

How Well Do You Know That?

Uncertainty Analysis in Earth Remote Sensing

Robin D. Morris

Universities Space Research Association

A. Kottas, M. Farah

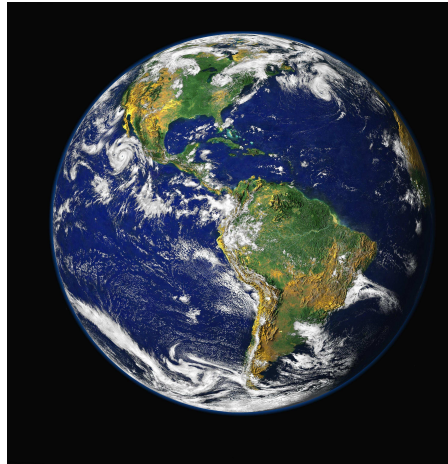
University of California, Santa Cruz

R. Furfaro

University of Arizona

Supported by NASA Grant Numbers NNG06GI77G and NNX07AV69G

Modeling The Biosphere Has Important Scientific and Public Policy Implications

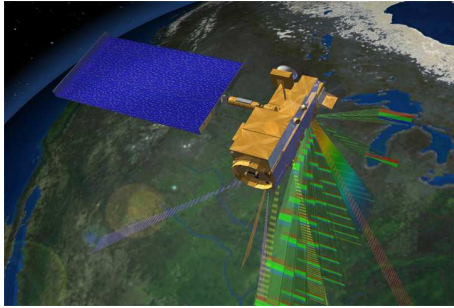


Modeling The Biosphere Has Important Scientific and Public Policy Implications



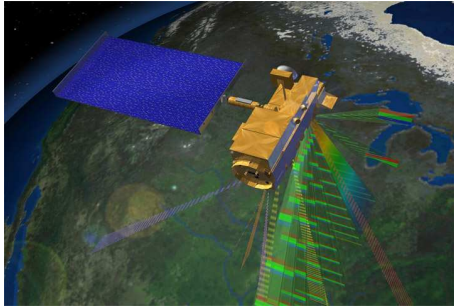
Forests are a major source for the transfer of mass and energy from land to the atmosphere

Modeling The Biosphere Has Important Scientific and Public Policy Implications



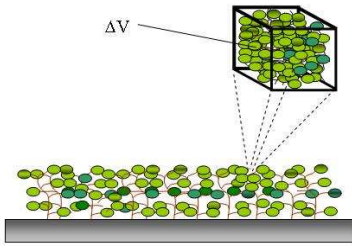
Also, you can see them from space!

Modeling The Biosphere Has Important Scientific and Public Policy Implications



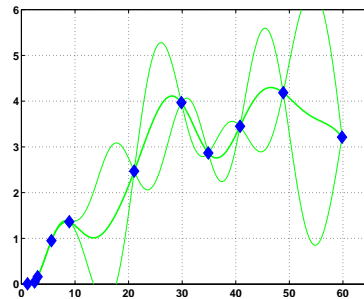
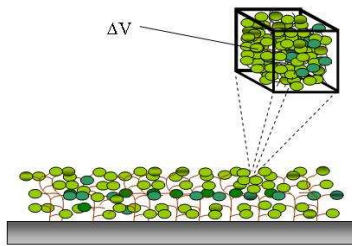
How well can we estimate forest parameters from remote-sensed data?

Statistical Inference Provides a Methodology to Analyse Models and Provide Well-Calibrated Estimates



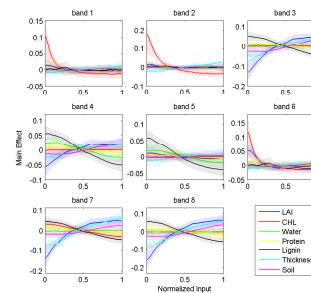
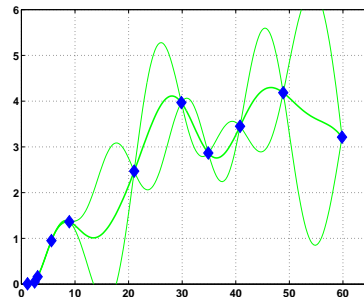
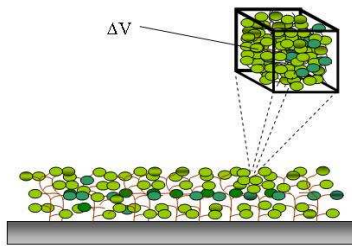
Analyse a computer model of light interaction with forest canopies.

