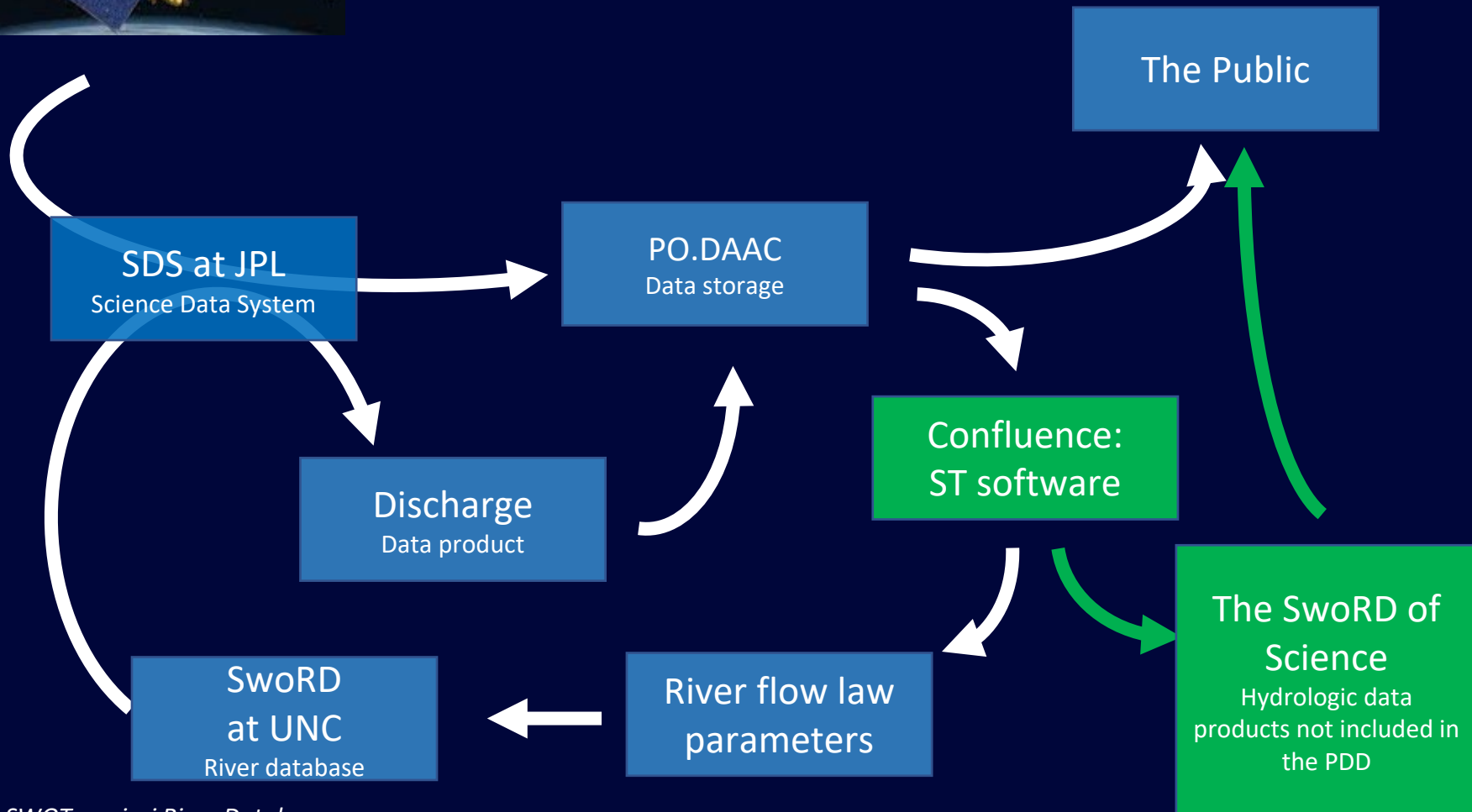


An aerial photograph of a city with a river winding through it. The river is dark green and flows from the top center towards the bottom left. It has several sharp turns, creating a complex path through the urban landscape. The city is densely packed with buildings, streets, and green spaces. The river eventually joins a larger body of water at the bottom left. The text 'Confluence: ST discharge software' is overlaid in white, sans-serif font in the upper half of the image.

# Confluence: ST discharge software

Colin Gleason  
Nikki Tebaldi  
UMass Amherst

# What is Confluence?



*SwoRD: SWOT a priori River Database*

*SDS: Science Data System*

*ST: Science team*





INPUT

PRE

FLPE

POST

INTEG.

