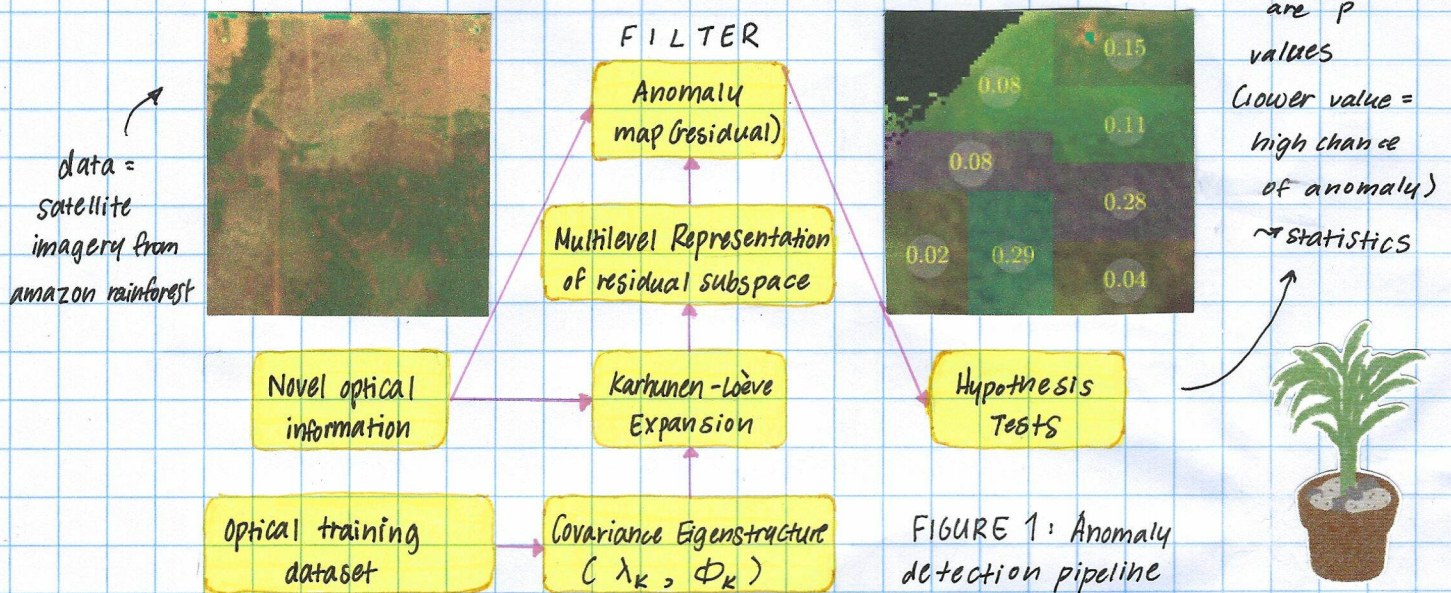


multilevel anomaly detection

FOR VECTOR FIELDS VIA DISTRIBUTION-FREE STOCHASTIC ANALYSIS - DEFORESTATION

Overview of Method:

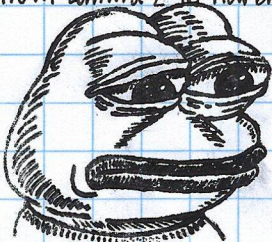
- Method doesn't make assumptions about data & works on vector field data
- Pipeline: optical training data are used to construct the eigenstructure. From the eigenstructure & the domain of the novel optical information, the Karhunen-Loève expansion is built. The KL module produces the truncation expansion of the random vector field of the training dataset. From the truncated KL expansion the multilevel representation of the space is constructed. The next step is to project the novel optical information onto the multilevel basis in the residual space, giving rise to the anomaly map. Hypothesis tests & anomaly magnitudes can then be computed.



Underlying Math & Algorithm:

- Consider the operator $G_V(u)(x) := \int_U \text{Cov}(v(x, \omega), v(y, \omega)) u(y) dy$ for all $u \in L^2(U; \mathbb{R}^d)$.