Statistical Inference Provides a Methodology to Analyse Models and Provide Well-Calibrated Estimates



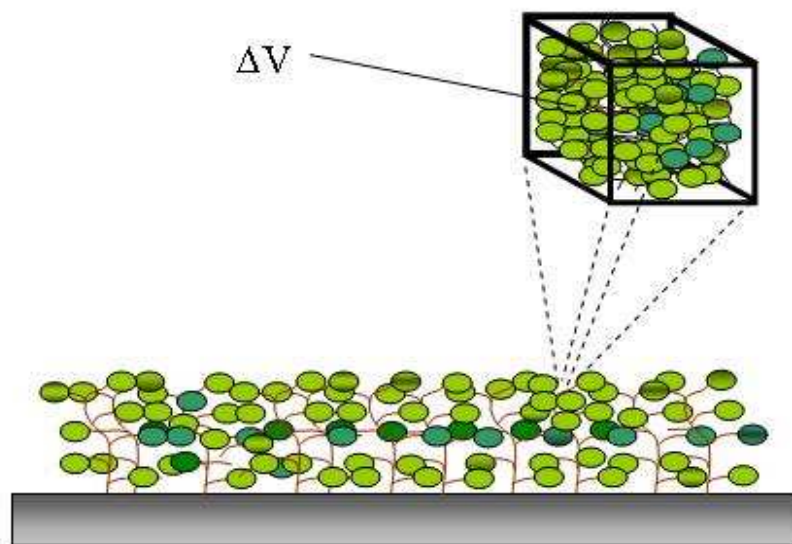
Emulate the computer model using a Gaussian Process.

Statistical Inference Provides a Methodology to Analyse Models and Provide Well-Calibrated Estimates



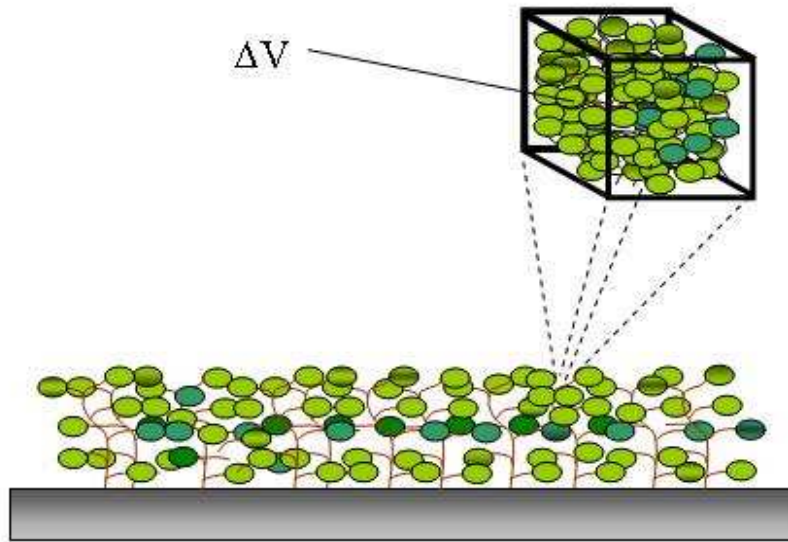
The Gaussian Process model is amenable to analysis.

