ALTARUM

DEPTH OF KNOWLEDGE • HEIGHTS OF INNOVATION

Robust Methods of Incorporating Prior Knowledge into Generalized RADAR Processing

Nikola Subotic

6 April, 2004

Using Prior Information

ML
vs.
MAP
vs.
Min. Risk
vs.
min/max
vs.
…

Data models:
Phenomenology
Pdfs
Nusiance parameters

Processing Formulation

Norm
L^2, L^p, L^{inf}

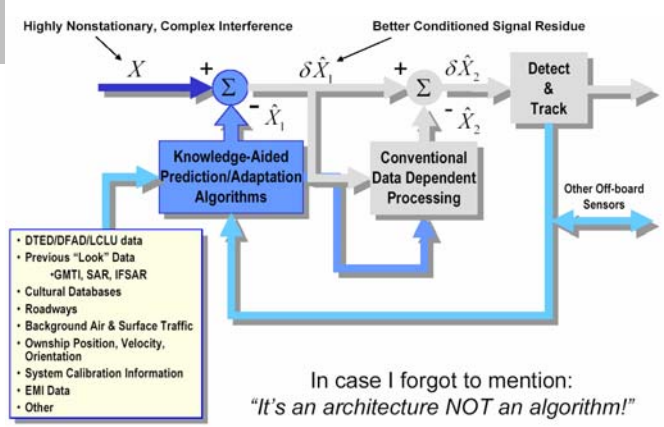
Training Schemes
Jackknife, cross validation,
semi parametric methods,
Tapers, etc.

Sensor
Antenna config.
Coherent integration
Waveform paramters

SAR GMTI

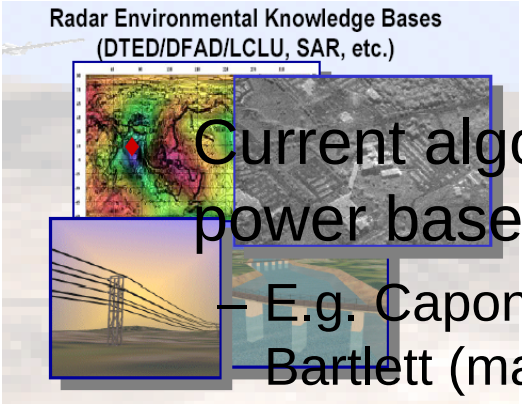
Where?
What?
How?

Prior information could be
-maps
-previous RADAR data
-models
-adjacent RADAR data
-traffic information

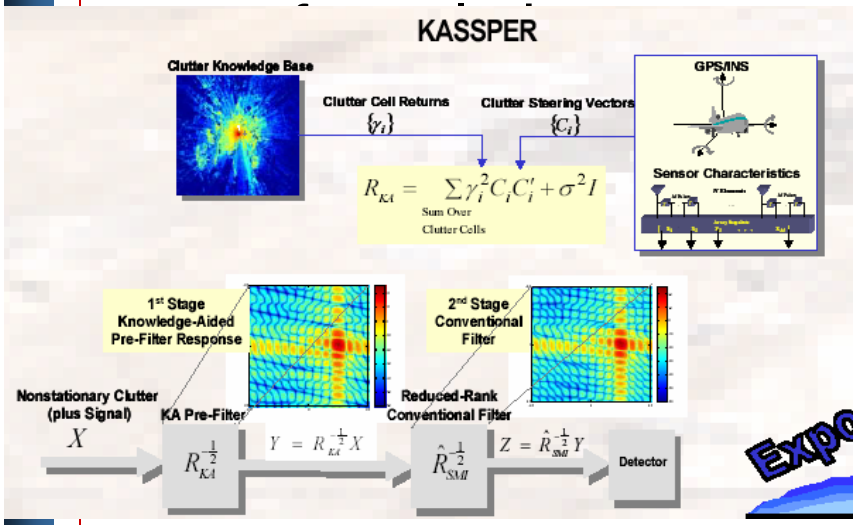


In case I forgot to mention:
"It's an architecture NOT an algorithm!"

