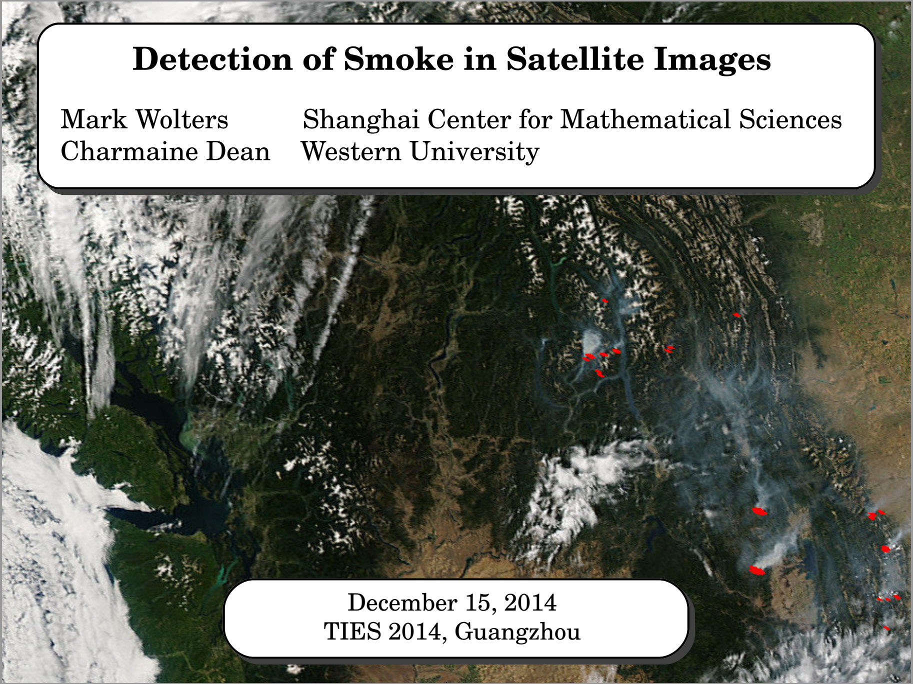
Detection of Smoke in Satellite Images

Mark Wolters — Shanghai Center for Mathematical Sciences
Charmaine Dean — Western University



December 15, 2014
TIES 2014, Guangzhou

Summary

Application

Smoke identification (*binary image segmentation*)

Data

MODIS images (*hyperspectral data*)

Methods

1. Logistic regression

   $\Rightarrow$ with *high-dimensional feature space*

2. Autologistic regression

   $\Rightarrow$ gives *spatial smoothness*
   $\Rightarrow$ but is *computationally challenging*

Application

Why interested in smoke?

Smoke from forest fires has **population health relevance**

Large area, hard to monitor.

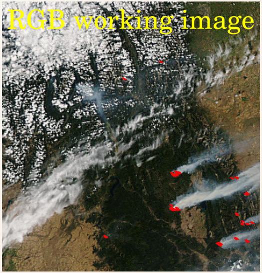
Applications:

- retrospective studies
- model validation
- model initialization, updating

Goal: **binary image segmentation**

- classes: smoke, nonsmoke
- supervised learning

Methodology generally applicable



Problems detecting smoke

Some spectral characteristics of smoke are known, e.g.:

- transparent in middle IR
- absorption in blue & near UV

But:

- aging effects
- smoke and cloud can coexist
- fire-to-fire variability
- “thick” vs. “thin” smoke

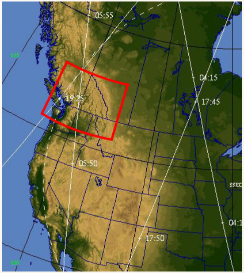
So:

- Get a large data set and see if machine learning can sort it out.

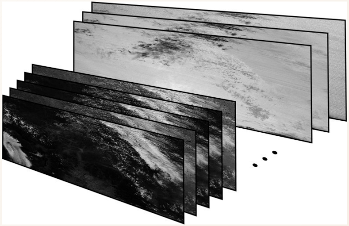
Data

MODIS data

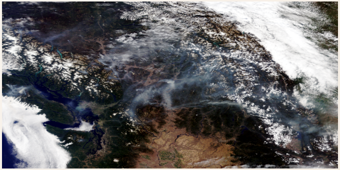
- ROI centered on Kelowna, BC
- 143 images
- 35 usable bands



(http://www.ssec.wisc.edu/datacenter/terra/)