Radiative Transfer Models Simulate the Interaction of Light With Forest Canopies



CANopy MODeI

Radiative Transfer Models Simulate the Interaction of Light With Forest Canopies



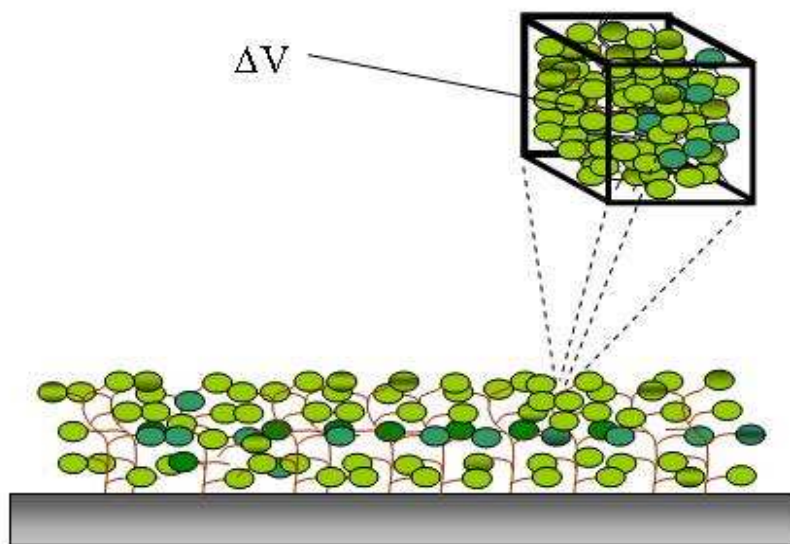
CANopy MODeI

Leaf Area Index (LAI)