Prior Knowledge and Adaptive Processing



- Current algorithms are likelihood based and/or use power based metrics
  - E.g. Capon's method (min. power with unit gain in Θ), Bartlett (max. power in Θ), ML with Gaussian pdf's.
- Use Bayes Risk or posterior probabilities for



$$\Re \equiv E\{C[a,\hat{a}(R)]\} = \int_{-\infty}^{\infty} dA \int_{-\infty}^{\infty} C[a,\hat{a}(R)]\, p_{a,r}(A,R)\, dR$$

$$\Re = \int_{-\infty}^{\infty} dA \int_{-\infty}^{\infty} C[a,\hat{a}(R)]\, p_a(A)\, p_{r|a}(R \mid A)\, dR$$

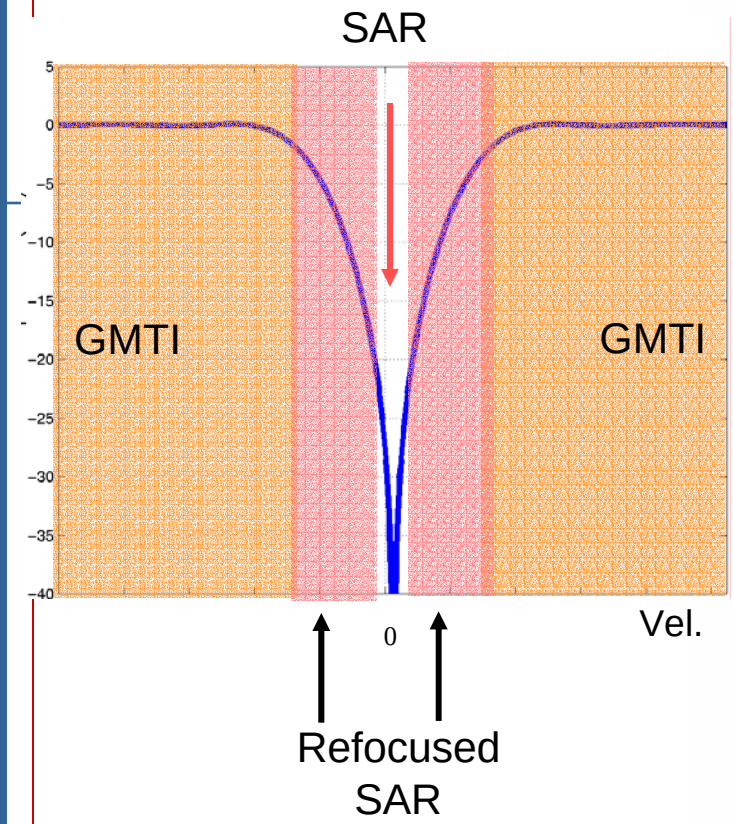
Metric: L^2, L^p

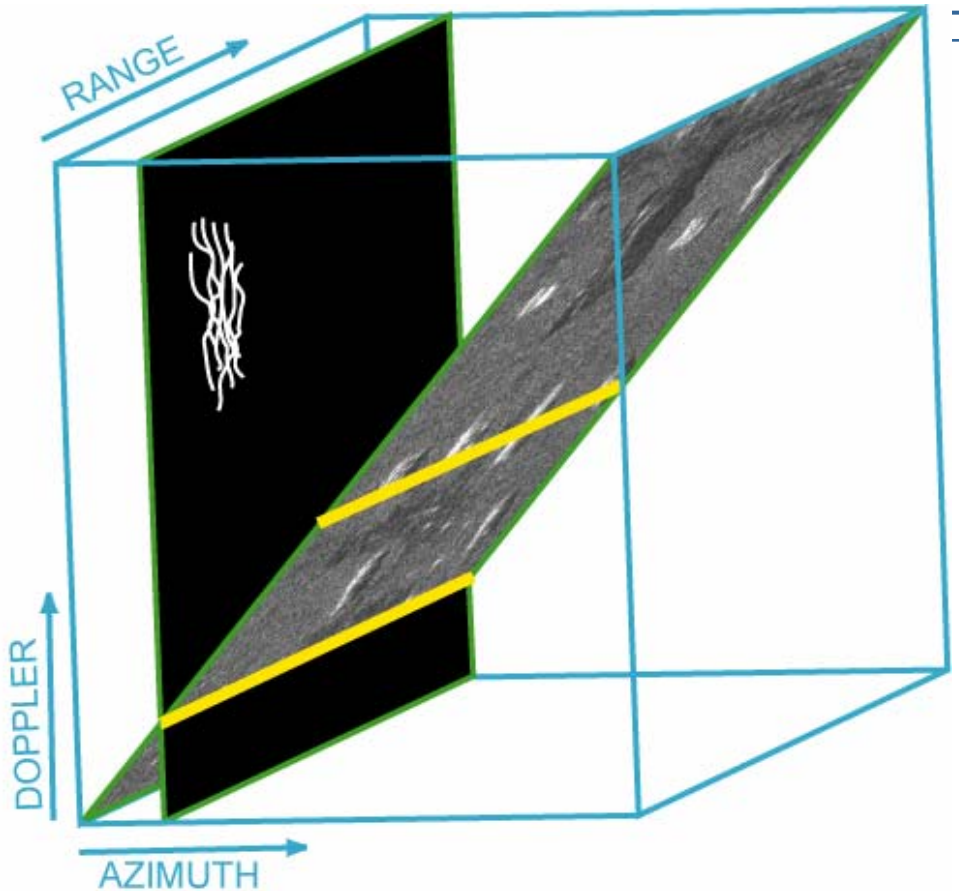
$$C[a,\hat{a}(R)] = C[A-\hat{a}(R)]^p$$

Prior Knowledge

Data

GMTI/SAR Handoff





Long term goal: Single RADAR processing paradigm instead of GMTI vs. SAR

Minimum Risk Classifier

- Based on the risk of declaring hypothesis H_k when H_j is true, j=1,…,J

$$\min_{H_k} R(H_k \mid \bar{x}) = \sum_{j=1}^{J} C_{kj} P(H_j \mid \bar{x}) = \sum_{j=1}^{J} C_{kj} P(H_j) P(\bar{x} \mid H_j)$$

- C_{kj} is the cost of declaring H_k when H_j is true
  - some target misclassification more costly than others

$$C_{kj} = \begin{cases} 0 & k = j \\ C_{kj} & k \neq j \end{cases}$$

- $P(H_j)$ is the prior probability of H_j