POST

OFFLINE

OUTPUT

## Amazon Web Services (AWS) Infrastructure

- Amazon Simple Storage Service (S3) buckets
- Amazon Elastic File System(s) (EFS)
- Docker containers
- AWS Batch

## Confluence workflow

- Program that creates and orchestrates jobs
- The workflow is made up of stages
- Each stage has one or more modules
- A module is considered a job

INPUT

PRE

FLPE

POST

INTEG.

POST

OFFLINE

OUTPUT

SWOT  
OBSERVATION

S

- HEIGHT
- WIDTH
- SLOPE
- dA
- FLAGS
- SHAPEFILE

INPUT

- PASS/CYCLE -> nx by nt
- TRANSFORM TO NETCDF
- STORE IN ELASTIC FILE SYSTEM

TO SOS  
NOTHING  
(TEMPORARY  
STORAGE)

Contributors:  
Nikki Tebaldi  
Steve Coss

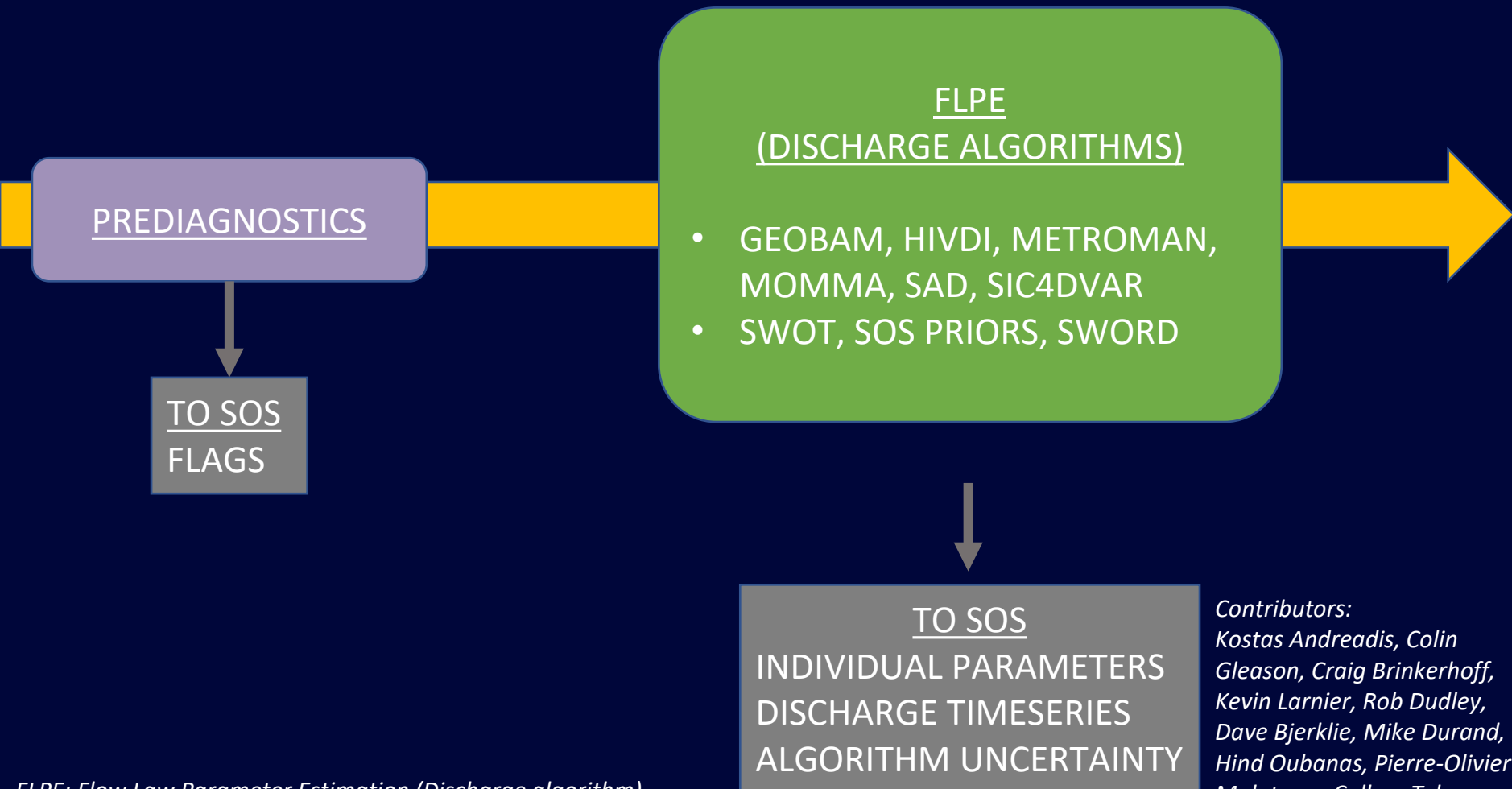
PREDIAGNOSTICS

- SEASAME STREET FILTER
- PASS-THROUGH SWOT FLAGS
- MISSING DATA TREATMENT

TO SOS  
FLAGS

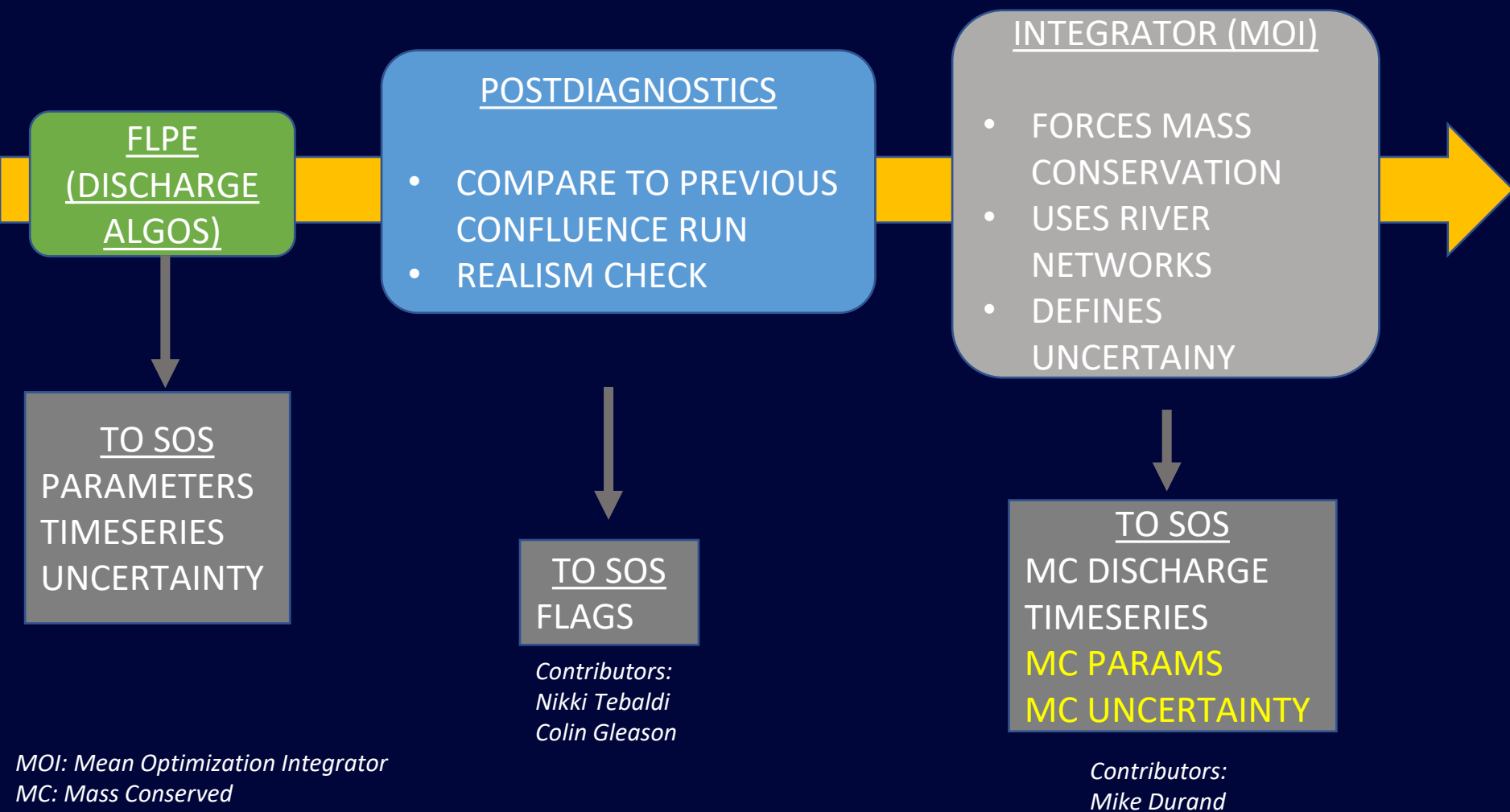
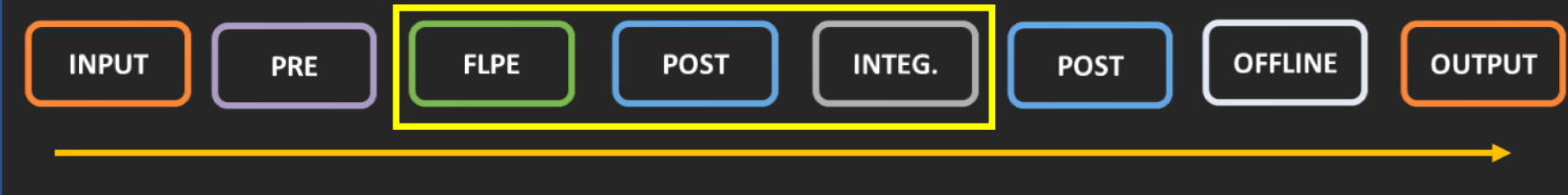
Contributors:  
Nikki Tebaldi  
Colin Gleason