Leaf Angle Distribution

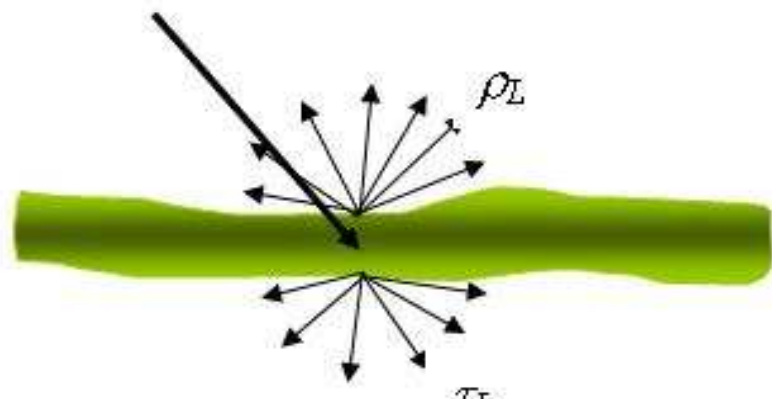
Soil Reflectance

Radiative Transfer Models Simulate the Interaction of Light With Forest Canopies



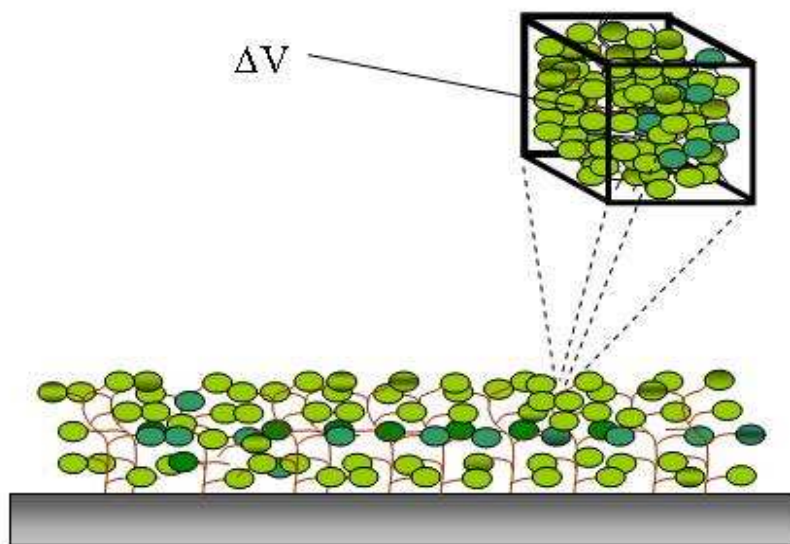
CANopy MODeI

Leaf Area Index (LAI)
Leaf Angle Distribution
Soil Reflectance



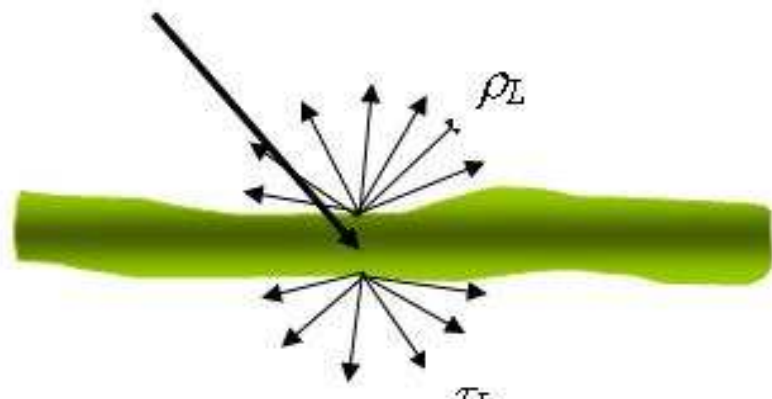
LEAF MODeI

Radiative Transfer Models Simulate the Interaction of Light With Forest Canopies



CANopy MODeI

Leaf Area Index (LAI)
Leaf Angle Distribution
Soil Reflectance



LEAF MODeI

Chlorophyll
Water Fraction
Protein
Lignin/Cellulose
Thickness

Global Sensitivity Analysis Reveals Important Characteristics of the Computer Model

Decompose the output of the LCM as

$$y = f(\mathbf{v}) = \mathbf{E}(Y) + \sum_{i=1}^d z_i(v_i) + \sum_{i < j} z_{i,j}(v_i, v_j) + \dots \\ + z_{1,2,\dots,d}(v_1, v_2, \dots, v_d)$$

Global Sensitivity Analysis Reveals Important Characteristics of the Computer Model

Decompose the output of the LCM as

$$y = f(\mathbf{v}) = \mathbf{E}(Y) + \sum_{i=1}^d z_i(v_i) + \sum_{i < j} z_{i,j}(v_i, v_j) + \dots \\ + z_{1,2,\dots,d}(v_1, v_2, \dots, v_d)$$

Global Mean

$$\mathbf{E}(Y) = \int_{v_j, j=1\dots d} f(\mathbf{v}) dH(\mathbf{v})$$

Global Sensitivity Analysis Reveals Important Characteristics of the Computer Model

Decompose the output of the LCM as

$$y = f(\mathbf{v}) = \mathbf{E}(Y) + \sum_{i=1}^d z_i(v_i) + \sum_{i < j} z_{i,j}(v_i, v_j) + \dots \\ + z_{1,2,\dots,d}(v_1, v_2, \dots, v_d)$$

Global Mean

$$\mathbf{E}(Y) = \int_{v_j, j=1\dots d} f(\mathbf{v}) dH(\mathbf{v})$$

Main Effects

$$z_i(v_i) = \mathbf{E}(Y|v_i) - \mathbf{E}(Y) \\ = \int_{\mathbf{v}_{-i}} f(\mathbf{v}) dH(\mathbf{v}_{-i}|v_i) - \mathbf{E}(Y)$$

Sensitivity Indices - The Expected Reduction in Output Uncertainty if an Input Was Known Exactly

Variances of the components of the output decomposition

$$V_i = \text{Var}\{\mathbf{E}(Y \mid v_i)\} = \mathbf{E} [(\mathbf{E}(Y \mid v_i))^2] - (\mathbf{E}(Y))^2.$$

Sensitivity Indices - The Expected Reduction in Output Uncertainty if an Input Was Known Exactly

Variances of the components of the output decomposition

$$V_i = \text{Var}\{\mathbf{E}(Y \mid v_i)\} = \mathbf{E} [(\mathbf{E}(Y \mid v_i))^2] - (\mathbf{E}(Y))^2.$$

Normalize to give the Sensitivity Indices

$$S_i = \frac{V_i}{\text{Var}(Y)}$$

How much of the variance of the output is due to input i .

