A step towards the automated production of Water Masks from Sentinel-1 SAR images

Pierre Fabry, Nicolas Bercher

ALONG-TRACK .com

Brest & Toulouse, France

Water masks from Sentinel-1 : Why ?

- **dynamic water masks (every 12 days) : all weather, night and day, high resolution !**
- Help **monitor water resource variations** over time
- Help **future missions** on Surface Waters (SWOT)
- Help evaluate **Retrackers' performance** in altimetry for hydrology
- Provide a priori information to retrackers
- Provide information to **better analyse radar altimeter waveforms** in hydrology
- **Combine radar altimetry and radar imaging** to derive **Hypsometric Curves** (Height-Surface and Height-Surface-Volume) curves as well as **bathymetry** over lakes in hydrology
- …

Algorithm selection criteria

- **No « threshold »** : algorithm shall adapt to the image content.
- Algorithm shall cope with **full resolution speckle noise**
- Directly provides countours (**region-wise segmentation** and not pixel-wise).
- **Fast** (versus the water-mask target resolution) and/or able to start from a **previous (over-) segmented mask**.
- No apriori choice on the **number of regions**.
- **Shall not over segment** (to ease the classification step)

Selected Algorithm

- The **Minimum Description Length Automated Segmentation Grid** (MDLSAG), which results from a long evolution :

1. *Statistical Region based Active Contours* (several authors, statistical framework)
2. *Statistical **Polygonal Snakes*** (Germain 1996, **Max. Likelihood** based allowing 2 classes only)
3. *Statistical Polygonal Active Grid* (Germain 2001, Bayesian, **multi-region** generalization of the Polygonal snake)
4. The **Minimum Description Length (MDL)** principle, introduced by Rissanen in 1978 was refreshed by Andrew Barron in 1998)
5. **MDL + Polygonal Active** *Grid* , PhD thesis Galland, 2003
6. In 2009, Galland et al. revisited the Automated MDL Polygonal Grid with the **assumption that intensity pixels of SAR images are appropriately modeled by Gamma PDF** with
   - **same order L** all over the SAR image
   - **mean intensity** depending on the regions.

SAR images Stochastic Model (1/2)

❑ Multiplicative Noise

SLC SAR Image ≡ 1 occurrence of random field $S_\lambda(x,y)$ all of the r.v. depend on random event :

Under basic assumptions (Goodman), the **r.v.** $S_\lambda(x,y)$ (**pixel intensity**) in an area with **constant mean reflectivity (μ)** follow an **exponential law** :

$$P_S(s) = \begin{cases} \frac{1}{\mu} \cdot e^{-\frac{s}{\mu}}, s > 0 \\ 0 \text{ sinon} \end{cases} \Rightarrow \begin{cases} E[S] = \mu \\ \operatorname{var}[S] = \mu^2 \end{cases}$$

$b \quad S_\lambda(x,y) = \mu \cdot b(x,y)$

=> ∃ **multiplicative noise** ,

$P_{S,x,y}(s)$

If we could repeat the random experiment then we could learn the pixels' PDF ,
instead we are going to assume **homogeneity** and compute a « spatial mean value »

SAR images Stochastic Model (2/2)

❑ **How to define « Homogeneous Region » ?**

- Strict definition at order 1 (impractical from 1 image occurence) :

$$\left\{ (x,y),\, P_{S,x,y}(s) = P_S(s) \right\}$$

- Wide sense definition (2nd order over neighborhoud, applicable to 1 image occ.) : $\left\{ (x,y),\, E\left[S_\lambda(x,y)\right] = E\left[S\right] \text{ et } E\left[S_\lambda(x,y)\cdot S_\lambda(x+\Delta x, y+\Delta y)\right] = f(\Delta x, \Delta y) \right\}$

In practice (SLC to GRD) → L multi-looking → Modification pixels ' PDF :

- Exponential Law → **Gamma Law** (Assumption-1 from Galland, 2009) :

$$P_{\bar{S}}(\bar{s}) = \begin{cases} \dfrac{L^L}{\mu\cdot\Gamma(L)}\cdot\left(\dfrac{\bar{s}}{\mu}\right)^{L-1} e^{-L\frac{\bar{s}}{\mu}},\ \bar{s} > 0 \\ 0 \text{ sinon} \end{cases} \Rightarrow \begin{cases} E\left[\bar{S}\right] = \mu \\ \operatorname{var}\left[\bar{S}\right] = \dfrac{\mu^2}{L} \end{cases}$$

- In practice **L = ENL** (example : L = 4.9 on Sentinel-1 IW HR GRD)

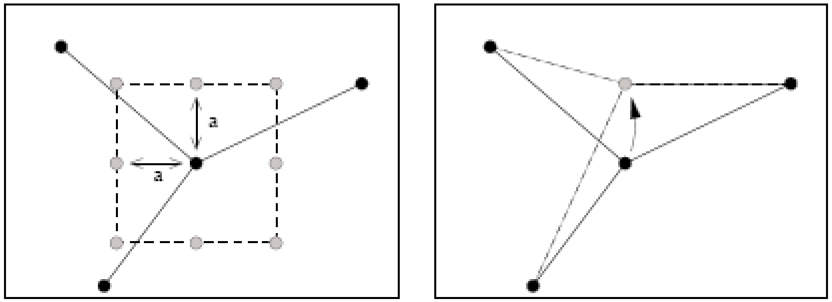
The MDL Principle (2/3)

Initial partition = **grid** = set of **NODES** linked together through **SEGMENTS** to define rectangular **REGIONS**

Iteratively perform **3 types of grid modifications** and **keep the changes that lower the « length of the description message »** (**Stochastic Complexity**)

1. Merge : test all possible region merges 2 by 2

2. **Move**: Try to Move the nodes in 8 directions (amplitude is a function of connected segments' length) with decrease.



(a) (b)

3. **Remove** : Go through each node … other nodes) and evaluate its SC reductio…

The MDL Principle (3/3)

- **Stochastic Complexity** (SC) as derived by Galland :

- $\Delta_G(w)$: Geometric term of the SC : nb of bits to encode partition w,

  depends on $\begin{cases} k \text{ : nb of nodes in the grid} \\ p \text{ : nb of segments in the grid} \end{cases}$

- $\Delta_P(\theta \mid w)$: Parameters term of the SC : nb of bits to encode all of the parameter vectors θ_r

  depends on $\begin{cases} \alpha\text{: nb of parameters in } \vec{\theta}_r \\ N_r \text{ : nb of pixels within region region r} \end{cases}$

- $\Delta_L(s \mid \theta, w)$: Data entropy term of the SC knowing partition w and parameters θ for all regions.

The MDL Principle (1/3)

- **Information Theory issue** : transmit the shortest message to describe the image through its « **homogeneous regions**»

- The $N_x \times N_y$ pixels image (random field with PDF $P_{S,\Omega_r}(s)$) summarizes as

$$s(x,y) = \sum_{r=1}^{R} a_r(x,y) \cdot \delta(w(x,y), r)$$

Where

$a_r(x,y)$: are the $N_x \times N_y$ image pixels

$\delta(a,b) = \begin{cases} 1, & a = b \\ 0, & \text{sinon} \end{cases}$: is the Kronecker symbol

w : is the partitionning function, $w(x,y) = r \quad \Leftrightarrow \quad (x,y) \in \Omega_r$

$\Omega_r, r \in [1, R]$: are the region images

Stochastic Complexity Terms (1/3)

- Data Entropic Code Length (ΔL)

$$\Delta_L\left(\mathrm{s}|\widehat{\theta},\mathbf{w}\right)=\sum_{r=1}^{R}\Delta_r\left(\widehat{\theta}_r\right)=\sum_{r=1}^{R}-\mathcal{L}_e\left(\Omega_r|\widehat{\theta}_r\right) \tag{4.8}$$

Galland [53] demonstrated that for a Gamma distribution with same known order L for all of the regions,

$$\Delta_L\left(\mathrm{s}|\widehat{\theta},\mathbf{w}\right)=L\sum_{r=1}^{R}N_r\ln\left(\widehat{\mu}_r\right)+K\left(\mathrm{s},L\right) \tag{4.10}$$

where N_r is the number of pixels in region r and $K\left(\mathrm{s},L\right)$ is independent from the partition $\mathbf{w}$ but depends on the image itself and on the order L of the Gamma distribution, and could therefore be of importance when performing several trials of L as described in A:

$$K\left(\mathrm{s},L\right)=N\cdot\left[\ln\Gamma\left(L\right)+L\cdot\left(1-\ln L\right)\right]-\left(L-1\right)\cdot\sum_{x=1}^{N_x}\sum_{y=1}^{N_y}\ln\mathrm{s}\left(x,y\right) \tag{4.11}$$

where $N=N_xN_y$ is the number of pixels of the whole image.