  - First order: $P(H_j)=1/J$
- When $C_{kj} = \begin{cases} 0 & k = j \\ 1 & k \neq j \end{cases}$ then minimum risk classifier becomes maximum likelihood (ML)

$$\max_{H_j} P(\bar{x} \mid H_j)$$

ML vs. MAP (Gaussian Case)

- Let $p_{r|a}\left(R \mid A(\theta)\right) = \dfrac{1}{\left(\pi\sigma^2\right)^L} e^{-\|r(t)-As(t)\|^2/\sigma^2}$

$$p_a(A(\theta)) = e^{-\Lambda_A(\theta)}$$

Prior knowledge of presence/location of interferors

$$\hat{\theta}_{ML} = \arg\left\{\min_\theta Tr\left\{\Pi_A^\perp \hat{R}\right\}\right\} = \max_\theta \frac{1}{a^+(\theta)\hat{R}^{-1}a(\theta)}$$

$$\hat{\theta}_{MAP} = \arg\left\{\min_\theta Tr\left\{\Pi_A^\perp \hat{R}\right\} + \Lambda_A(\theta)\right\} = \max_\theta \left\{\frac{1}{a^+(\theta)\hat{R}^{-1}a(\theta)} - \Lambda_A(\theta)\right\}$$

Prior knowledge will impact estimator

Bartlett vs. Minimum MSE

- Bartlett: find $\max_{w} E\left\{\left|w^{+} r\right|\right\}$ subject to $|w| = 1$

$$\hat{\theta}_{Bartlett} = \max_{\theta} \frac{a^{+}(\theta)\hat{R}a(\theta)}{a^{+}(\theta)a(\theta)}$$

- Alternatively minimize

$$\Re = \int_{-\infty}^{\infty} dA \int_{-\infty}^{\infty} (A - \hat{a}(R))^{2}\, p_{a}(A) p_{r|a}(R \mid A) dR$$