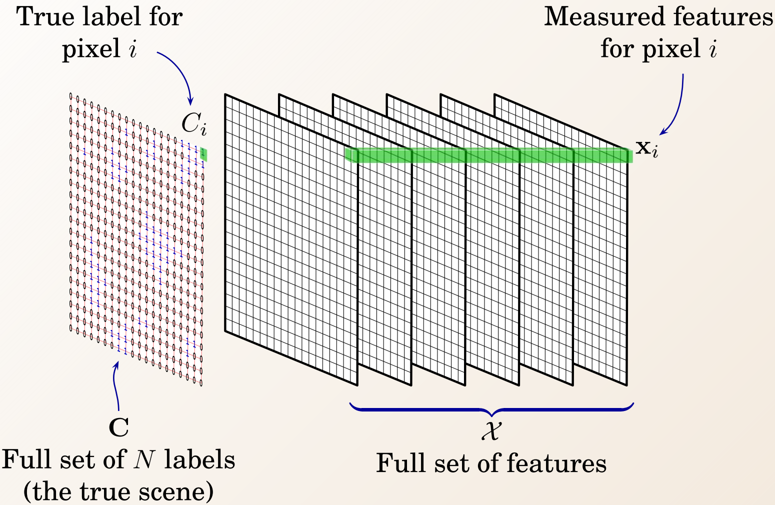


Each image
36 image planes
Bands: 0.4–14μm
About 1.2 Mpix

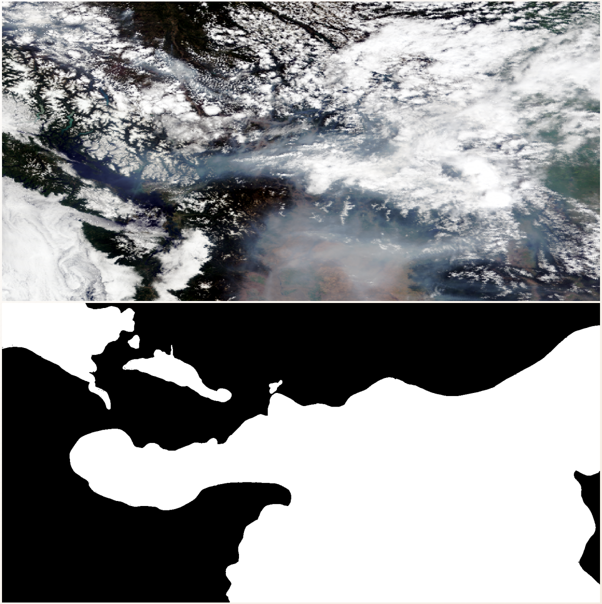


Bands 1, 4, 3 used to produce RGB images

Notation for an image



Obtaining the masks



Use RGB images with fire locations.

Assign each pixel to class 0 (nonsmoke) or class 1 (smoke)

Not a satisfactory process!

143 Images:

- 70 training
- 36 validation
- 37 test

Methods

1. Logistic regression classifier

$\pi_i = \mathrm{P}(\text{pixel } i \text{ is smoke})$ $\qquad$ $\log\left(\dfrac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^T\boldsymbol{\beta}$

x is pixel i's feature vector.

Here, use:

$$\left.\begin{array}{ll} (1) & \text{35 observed bands} \\ (2) & \text{their squares} \\ (3) & \text{their square roots} \\ (4) & \text{interactions among } \{(1), (2)\} \\ (5) & \text{interactions among } \{(1), (3)\} \end{array}\right\} \text{4340 features}$$

Use *genetic algorithm* or *lasso* to find a good small model.

- minimize deviance on validation set

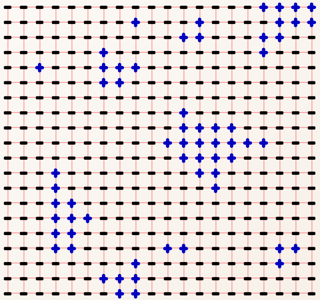
2. Autologistic regression classifier

$\pi_i = \mathrm{P}(\text{pixel } i \text{ is smoke} \mid \text{all other pixels})$

Logit form: $$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^T\boldsymbol{\beta} + \lambda \underbrace{\sum_{j\sim i} c_j}_{\text{sum of neighbors}}$$

unary coeffs

pairwise coeffs



- It's an MRF model on a graph
- Use rectangular grid
- each pixel has 4 neighbours
- *use plus / minus coding*

Joint distribution of C

$$\mathrm{P}(\mathbf{C} = \mathbf{c} | \boldsymbol{\beta}, \lambda) = \frac{1}{Z(\boldsymbol{\beta}, \lambda)} \exp\left(\frac{1}{2} \sum_{i \in \mathcal{V}} \mathbf{x}_i^T \boldsymbol{\beta} c_i + \frac{\lambda}{2} \sum_{(i,j) \in \mathcal{E}} c_i c_j \right)$$

Larger $\mathbf{x}_i^T \boldsymbol{\beta}$ values favor $+1$ (smoke).