Stochastic Complexity Terms (2/3)

- Statistical Model Parameters code length (ΔP)

Let α be the dimension of the vector of parameters ($\alpha = 2$ for the Gamma law); so one has to encode α scalar parameters for each region. Since each of them are estimated on a sample of N_r pixels, an approximation of the code length associated to the parameter vector is $\alpha \ln \sqrt{N_r}$ according to [55] and therefore :

$$\Delta_P(\theta|\mathbf{w}) = \sum_{r=1}^{R} \frac{\alpha}{2} \ln (N_r) \tag{4.12}$$

- ➔ *:* known parameters
- *R :* variable but known
- ***N_r :* need to be estimated for each region**

Stochastic Complexity Terms (3/3)

- Geometrical partition code length (Delta_G)

The code length to encode the grid is then deduced from 4.13

$$\Delta_G(\mathbf{w}) = n_{SN} \cdot (\ln N + \ln p) + p\left(\ln 2\widehat{m}_x + \ln 2\widehat{m_y} + 2\right) + \ln p \qquad (4.22)$$

p = total number of segments : is variable but known

m_x, m_y = mean segment length in both axes : **shall constantly be updated**

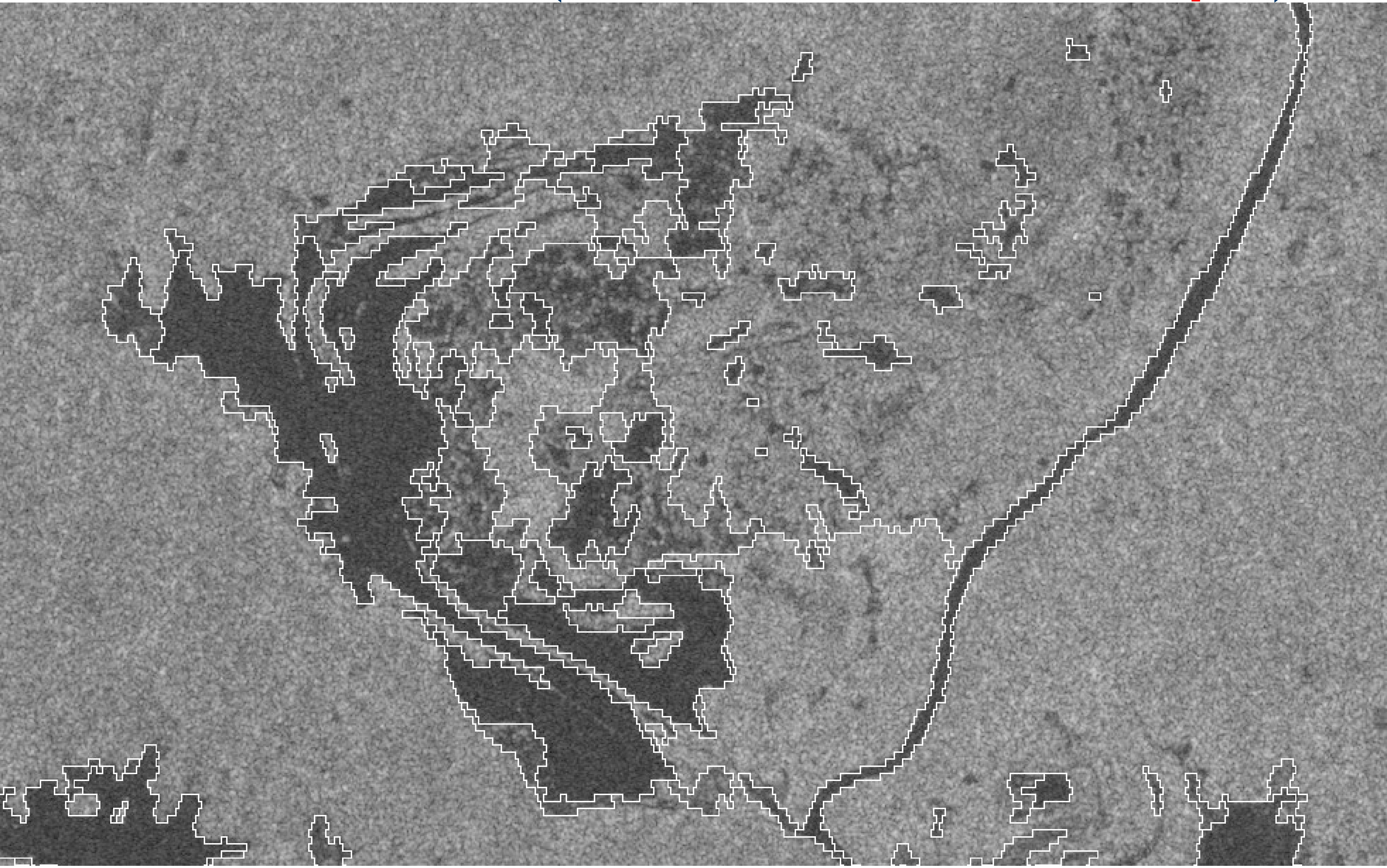
Burman Lake

- ENL = 4.9 (IW GRD HR) → L=5
- Image Size (Az x Rg) in pixels : 3424 x 2760
- Pixel size (Az x rg) : 10x10 m
- Proc. Time (8x8 pixels grid): 1237s (20min)
- Proc. Time (5x5 pixels grid): 1803s (30min)