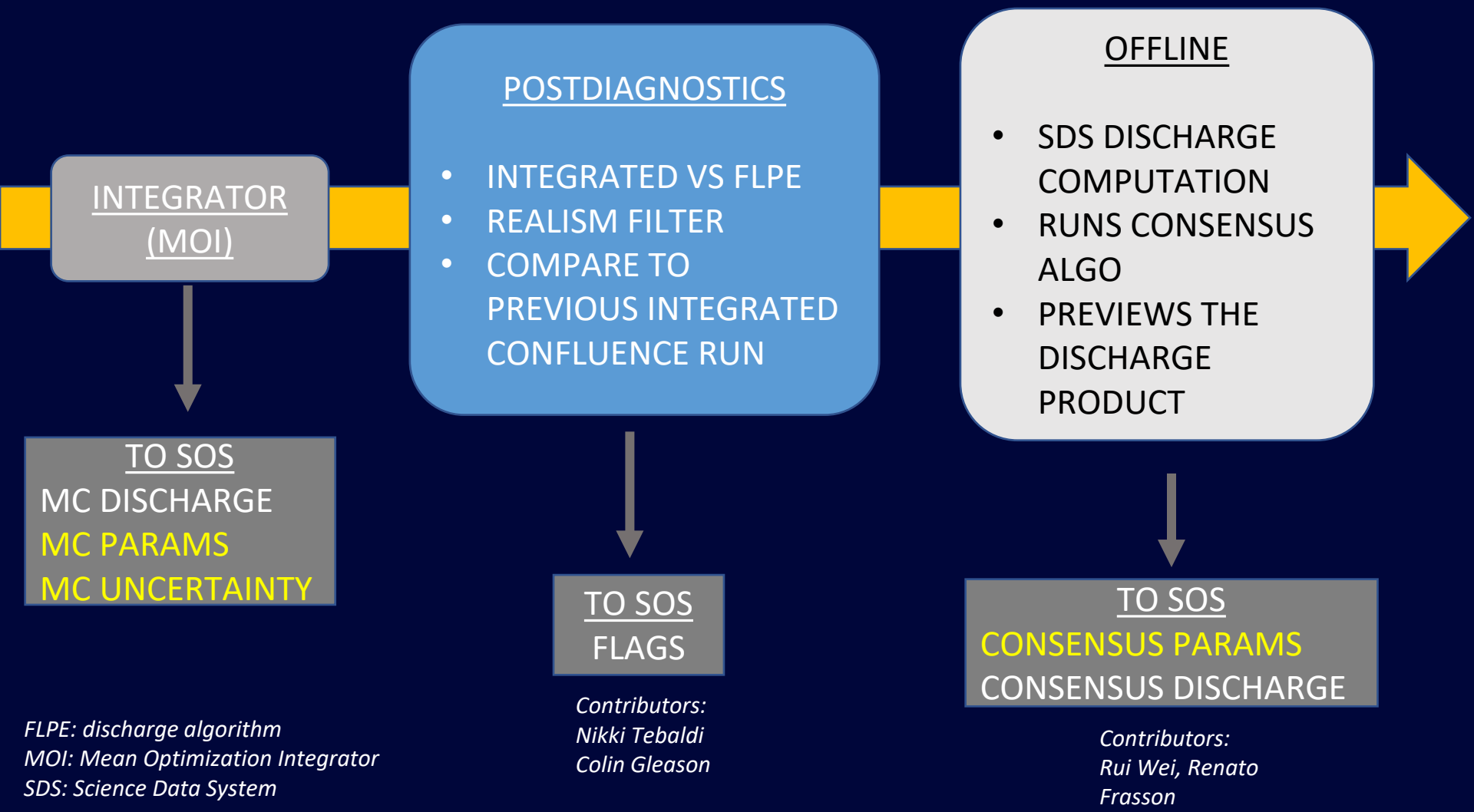
dA: change in river cross section area  
nx: number of spatial units  
nt: number of observations  
Seasame Street: outlier filter  
SoS: Sword of Science



*Contributors:*  
Kostas Andreadis, Colin Gleason, Craig Brinkerhoff, Kevin Larnier, Rob Dudley, Dave Bjerklie, Mike Durand, Hind Oubanas, Pierre-Olivier Malaterre, Callum Tyler

*FLPE: Flow Law Parameter Estimation (Discharge algorithm)  
Parameters: discharge parameters as described in PDD*







INPUT

PRE

FLPE

POST

INTEG.