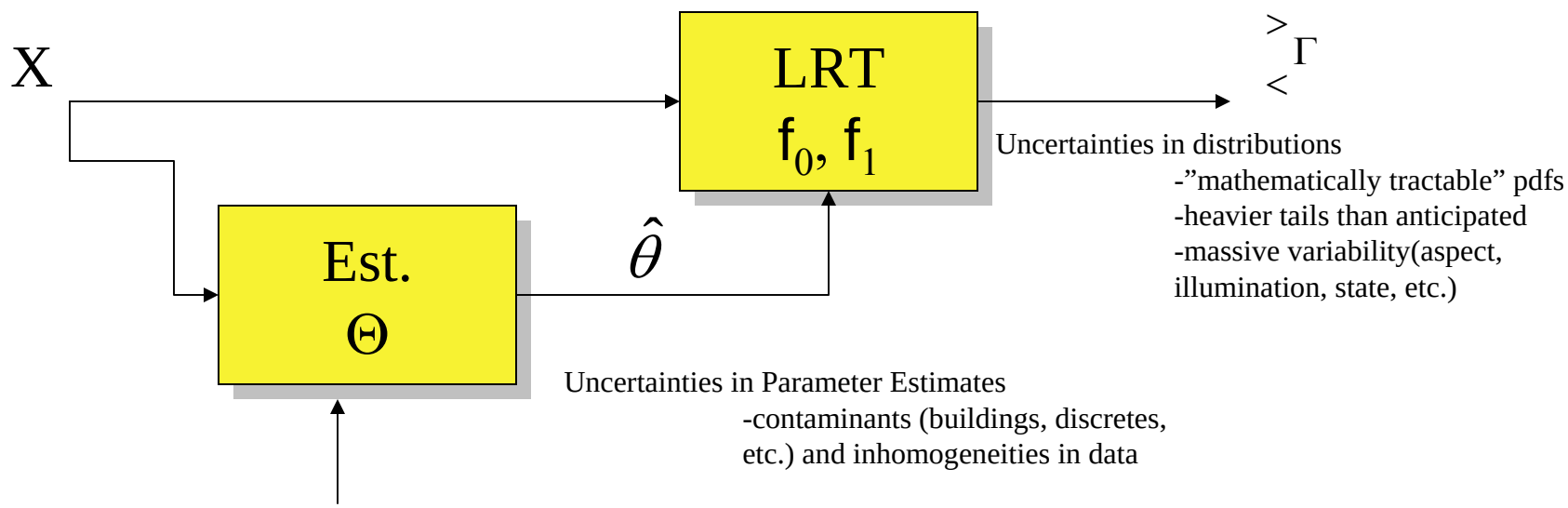
$$\hat{\theta}_{MSE} = \max_{\theta} \left(\frac{a^{+}(\theta)\hat{R}a(\theta)}{a^{+}(\theta)a(\theta)} - \Lambda_{A}(\theta) \right)$$

Sampling Strategies for Training

- General class of bootstrap methods
  - ‘leave one out’ is a special case (the moving hole)
- Bayesian bootstraping methods
  - Sampling where the priors instruct
  - Randomized subsetting and monitoring of performance
    - E.g. fall off rate of SVD coefficients vs. sample set
- Boosting the margins (well known robust approach)
  - Use the priors to weight the training data
  - Modify weights to attain consistency

Model Mismatch Can Adversely Effect Detector Performance

- Likelihood ratio test is fragile for unknown Θ
- Mazimum likelihood estimators are fragile for unknown pdf's
- Unknown pdf's and Θ both detract from GLRT performance

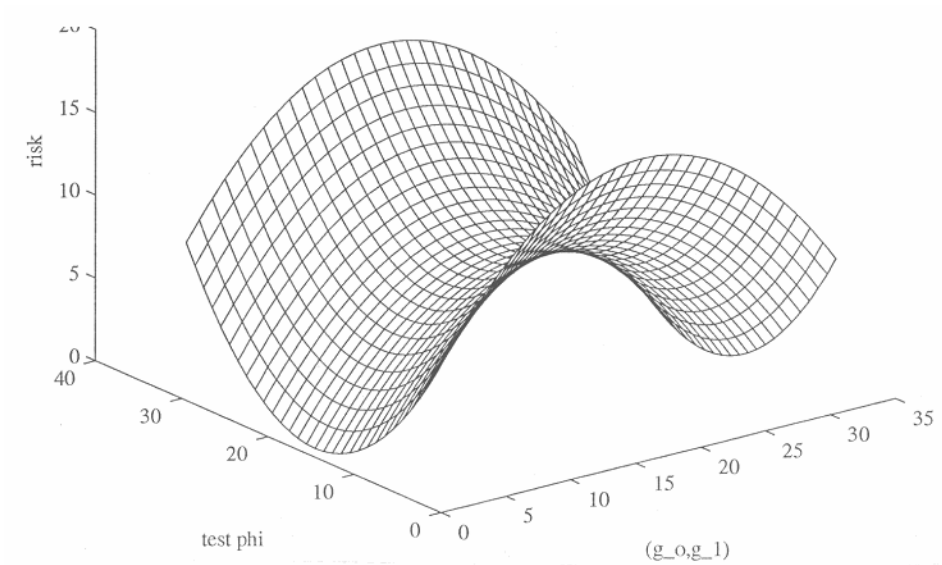


Robust GLRT

- Minimax formulation
  - "minimize worst case error"