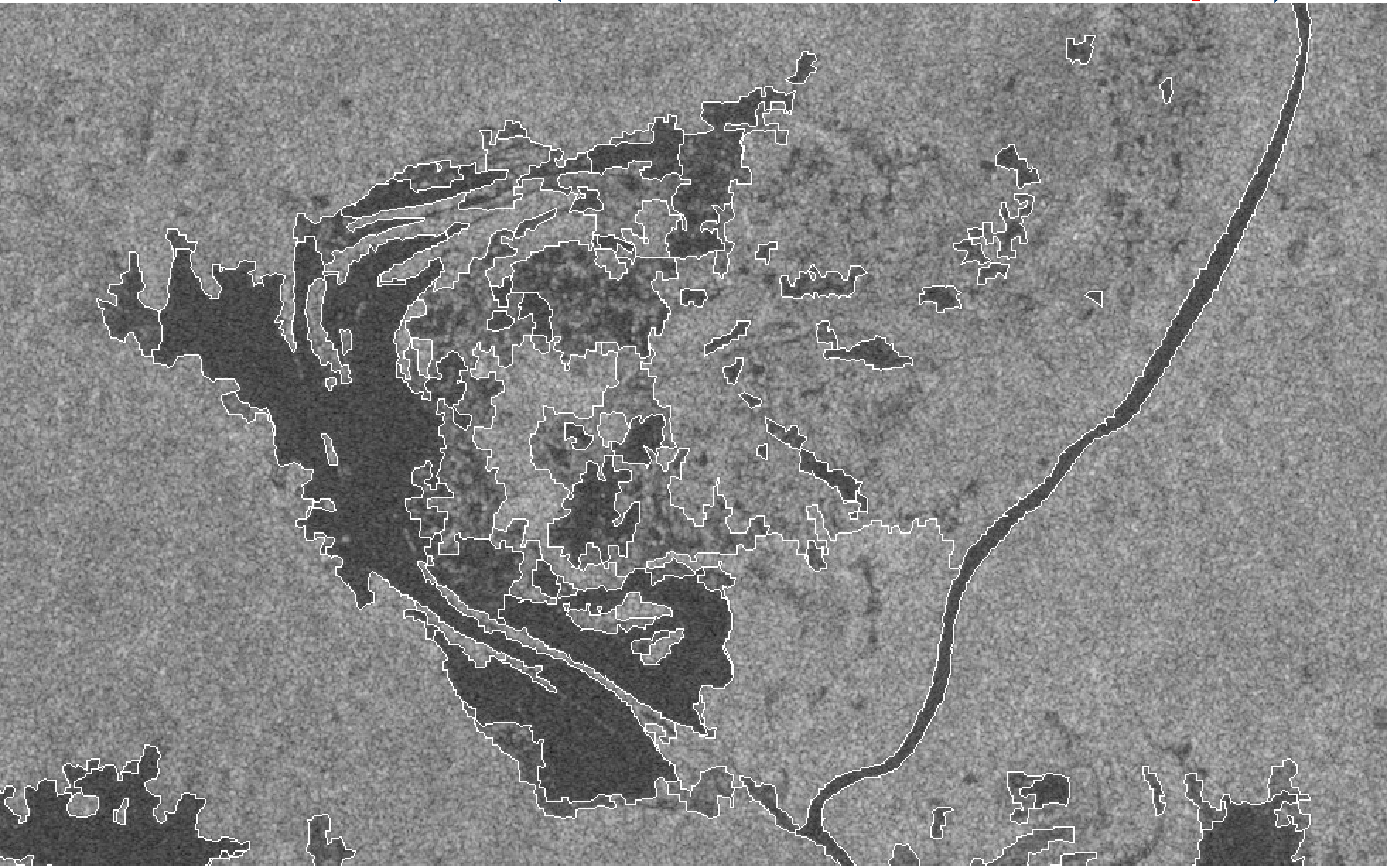
Processed on a core i7 laptop

Burman Lake (8x8, VH, zoom, initial)

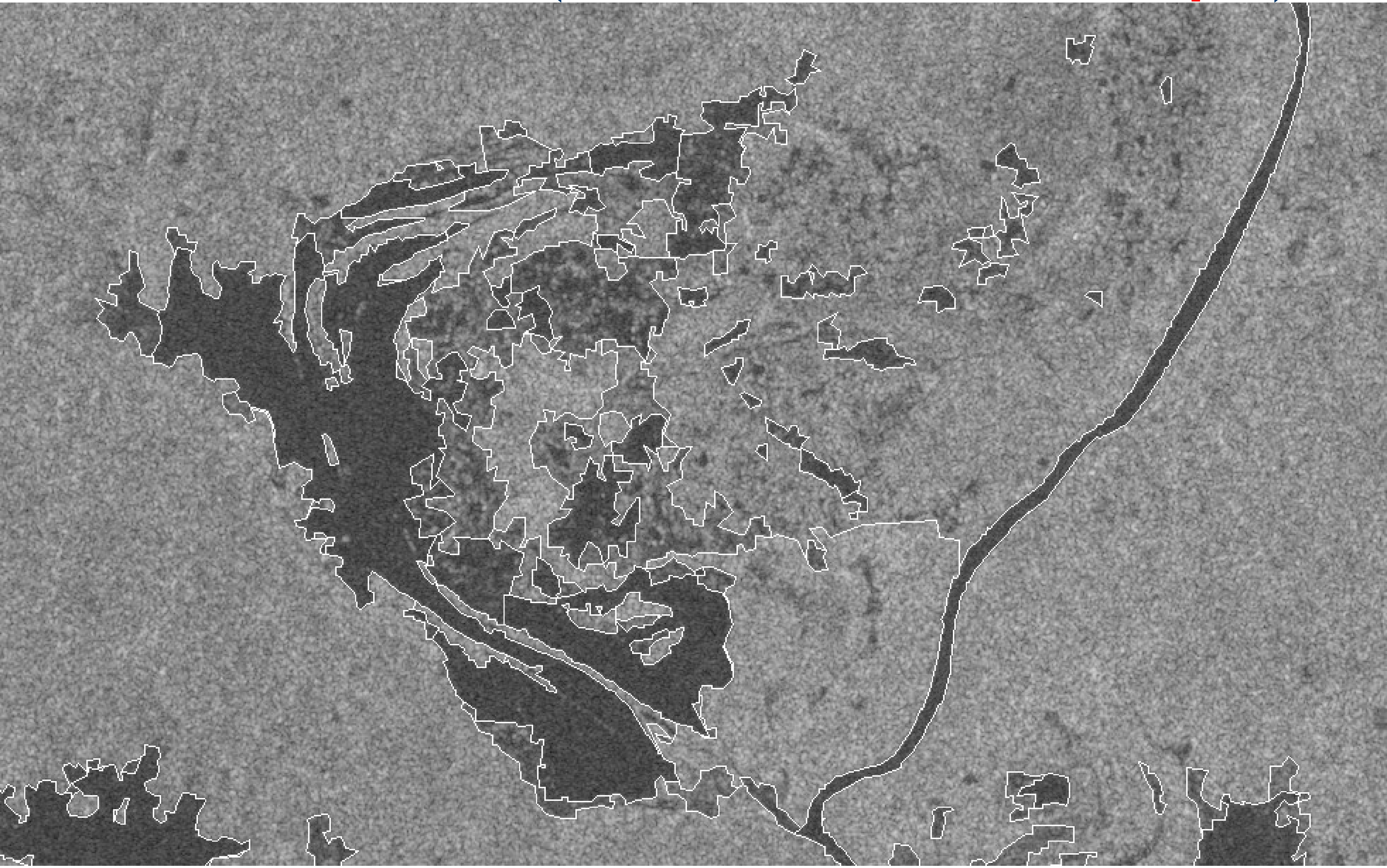
Burman Lake (5x5, VH, zoom, loop1)



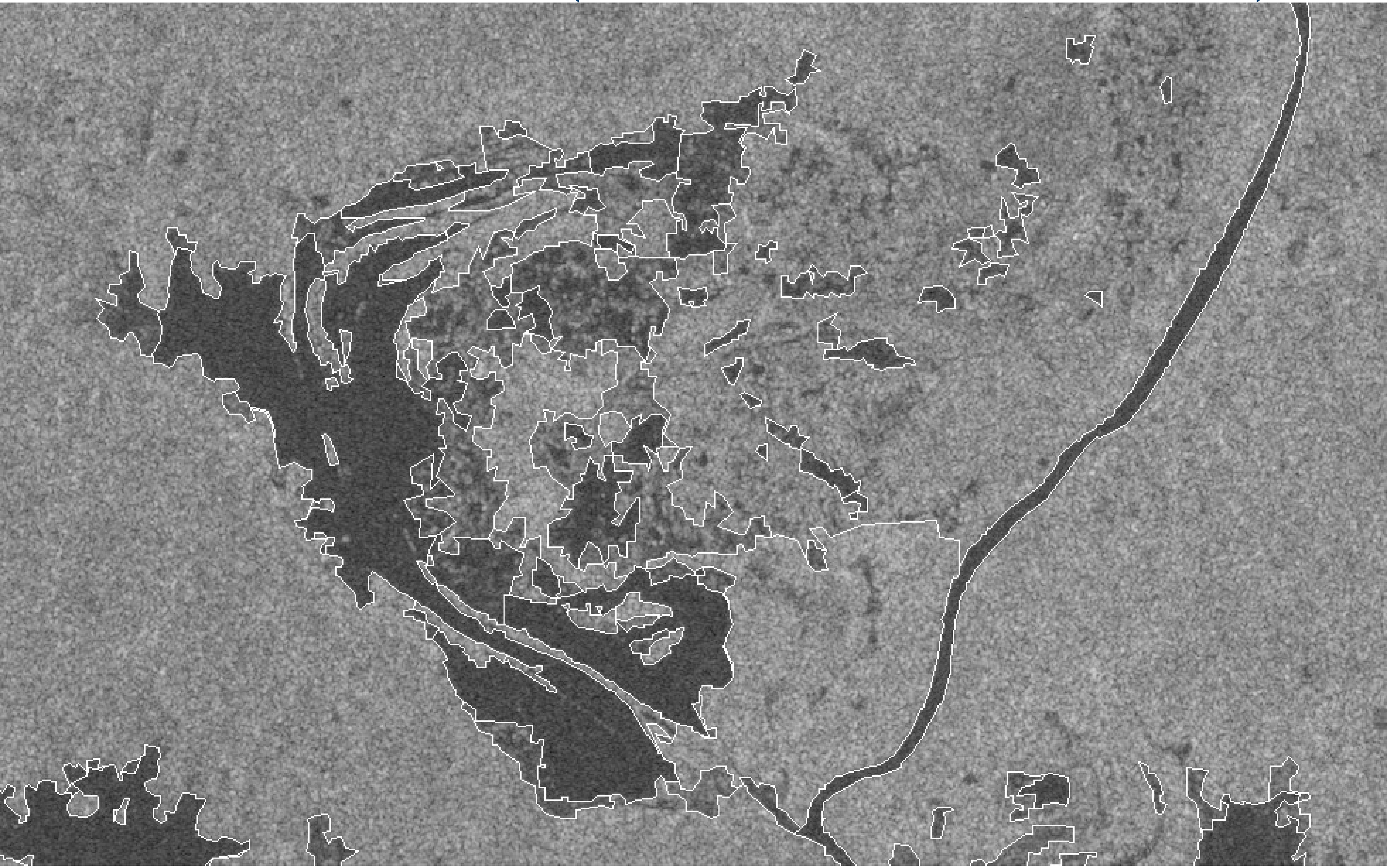
Burman Lake (5x5, VH, zoom, loop2)