POST

OFFLINE

OUTPUT

OFFLINE

VALIDATION

- COMPARES TO GAUGES
- CALCULATES ERROR STATS
- GENERATES HYDROGRAPHS

OUTPUT

- WRITE TO S3
- VERSION CONTROL

*Contributors:  
Nikki Tebaldi*

SwoRD OF SCIENCE  
S3 BUCKET

- PUBLIC ACCESS
- UNIQUE VERSION FOR EACH RUN

TO SOS  
CONSENSUS  
PARAMS  
CONSENSUS  
DISCHARGE

TO SOS  
STATS  
HYDROGRAPHS

*Contributors:  
Steve Coss*

TO SwoRD  
PARAMS  
UNCERTAINTY

*Contributors:  
Nikki Tebaldi*

# What's next?

- Move to release versions
- Optimize SoS
- Handle exceptions and job failures
- Optimize modules for runtime and cost
- Test!!!

# Recap

- Confluence produces global discharge for DAWG
- We want to share
- Working Beta in AWS
- Containerized and open source
- Run often (ideally weekly) after launch
- Consider costs- without additional support, only UMass can run this, and only locally

The DAWG will be ready for launch



INPUT

PRE

FLPE

POST

INTEG.

POST

OFFLINE

OUTPUT

## Input

Description: Takes SWOT observation data and transforms shapefile data into time series NetCDF files with one file per reach.

Input: SWOT shapefiles

Output: Time series data (width, height, slope, dA, and flags) in NetCDF format

## Prediagnosics

Description: Parses the SWOT observation data formatted by the Input module to track quality indicators and perform consistency checks. Replaces flags with missing values.

Input: SWOT NetCDF files

Output: Overwrites SWOT NetCDF files and outputs flags.

## Reach-level Flow Law Parameter Estimation (FLPE)

Description: Take the SWOT observation data formatted by the Input module and processed by the Prediagnosics module as well as SoS priors and SWORD data to produce discharge and discharge parameters.

Input: SWOT NetCDF files, SoS NetCDF files, SWORD NetCDF files

Output: Discharge parameters, discharge time series, uncertainty as NetCDF files.



INPUT

PRE

FLPE

POST

INTEG.

POST

OFFLINE

OUTPUT

### Postdiagnostics FLPE

Description: Compares previous and current discharge values while also performing a realism check on the results of each FLPE algorithms' execution.

Input: FLPE algorithm NetCDF files and SoS NetCDF files

Output: Flags produced from comparisons and stored in NetCDF files

### Integrator (MOI)

Description: Takes the formatted SWOT observation data and reach-level FLPE output and integrates the results. It uses river network topology to force mass conservation and also defines uncertainty.

Input: SWOT NetCDF files

Output: Discharge time series for each FLPE algorithm, mass conserved parameters, mass conserved uncertainty stored in NetCDF files.

### Postdiagnostics Integrator

Description: Compares FLPE and integrated discharge values while also performing a realism check on the results of Integrator execution. Also compares previous and current discharge.

Input: MOI NetCDF files and SoS NetCDF files

Output: Flags produced from comparisons and stored in NetCDF files

INPUT

PRE

FLPE

POST

INTEG.

POST

OFFLINE

OUTPUT

## Offline

Description: Executes the RiverObs SWOT Algorithm that will be run by the Science Data System (SDS) to produce discharge for each FLPE algorithm and executes a consensus algorithm.

Input: SWORD NetCDF files

Output: Discharge values, consensus parameters and consensus discharge stored in NetCDF files

## Validation

Description: Compares the results of FLPE and integrator algorithms to gages, calculates stats and creates hydrographs.

Input: FLPE algorithm NetCDF files, Integrator algorithm NetCDF files, SoS NetCDF files

Output: Statistics stored in NetCDF files and hydrographs

## Output

Description: Appends the results of all stages and modules to a new version of the SoS.

Input: NetCDF files from every stage

Output: SoS NetCDF file uploaded to SoS S3 bucket