If I learn the value of input i exactly, by how much is the variance of the output reduced?

There are Two Strategies for Computing the Integrals Involved in the Main Effects and Sensitivity Indices

Monte Carlo approximation of the analytic integral.

There are Two Strategies for Computing the Integrals Involved in the Main Effects and Sensitivity Indices

Monte Carlo approximation of the analytic integral.

Analytic integration of an approximation to the function.

There are Two Strategies for Computing the Integrals Involved in the Main Effects and Sensitivity Indices

Monte Carlo approximation of the analytic integral.

Analytic integration of an approximation to the function.

Trade-off depends on the dimensionality.

Use a Gaussian Process Emulator in Place of the LCM

A GP is a distribution over functions.

Use a Gaussian Process Emulator in Place of the LCM

A GP is a distribution over functions.

Specified by the

mean function

$$\mathbf{E}(f(\mathbf{v}))$$

$$= \mu$$

covariance function

$$\mathbf{Cov}(f(\mathbf{v}), f(\mathbf{v}'))$$

$$= \sigma^{-2} \exp\left(-\sum_{l=1}^k \phi_l |v_l - v'_l|^\alpha\right)$$

Use a Gaussian Process Emulator in Place of the LCM

A GP is a distribution over functions.

Specified by the

mean function

$$\mathbf{E}(f(\mathbf{v}))$$

$$= \mu$$

covariance function

$$\mathbf{Cov}(f(\mathbf{v}), f(\mathbf{v}'))$$

$$= \sigma^{-2} \exp\left(-\sum_{l=1}^k \phi_l |v_l - v'_l|^\alpha\right)$$

Joint distribution of any finite set of points is multivariate Gaussian.

Parameters (μ, σ, ϕ) learned using maximum likelihood from a set of training runs of the LCM.

Taking Into Account the Uncertainty Introduced by the GP Approximation

Recall Main Effects $z_i(v_i) = \mathbf{E}(Y|v_i) - \mathbf{E}(Y)$

Taking Into Account the Uncertainty Introduced by the GP Approximation

Recall Main Effects $z_i(v_i) = \mathbf{E}(Y|v_i) - \mathbf{E}(Y)$

We must include the uncertainty introduced by using the GP emulator for $y = f(\mathbf{v})$

Taking Into Account the Uncertainty Introduced by the GP Approximation

Recall Main Effects $z_i(v_i) = \mathbf{E}(Y|v_i) - \mathbf{E}(Y)$

We must include the uncertainty introduced by using the GP emulator for $y = f(\mathbf{v})$

$$\mathbf{E}^* \{ \mathbf{E}(Y) \} = \hat{\mu} + \mathbf{T}^T \mathbf{C}^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}_n)$$

$\mathbf{E}^* \{ \}$ is expectation wrt GP

\mathbf{T} – $n \times 1$ vector with elements

$$\prod_{\ell=1}^d \left\{ \int_{a_\ell}^{b_\ell} \exp(-\phi_\ell |v_\ell - x_{i\ell}|^\alpha) (b_\ell - a_\ell)^{-1} dv_\ell \right\}$$

Taking Into Account the Uncertainty Introduced by the GP Approximation

Recall Main Effects $z_i(v_i) = \mathbf{E}(Y|v_i) - \mathbf{E}(Y)$

We must include the uncertainty introduced by using the GP emulator for $y = f(\mathbf{v})$

$$\mathbf{E}^* \{ \mathbf{E}(Y) \} = \hat{\mu} + \mathbf{T}^T \mathbf{C}^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}_n)$$