Burman Lake (5x5, VH, zoom, loop3)

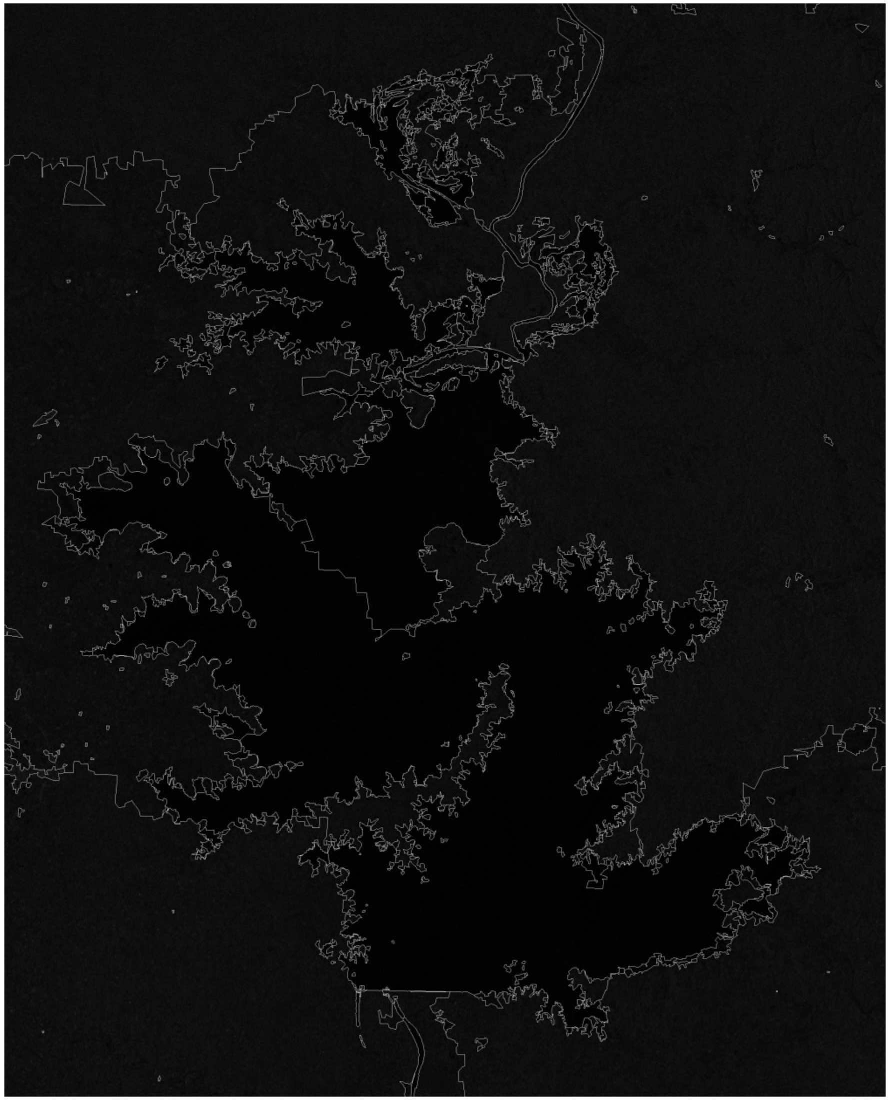


Burman Lake (5x5, VH, zoom, final)



Burman Lake (5x5, VH, global, final)

Burman Lake (5x5, VV, global, final)



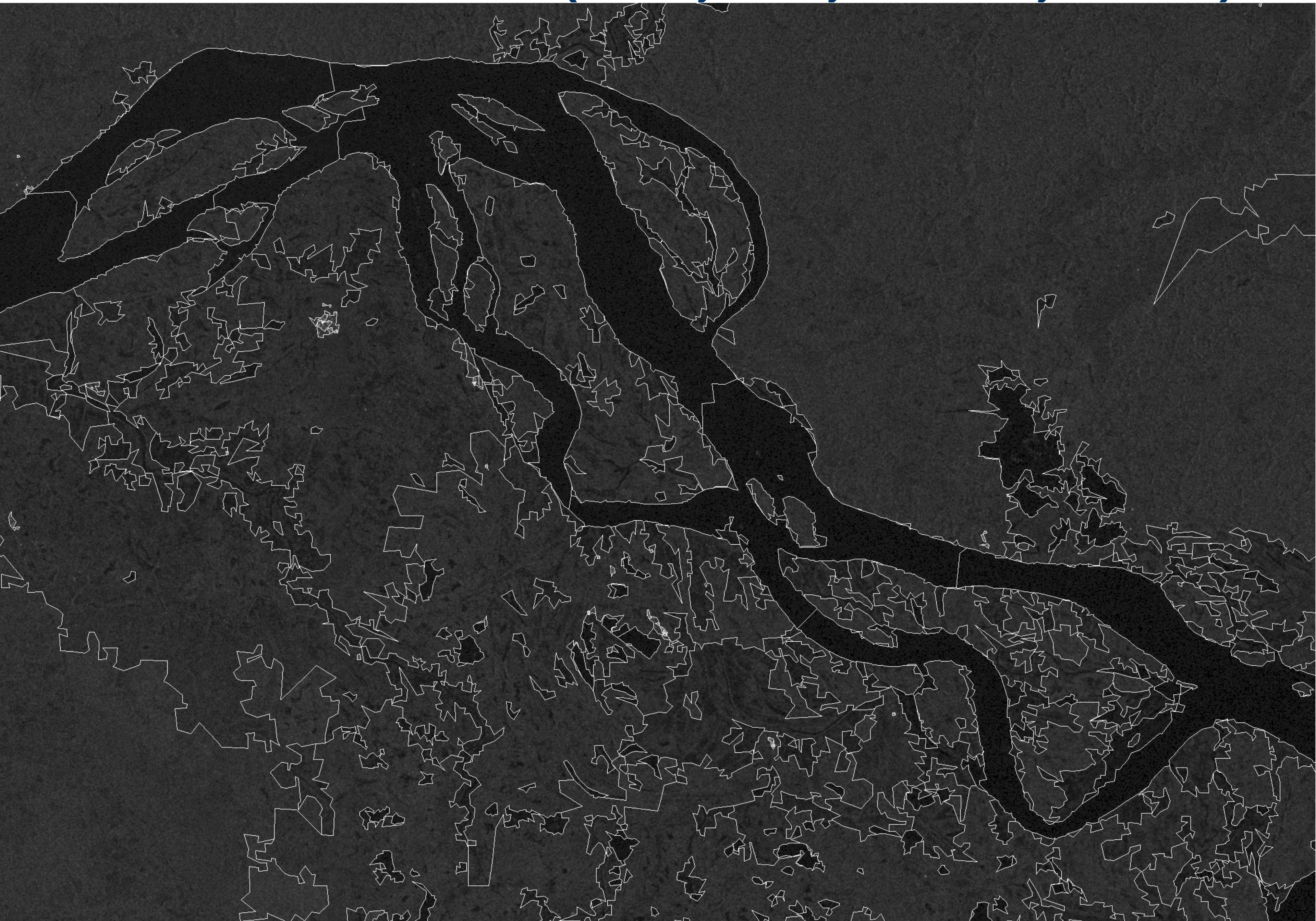
Burman Lake (existing Google and SWBD)