From Lemma 2 in Harbrecht et al. (2016) the operator $G_V: L^2(U; \mathbb{R}^d) \rightarrow L^2(U; \mathbb{R}^d)$ is a



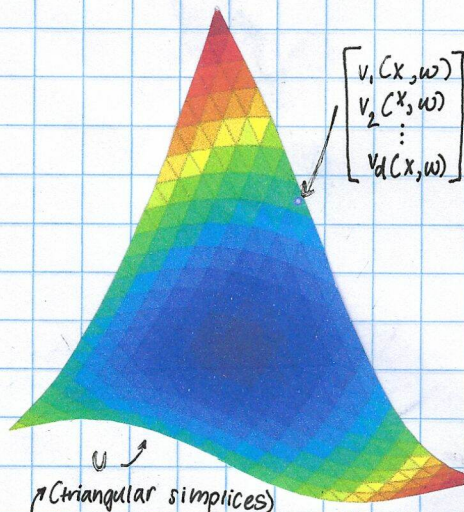
non-negative symmetric trace class operator. From Theorem 1 in Harbrecht et al. (2016) there exists an orthonormal set of eigenfunctions $\{\Phi_k\}_{k \in \mathbb{N}}$, where $\Phi_k \in L^2(U; \mathbb{R}^d)$, and eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ such that $G_V \Phi_k = \lambda_k \Phi_k$ for all $k \in \mathbb{N}$. We can now form the vector field KL expansion.

Theorem 2: Suppose $v \in L^2_p(\Omega; L^2(U; \mathbb{R}^d))$ then $v(x, \omega) = \mathbb{E}[v(x, \omega)]$

+ $\sum_{k \in \mathbb{N}} \lambda_k^{1/2} \Phi_k(x) \gamma_k(\omega)$, where $\gamma_k(\omega) = \frac{1}{\sqrt{\lambda_k}} \int_U (v(x, \omega) - \mathbb{E}[v(x, \omega)])^T \Phi_k(x) dx$, $\mathbb{E}[\gamma_k \gamma_l] = \delta_{kl}$ and $\mathbb{E}[\gamma_k] = 0$ for all $k, l \in \mathbb{N}$. The KL expansion has the useful property of being optimal in the set of all product expansions. It's currently restricted to vectors, but can apply to a continuous space. Mean = 0, & covariance, Variance = 1 & nominal set.

Underlying Math & Algorithm: continued

→ Algorithm works by constructing a kd-tree from the triangular simplices in $\mathcal{T}$.

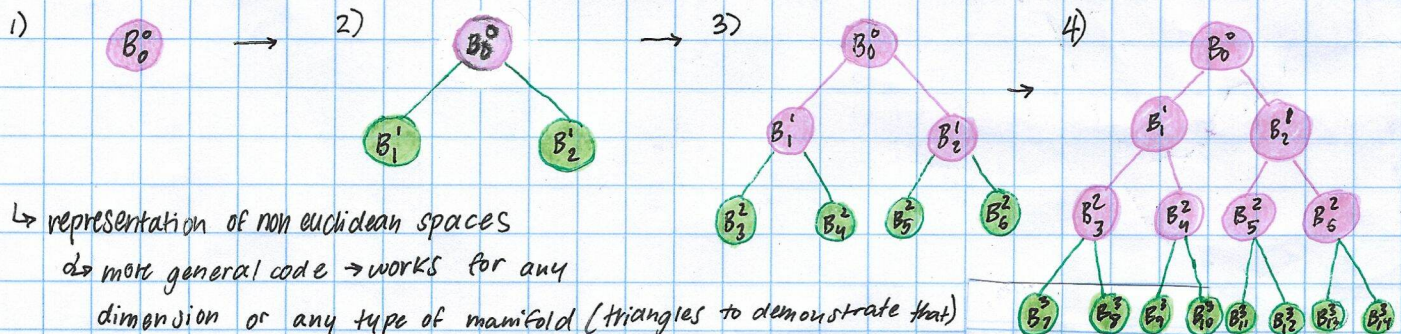


↑ (triangular simplices)

FIGURE 2: Surface domain U constructed from 2-simplices (triangles) in $\mathbb{R}^3$. The vector random field $v(x, w) \in L^2(\Omega; L^2(U; \mathbb{R}^d))$

is defined over the domain U . As an example this could be satellite multi-spectral data over land.

→ Construction of kd-tree from



→ representation of non euclidean spaces

↳ more general code → works for any

dimension or any type of manifold (triangles to demonstrate that)

→ As M increases, not only is the approximation error of the KL expansion reduced & dominated by the sum of eigenvalues, but the variance of the coefficients $d_{\alpha}^{l,k}$ for the corresponding cell B_k^l is also controlled by the same quantity. We shall use this property to construct a reliable hypothesis for detection of anomalous signals in any of the cells $B_k^l \in \mathcal{B}^l$ for $l = 0, \dots, n$.