$\mathbf{E}^* \{ \}$ is expectation wrt GP

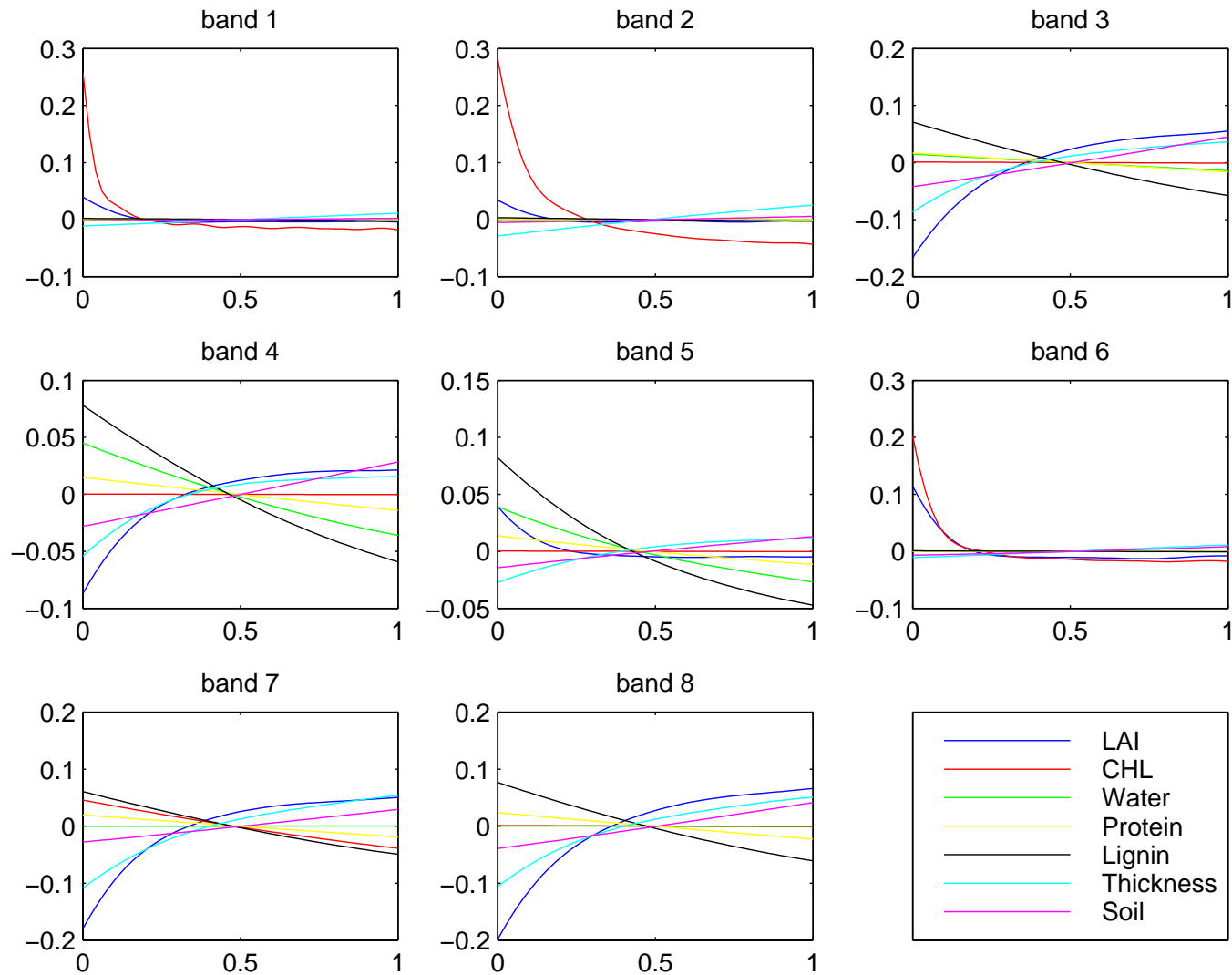
\mathbf{T} – $n \times 1$ vector with elements

$$\prod_{\ell=1}^d \left\{ \int_{a_\ell}^{b_\ell} \exp(-\phi_\ell |v_\ell - x_{i\ell}|^\alpha) (b_\ell - a_\ell)^{-1} dv_\ell \right\}$$