Burman River

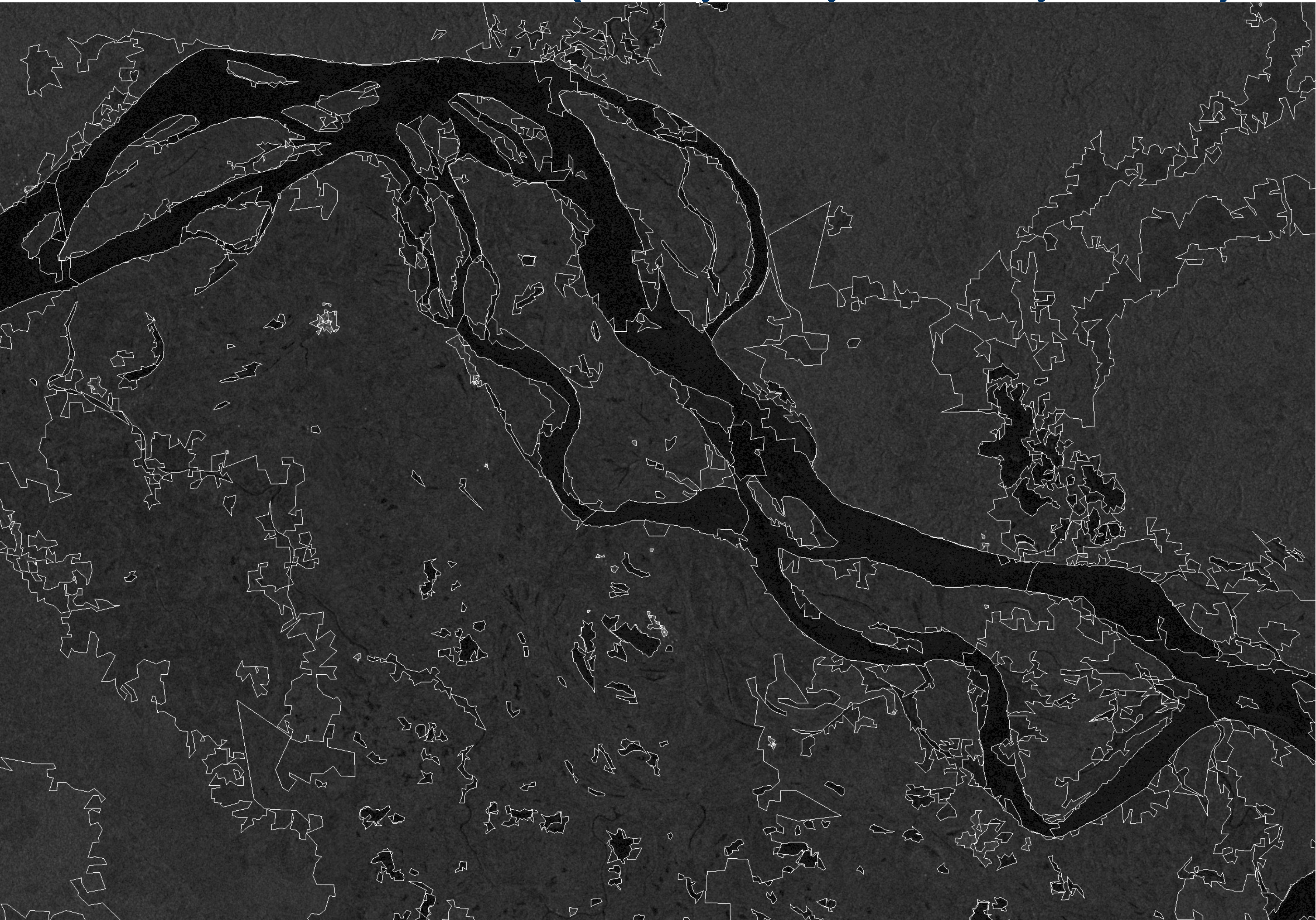
- ENL = 4.9 (IW GRD HR) → L=5
- Image Size (Az x Rg) in pixels : 1614 x 4164
- Pixel size (Az x rg) : 10x10 m
- Proc. Time (8x8 pixels grid): 572 s (10 min)
- Proc. Time (5x5 pixels grid): 2550 s (42min)

Processed on a core i7 laptop

Burman River (8x8, VH, zoom, final)



Burman River (8x8, VV, zoom, final)



Conclusions

- ❑ Was a first try with L=ENL, but could be refined
- ❑ Speed issue
- ❑ The differences in the two polars
  - VH has strong water / non water contrast
  - VV has lower water / non water contrast but should help over windy lakes (and current in rivers)
- ❑ Still over segmented : needs a robust classification step

Perspectives and Follow On

- **Speed issue** :
  - pre-process with a **fast over-segmenting algorithm**

- **Robustness** :
  - extend the Stochastic Complexity criterion to both polar :

$$SC = \Delta G + \Delta P(vh) + \Delta P(vv) + \Delta L(vh) + \Delta L(vv)$$

- **Final result**:
  - post-process with a learning /classification stage

Many Thanks for your Attention

A step because we juste address the question of image segmentation but not the classification (or learning stage) to discrinate water from the rest. There may also be several classes of water (windy, quiet, eutrophised, shallow …)

Water masks from Sentinel-1 : Why ?

- **dynamic water masks (every 12 days) : all weather, night and day, high resolution !**
- Help **monitor water resource variations** over time
- Help **future missions** on Surface Waters (SWOT)
- Help evaluate **Retrackers' performance** in altimetry for hydrology
- Provide a priori information to retrackers
- Provide information to **better analyse radar altimeter waveforms** in hydrology
- **Combine radar altimetry and radar imaging** to derive **Hypsometric Curves** (Height-Surface and Height-Surface-Volume) curves as well as **bathymetry** over lakes in hydrology
- …

Selected Algorithm

- The **Minimum Description Length Automated Segmentation Grid** (MDLSAG), which results from a long evolution :

1. *Statistical Region based Active Contours* (several authors, statistical framework)
2. *Statistical **Polygonal Snakes*** (Germain 1996, **Max. Likelihood** based allowing 2 classes only)
3. *Statistical Polygonal Active Grid* (Germain 2001, Bayesian, **multi-region** generalization of the Polygonal snake)
4. The **Minimum Description Length (MDL)** principle, introduced by Rissanen in 1978 was refreshed by Andrew Barron in 1998)
5. **MDL + Polygonal Active** *Grid* , PhD thesis Galland, 2003
6. In 2009, Galland et al. revisited the Automated MDL Polygonal Grid with the **assumption that intensity pixels of SAR images are appropriately modeled by Gamma PDF** with
   - **same order L** all over the SAR image
   - **mean intensity** depending on the regions.