$\lambda > 0$ favors locally smooth configurations.

Normalizing constant is **intractable**.

- can't compute likelihood
- can't compute marginals
- estimation and prediction are hard

Usual practice

Estimation

- pseudolikelihood: $\mathrm{PL}(\boldsymbol{\beta}, \lambda) = \prod_{i=1}^{N} \pi_i$
- sampling/MCMC approaches

Prediction for new images

- MAP: find $\mathbf{c}^*$ to maximize $\mathrm{P}(\mathbf{C} = \mathbf{c}^* | \hat{\boldsymbol{\beta}}, \hat{\lambda})$

  $\Rightarrow$ graph-cut methods available (sometimes)
  $\Rightarrow$ is it what we really want?

Claim: using plus/minus coding creates alternatives

Estimation

1. estimate β using independence model ($\lambda = 0$, logistic regression)
2. plug $\hat{\beta}$ into autologistic model
3. choose $\hat{\lambda}$ to minimize prediction error

Prediction

- approximate marginal P(pixel i is smoke) by *Gibbs sampling*

$\Rightarrow$ With $+/-$ coding, this is well-behaved

Gibbs sampler videos

Results

Predictive performance

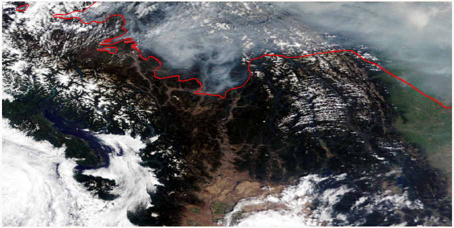
Validation runs: $\hat{\lambda} = 3$ suitable for autologistic.

Prediction errors on the 37 test images:

Model	Error rate (%)		
	overall	nonsmoke	smoke
"everything is nonsmoke"	**10**	0	100
Independence: GA (50 variables)	**8.1**	1.9	64
Independence: lasso (109 variables)	**7.8**	1.2	66
Autologistic (50 variables)	**7.4**	0.89	65
Autologistic (109 variables)	**7.3**	0.63	67

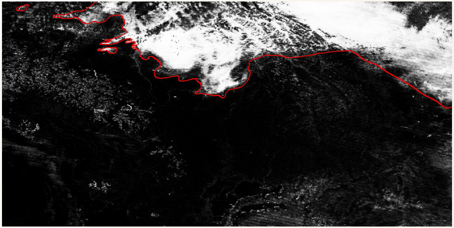
Qualitative assessment

RGB image



1. Smoke-free areas: OK
2. Clouds vs. smoke: OK
3. Snow vs. smoke: OK
4. Spatial smoothing: OK
5. Smoke + Cloud: Problem

Logistic model

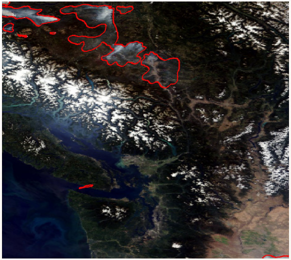


Autologistic model

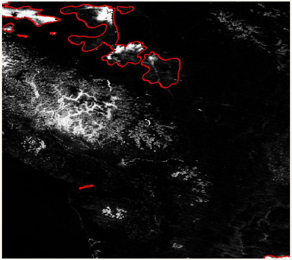
mwolters@fudan.edu.cn

Qualitative assessment (continued)

RGB



Logistic



Autologistic

6. “Thin” smoke: Problem
7. Original masks (training data): Problem

Conclusions

We've made the autologistic model into a relatively painless smoother.

- But is it worth it?

Still need an adequate independence model.

Model-based approach: *many potential extensions and improvements.*

- In the data itself:
  - ⇒ Label smoke more conservatively
  - ⇒ Disregard smoke + cloud pixels
- In the logistic model:
  - ⇒ Use additive model instead of polynomial terms
  - ⇒ Include indicators for ground-cover type
- In an autologistic/spatial model:
  - ⇒ Inlcude more classes (auto*binomial* model)
  - ⇒ Adaptive smoothing: let λ vary with location.

Advertisement

- Got feedback?
- Got data?
- Got applications?
- Got remote sensing expertise?
- Interested in a software package?

Please contact me! *mwolters@fudan.edu.cn*