$$\min_{\phi}\left(\max_{(g_0,g_1)\in F_0\times F_1} e(g_0,g_1;\phi)\right)$$

  - Suffices for find "least favorable pair" (g_0, g_1) – saddle point

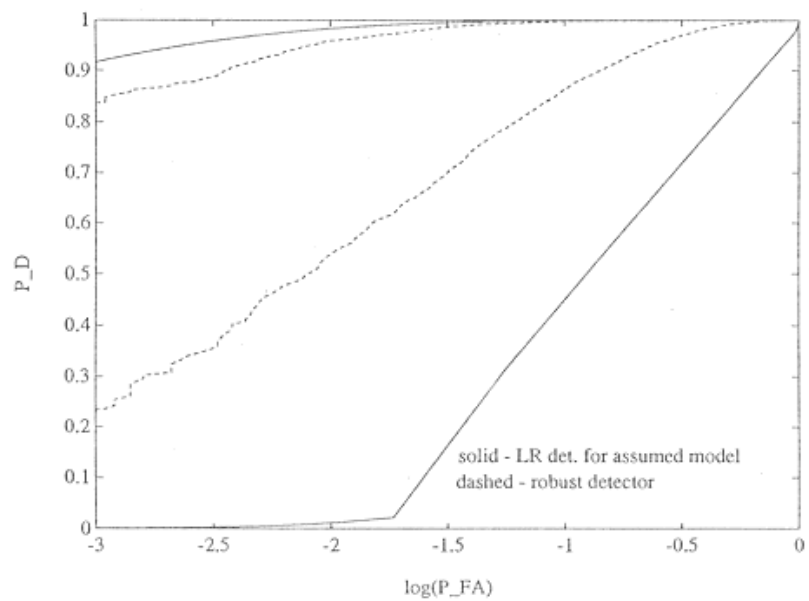


Minimax Robust Detection

- Example: $X_1,\ldots,X_N$ are *iid*

$$H_0 : X_i \sim \aleph(0,1) = f_0 \forall i$$
$$H_1 : X_i \sim \aleph(1,1) = f_1 \forall i$$

- Let $F_0 = \{f : f = (1-\varepsilon_0) f_0 + \varepsilon_0 h\}$, and $F_1 = \{f : f = (1-\varepsilon_1) f_1 + \varepsilon_1 h\}$ with $\varepsilon_0 = \varepsilon_1 = .1$



Minimax Robust Parameter Estimation

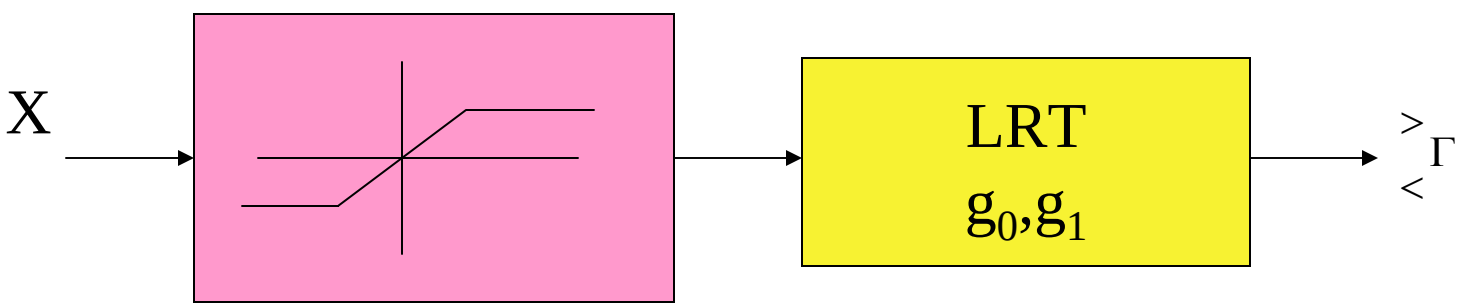
- Classical parameter estimation
  - ML is most efficient
  - ML is very fragile
- Example: $X \sim \aleph(\mu,1)$, estimate μ from $X_1,\ldots,X_N$ (Poor, 19985)
  - ML estimate $\hat{\mu}_{ML} = \frac{1}{N}\sum_{i=1}^{N} X_i$ (sample average)
  - Let $F = \{f : f = (1-\varepsilon)\aleph(\mu,1) + \varepsilon h(\cdot - \mu)\}$
    - For $\varepsilon \neq 0, \sigma^2_{\hat{\mu}_{ML}} \longrightarrow \infty$ (outliers, heavy tailed h)
  - Minimax robust estimator, $\hat{\mu}_R$: soft limit, sample average