SAR images Stochastic Model (1/2)

❑ Multiplicative Noise

SLC SAR Image ≡ 1 occurrence of random field $S_\lambda(x,y)$ all of the r.v. depend on random event :

Under basic assumptions (Goodman), the **r.v**. $S_\lambda(x,y)$ (**pixel intensity**) in an area with **constant mean reflectivity (μ)** follow an **exponential law** :

$$P_S(s) = \begin{cases} \frac{1}{\mu} \cdot e^{-\frac{s}{\mu}}, s > 0 \\ 0 \text{ sinon} \end{cases} \Rightarrow \begin{cases} E[S] = \mu \\ \text{var}[S] = \mu^2 \end{cases} \qquad b \quad S_\lambda(x,y) = \mu \cdot b(x,y)$$

=> ∃ **multiplicative noise** , $P_{S,x,y}(s)$

If we could repeat the random experiment then we could learn the pixels' PDF ,
instead we are going to assume **homogeneity** and compute a « spatial mean value »

SAR images Stochastic Model (2/2)

❑ **How to define « Homogeneous Region » ?**

- Strict definition at order 1 (impractical from 1 image occurence) :

$$\left\{(x,y), P_{S,x,y}(s) = P_S(s)\right\}$$

- Wide sense definition (2nd order over neighborhoud, applicable to 1 image occ.) : $\left\{(x,y), E\left[S_\lambda(x,y)\right] = E[S] \text{ et } E\left[S_\lambda(x,y) \cdot S_\lambda(x+\Delta x, y+\Delta y)\right] = f(\Delta x, \Delta y)\right\}$

In practice (SLC to GRD) → L multi-looking → Modification pixels ' PDF :

- Exponential Law → **Gamma Law** (Assumption-1 from Galland, 2009) :

$$P_{\bar{S}}(\bar{s}) = \begin{cases} \dfrac{L^L}{\mu \cdot \Gamma(L)} \cdot \left(\dfrac{\bar{s}}{\mu}\right)^{L-1} e^{-L\frac{\bar{s}}{\mu}}, \ \bar{s} > 0 \\ 0 \text{ sinon} \end{cases} \Rightarrow \begin{cases} E\left[\bar{S}\right] = \mu \\ \operatorname{var}\left[\bar{S}\right] = \dfrac{\mu^2}{L} \end{cases}$$

- In practice **L = ENL** (example : L = 4.9 on Sentinel-1 IW HR GRD)

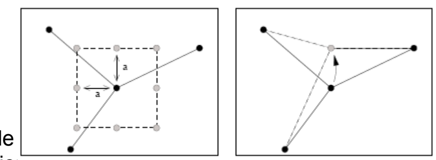
The MDL Principle (2/3)

Initial partition = **grid** = set of **NODES** linked together through **SEGMENTS** to define rectangular **REGIONS**

Iteratively perform **3 types of grid modifications** and **keep the changes that lower the « length of the description message » (Stochastic Complexity)**

1. Merge : test all possible region merges 2 by 2

2. **Move**: Try to Move the nodes in 8 directions (amplitude is a function of connected segments' length) with decrease.



(a) (b)

3. **Remove** : Go through each node ... other nodes) and evaluate its SC reductio

The MDL Principle (3/3)

- **Stochastic Complexity** (SC) as derived by Galland :

- $\Delta_G(w)$: Geometric term of the SC : nb of bits to encode partition w,

  depends on $\begin{cases} k \text{ : nb of nodes in the grid} \\ p \text{ : nb of segments in the grid} \end{cases}$

- $\Delta_P(\theta \mid w)$: Parameters term of the SC : nb of bits to encode all of the parameter vectors $\vec{\theta}_r$

  depends on $\begin{cases} \alpha\text{: nb of parameters in } \vec{\theta}_r \\ N_r \text{ : nb of pixels within region region r} \end{cases}$

- $\Delta_L(s \mid \theta, w)$: Data entropy term of the SC knowing partition w and parameters θ for all regions.

The MDL Principle (1/3)

- **Information Theory issue** : transmit the shortest message to describe the image through its « **homogeneous regions**»

- The $N_x \times N_y$ pixels image (random field with PDF $P_{S,\Omega_r}(s)$) summarizes as

$$s(x,y) = \sum_{r=1}^{R} a_r(x,y) \cdot \delta(w(x,y), r)$$

Where

$a_r(x,y)$: are the $N_x \times N_y$ image pixels

$\delta(a,b) = \begin{cases} 1, & a = b \\ 0, & \text{sinon} \end{cases}$: is the Kronecker symbol

w : is the partitionning function, $w(x,y) = r \Leftrightarrow (x,y) \in \Omega_r$

$\Omega_r, r \in [1, R]$: are the region images

Stochastic Complexity Terms (1/3)

- Data Entropic Code Length (ΔL)

$$\Delta_L\left(\mathbf{s}|\widehat{\theta},\mathbf{w}\right)=\sum_{r=1}^{R}\Delta_r\left(\widehat{\theta}_r\right)=\sum_{r=1}^{R}-\mathcal{L}_e\left(\Omega_r|\widehat{\theta}_r\right) \tag{4.8}$$

Galland [53] demonstrated that for a Gamma distribution with same known order L for all of the regions,

$$\Delta_L\left(\mathbf{s}|\widehat{\theta},\mathbf{w}\right)=L\sum_{r=1}^{R}N_r\ln\left(\widehat{\mu}_r\right)+K\left(\mathbf{s},L\right) \tag{4.10}$$

where N_r is the number of pixels in region r and $K(\mathbf{s},L)$ is independent from the partition $\mathbf{w}$ but depends on the image itself and on the order L of the Gamma distribution, and could therefore be of importance when performing several trials of L as described in A:

$$K\left(\mathbf{s},L\right)=N\cdot\left[\ln\Gamma\left(L\right)+L\cdot\left(1-\ln L\right)\right]-\left(L-1\right)\cdot\sum_{x=1}^{N_x}\sum_{y=1}^{N_y}\ln\mathbf{s}\left(x,y\right) \tag{4.11}$$

where $N=N_xN_y$ is the number of pixels of the whole image.

Stochastic Complexity Terms (2/3)

❑ Statistical Model Parameters code length (ΔP)

Let α be the dimension of the vector of parameters ($\alpha = 2$ for the Gamma law); so one has to encode α scalar parameters for each region. Since each of them are estimated on a sample of N_r pixels, an approximation of the code length associated to the parameter vector is $\alpha \ln \sqrt{N_r}$ according to [55] and therefore :

$$\Delta_P(\theta|\mathbf{w}) = \sum_{r=1}^{R} \frac{\alpha}{2} \ln (N_r) \tag{4.12}$$

- ➔ *:* known parameters
- *R :* variable but known
- ***N_r :* need to be estimated for each region**

Stochastic Complexity Terms (3/3)

- Geometrical partition code length (Delta_G)

The code length to encode the grid is then deduced from 4.13

$$\Delta_G(\mathbf{w}) = n_{SN} \cdot (\ln N + \ln p) + p\left(\ln 2\widehat{m}_x + \ln 2\widehat{m_y} + 2\right) + \ln p \qquad (4.22)$$

p = total number of segments : is variable but known
m_x, m_y = mean segment length in both axes : **shall constantly be updated**