→ Theorem (Detection: Hypothesis test): Suppose that $u(x, w) = v(x, w) + w(x, w)$ &

$H_0: u(x, w) = v(x, w)$, $H_a: u(x, w) \neq v(x, w)$. Let $1 \geq \alpha \geq 0$ be the significance

level. If null hypothesis H_0 is true: it follows that (a) $\mathbb{P}(|d_{\alpha}^{l,k}| \geq \alpha^{-1/2} \sum_{i=1}^M \lambda_i) \leq \alpha$.

& b) for any cell $B_k^l \in \mathcal{B}$, we have $\mathbb{P}(\sum_{p \in \mathcal{P}_k^l} (d_{\alpha}^{l,k} c_{p,k}^l)^2 \geq \alpha^{-1} \sum_{i=1}^M \lambda_i \sum_{p \in \mathcal{P}_k^l} c_{p,k}^l)^2) \leq \alpha$.

where $b_{i,k}^{l,k} := \int_U \Phi_i^T \mathcal{P}_p^{l,k} dx$.

→ p values produced based on markov inequality

→ decomposition of each vector into other vectors & picking up coefficients with the D values, size of d coefficients correspond to size of anomaly

Underlying Math & Algorithm: Continued

- ↳ With this hypothesis test the coefficients $d_p^{b,k}$ can be used as detectors of anomalous signals in the cell B_K^b . Here are the key features of this detector:
 - ↳ decay of the eigen values with M controlling the sharpness of the bound.
 - ↳ For validity of the hypothesis test only a good estimate of the covariance function is needed.
 - ↳ No assumptions on independence nor underlying distribution (Normal, Poisson, etc.) of the data.

Experimental Results:

- ↳ detect differences within the image, each pixel has different bands for that pixel's location, creates vector field where each pixel is a vector, tested on both scalar & vector field

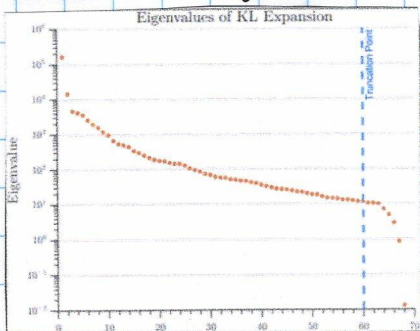
- ↳ Potta comes from Sentinel-2 satellite imagery of the Amazon

- ↳ 71 days of imagery used as a training set

- ↳ Scalar process uses Enhanced Vegetation Index (EVI) data

- ↳ Vector process uses 6 optical bands Blue, green, red, near infra-red, shortwave infrared 1, shortwave infrared 2

- ↳ Treat first 60 eigenfunctions as principal components



60 is truncation point

- ↳ grey circle = size of anomaly
- ↳ yellow # = pval.
- ↳ (if no #, no anomaly chance)



Training Image, Last Day

SCALAR DATA RESULTS

- ↳ Raw hypothesis test on 3 days of data

- ↳ H_0 = no anomalous process, H_A = anomalous process

- ↳ Day 3344 → no significant change from training set

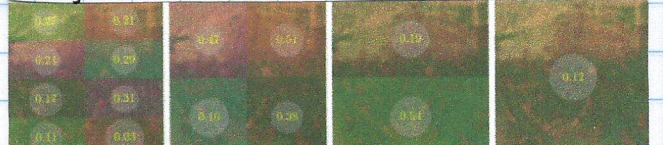
- ↳ Day 3484 → noticeable deforestation

- ↳ Day 3704 → new vegetation grows to fill in cleared areas

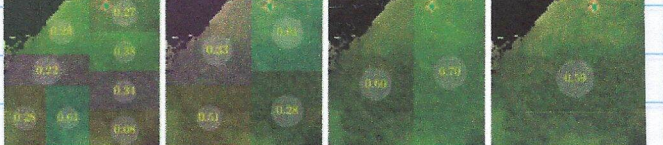
W_3^{3344} W_2^{3344} W_1^{3344} W_0^{3344}