$$\hat{\mu}_R \text{ is } 92\% \text{ efficient with } \sigma^2_{\hat{\mu}_R} = 1.5$$

- Same situation for covariance estimates

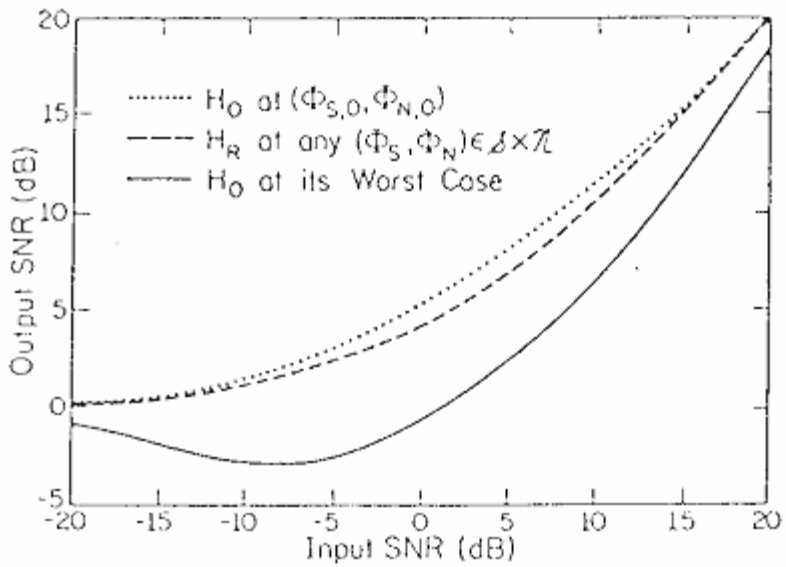
Characteristics of Robust Systems

- Test statistic and estimators often have the form $\sum_{i=1}^{N} l(x_i)$ where *l* is unbounded
- Robust estimators guard against unwaranted outliers and/or heavier pdf tails
  - Can be viewed as a “soft limiting” system
- Gives up some perofrmance (e.g. optimality) to be less fragile with respect to unknown information



Minimax Robust Linear Filtering

- Linear filter designed to maximize SNR for known assumed power spectral densities of signal and noise
- Assume only knowledge of total signal and noise powers and fractional signal and noise powers in a specific frequency band (Vastola and Poor, 1983)