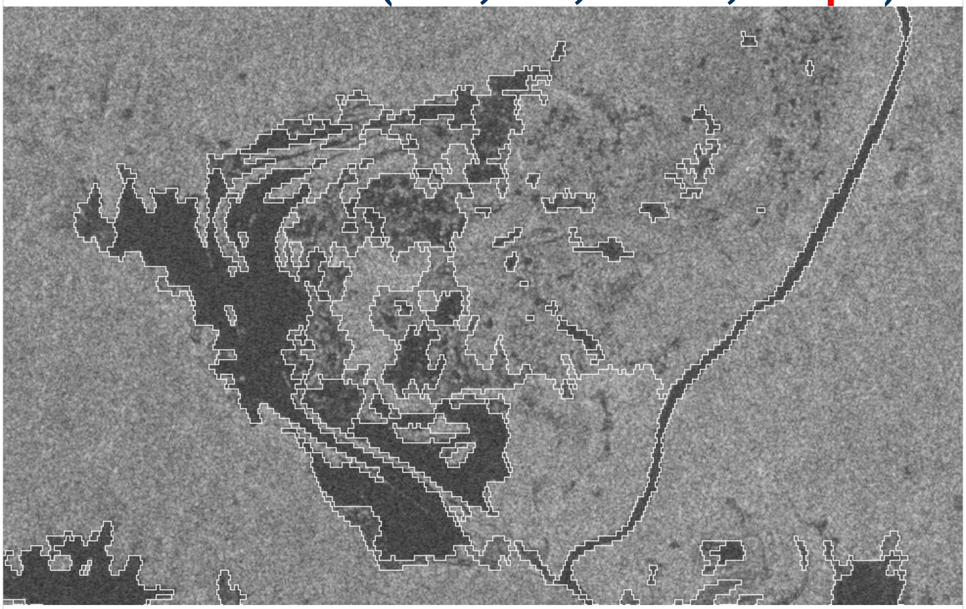
Burman Lake

- ENL = 4.9 (IW GRD HR) → L=5
- Image Size (Az x Rg) in pixels : 3424 x 2760
- Pixel size (Az x rg) : 10x10 m
- Proc. Time (8x8 pixels grid): 1237s (20min)
- Proc. Time (5x5 pixels grid): 1803s (30min)

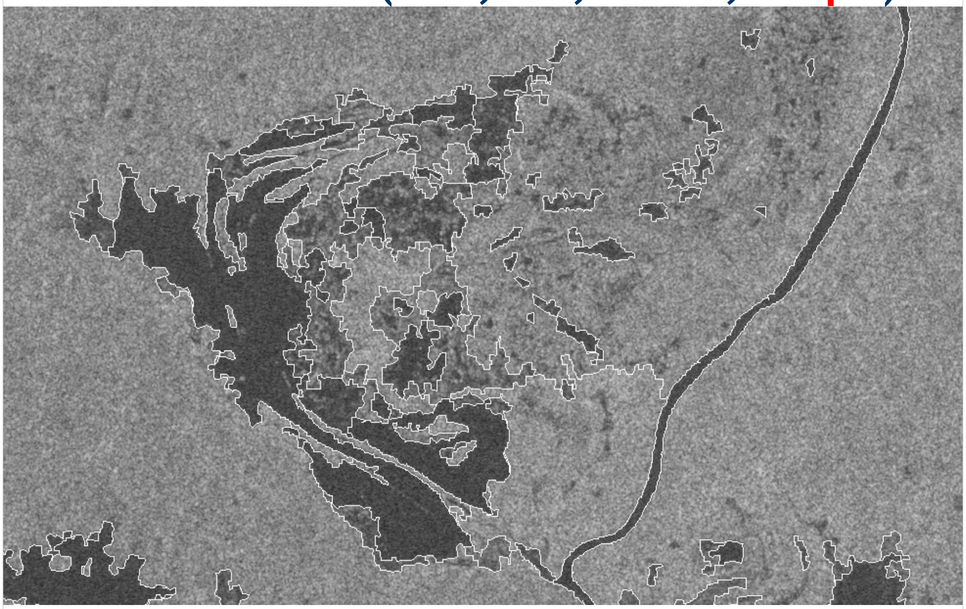
Processed on a core i7 laptop

Burman Lake (8x8, VH, zoom, initial)

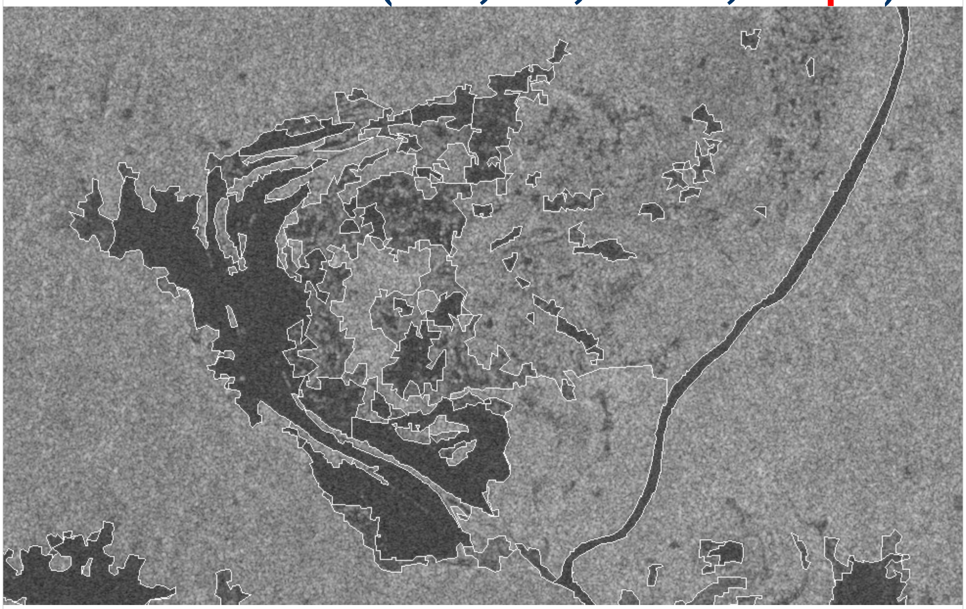
Burman Lake (5x5, VH, zoom, loop1)



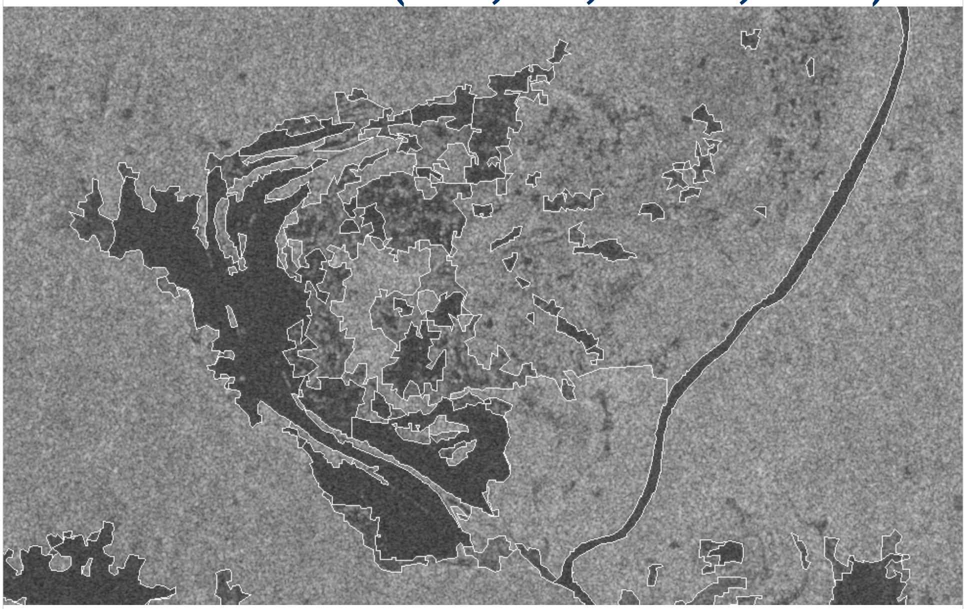
Burman Lake (5x5, VH, zoom, loop2)



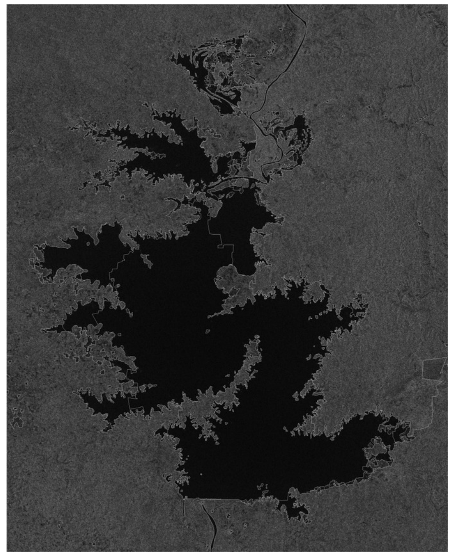
Burman Lake (5x5, VH, zoom, loop3)