W_3^{3484} W_2^{3484} W_1^{3484} W_0^{3484}



W_3^{3704} W_2^{3704} W_1^{3704} W_0^{3704}

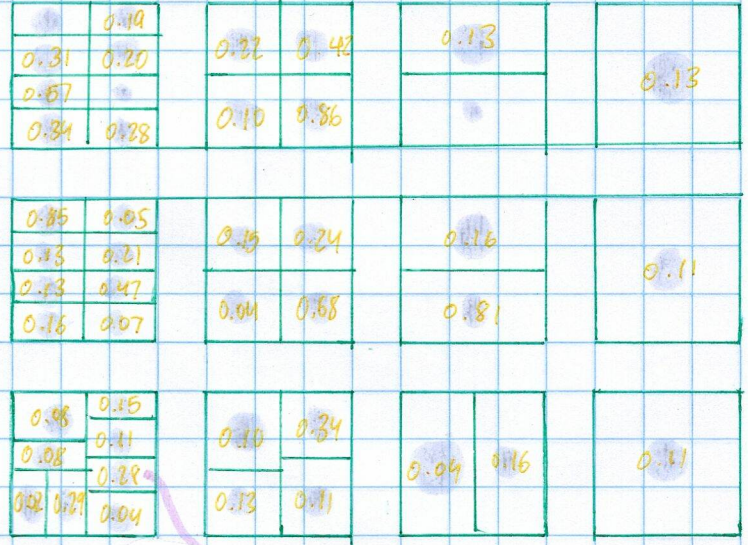
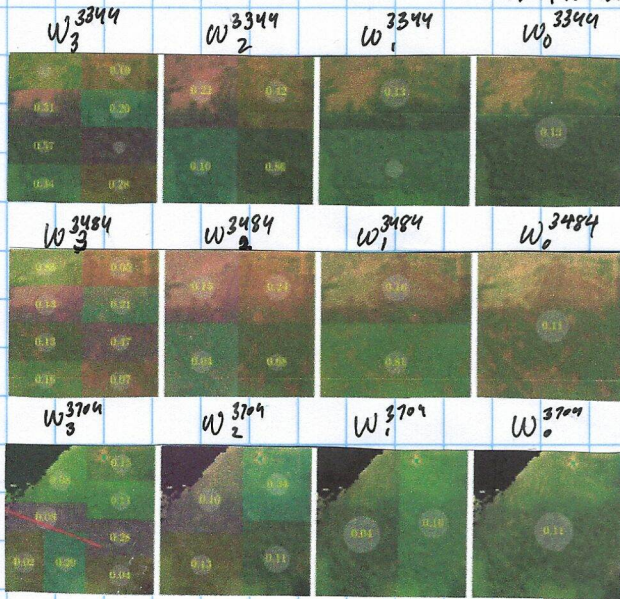


Experimental Results: Continued

VECTOR FIELD RESULTS

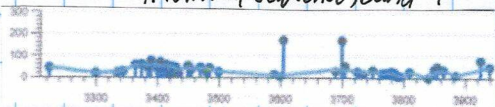
- Model detects notable anomalies on days 3484 & 3704
- Anomaly on day 3704 attributed to near-infrared (band 4)
- Model detects new chlorophyll as healthy leaves reflect near infrared light

see graph below

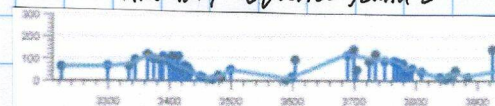


Coordinates (38, 38)

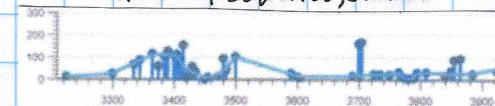
Anomaly Sequence, Band 1



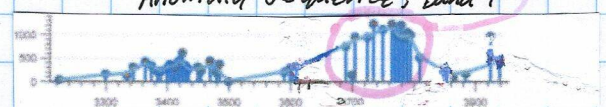
Anomaly Sequence, Band 2



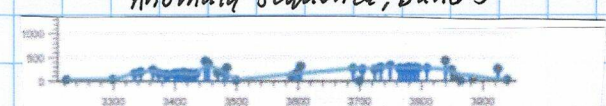
Anomaly Sequence, Band 3



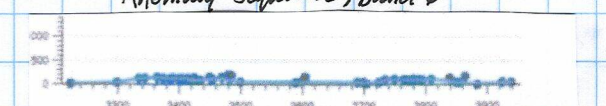
Anomaly Sequence, Band 4



Anomaly Sequence, Band 5



Anomaly Sequence, Band 6



Synthetic Data Test:

- Generated a series of functions based on a stochastic process $\rightarrow v_m(x, \omega) = 1 + y_1(\omega) \left(\frac{\sqrt{\pi} L}{2} \right)^{\frac{1}{2}} + \sum_{k=2}^M \lambda_k^{\frac{1}{2}} + \phi(x) y_k(\omega)$
- Each process is a sum of sine & cosine functions
- $L = L_p = 1/4$
- $M = 100$
- Products of eigenfunctions & values of sines & cosines

Realization of Stochastic Process