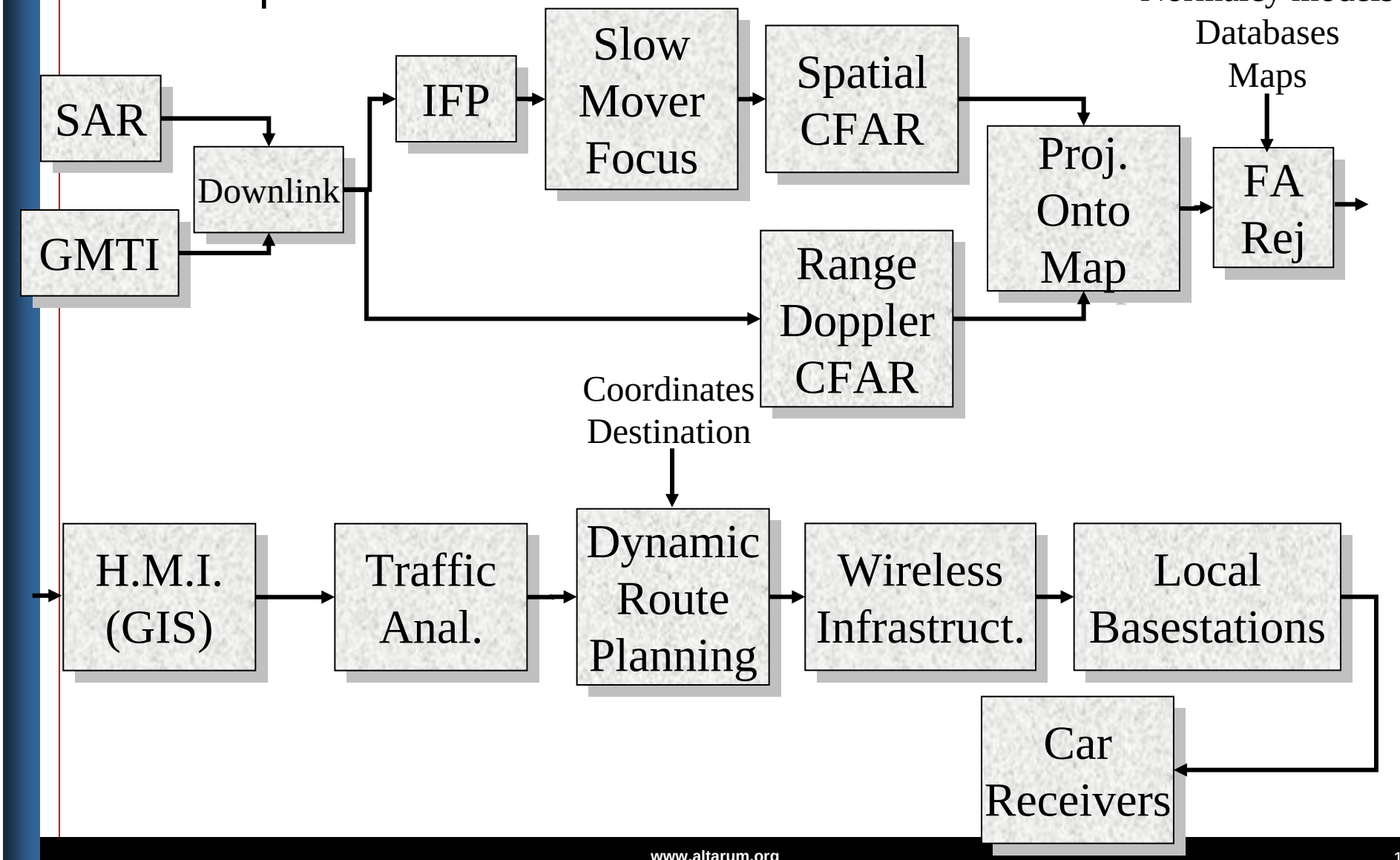


Robust Minimax Detection-Summary

- Incorporates uncertainty in both statistical models and unavoidable data anomalies
- Useful in both estimation and final detection in a GLRT
  - Estimation: power spectral density models, cultural discretes, inhomogeneities
  - Detection: heavier tailed distributions than anticipated, target uncertainties
- Gives up some performance in the 'perfect' scenario to guard against real-world uncertainties
  - Optimum detectors are non-robust to real-world uncertainties
- Average performance of the robust detector over all the uncertainties may be significantly better than the optimum detector
- Robust detectors allow the incorporation of some target knowledge into the detector

Processing Architecture

Assume separate modes



Proposed Tasks

- **Risk based formulation of spectral analysis techniques:** Re-derive STAP algorithms (e.g. post-Doppler STAP) using a minimum Risk formulation. Identify how the prior information (maps, signatures, etc.) enter into the formulation. Analyze the performance of the minimum risk STAP algorithm using the ISL derived data cube. Metrics of performance include SINR loss as a function of speed, cross section, etc.
- **Small sample size robust strategies:** Incorporate sampling strategies such as jackknifing, boosting, and cross validation for both classical STAP algorithms and minimum risk formulations. These strategies will be compared to classic sampling strategies from the perspective of statistical efficiency, bias and convergence to the Cramer-Rao bound.
- **Robust formulations:** Explore the formulation of STAP algorithms based on fundamentally robust formulations such as min/max, maximal invariance and exact robustness theory. These algorithms will be studied for the scenario of applicability and their performance over a class of known and unknown interferers.
- **Support for the architecture working group:** We will support the working group which is defining the overall KASSPER architecture to determine where and how any prior information is being incorporated..