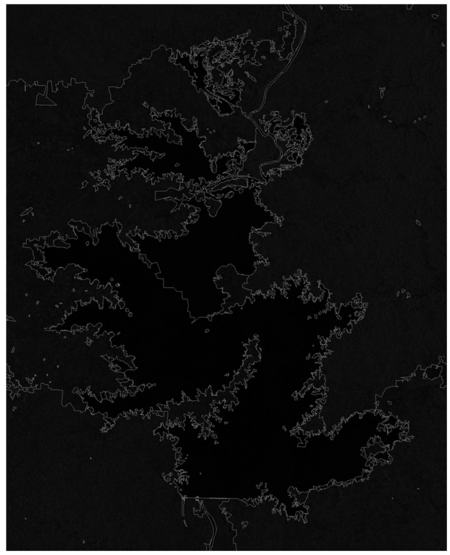
Burman Lake (5x5, VH, zoom, final)



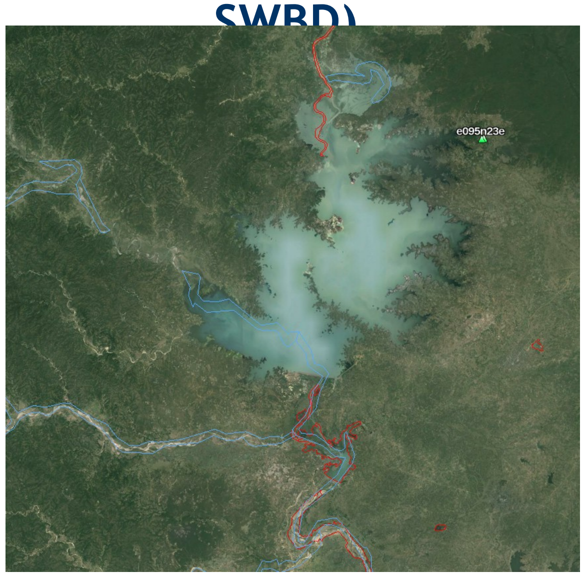
Burman Lake (5x5, VH, global, final)



Burman Lake (5x5, VV, global, final)



Burman Lake (existing Google and SWBD)

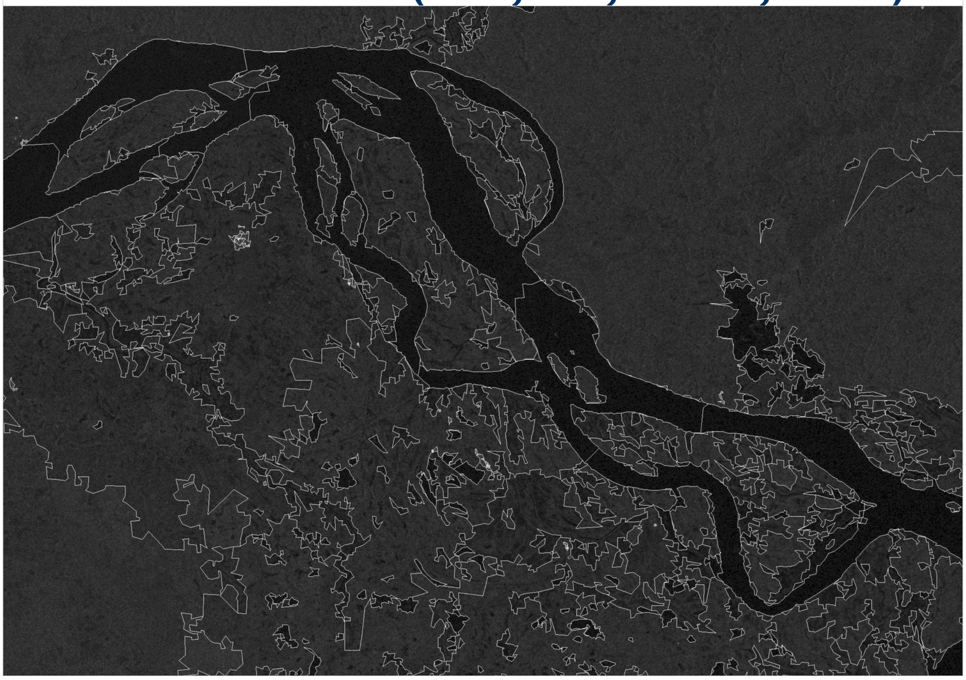


Burman River

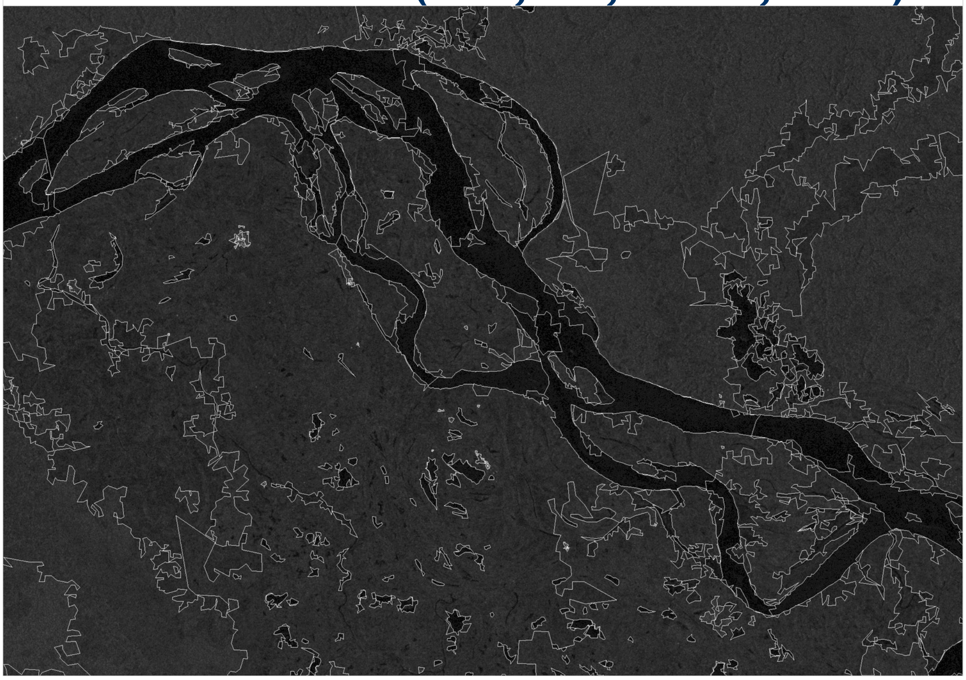
- ENL = 4.9 (IW GRD HR) → L=5
- Image Size (Az x Rg) in pixels : 1614 x 4164
- Pixel size (Az x rg) : 10x10 m
- Proc. Time (8x8 pixels grid): 572 s (10 min)
- Proc. Time (5x5 pixels grid): 2550 s (42min)

Processed on a core i7 laptop

Burman River (8x8, VH, zoom, final)



Burman River (8x8, VV, zoom, final)



Conclusions

- ❑Was a first try with L=ENL, but could be refined
- ❑ Speed issue
- ❑ The differences in the two polars
  - VH has strong water / non water contrast
  - VV has lower water / non water contrast but should help over windy lakes (and current in rivers)
- ❑ Still over segmented : needs a robust classification step

Perspectives and Follow On

- **Speed issue** :
  - pre-process with a **fast over-segmenting algorithm**
- **Robustness** :
  - extend the Stochastic Complexity criterion to both polar :

    $SC = \Delta G + \Delta P(vh) + \Delta P(vv) + \Delta L(vh) + \Delta L(vv)$
- **Final result**:
  - post-process with a learning /classification stage

Many Thanks for your Attention