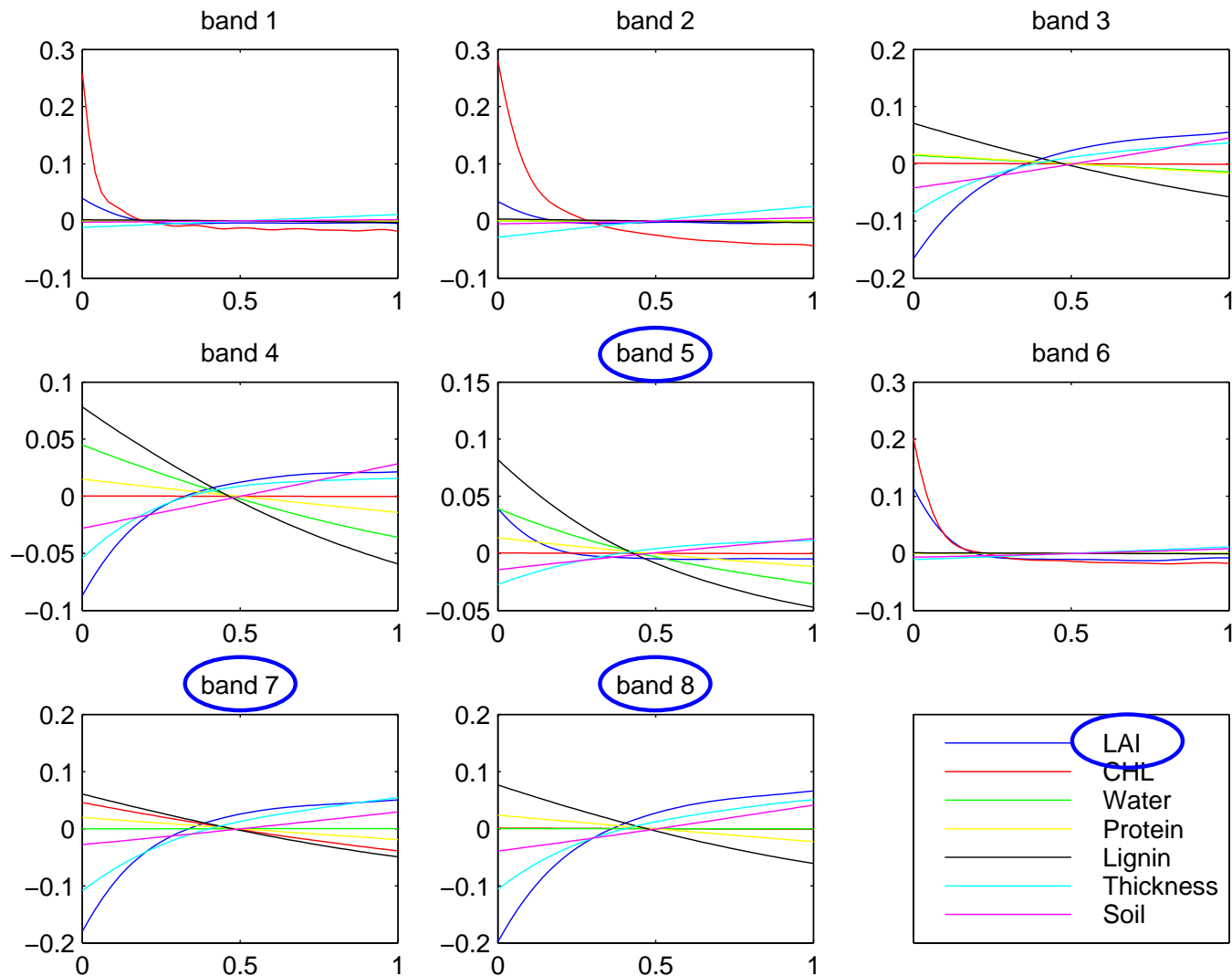
$\mathbf{E}^* \{ \mathbf{E}(Y | u_j) \}$ can be similarly derived

This gives point estimates for the main effects.

Main Effects for the Leaf-Canopy Model

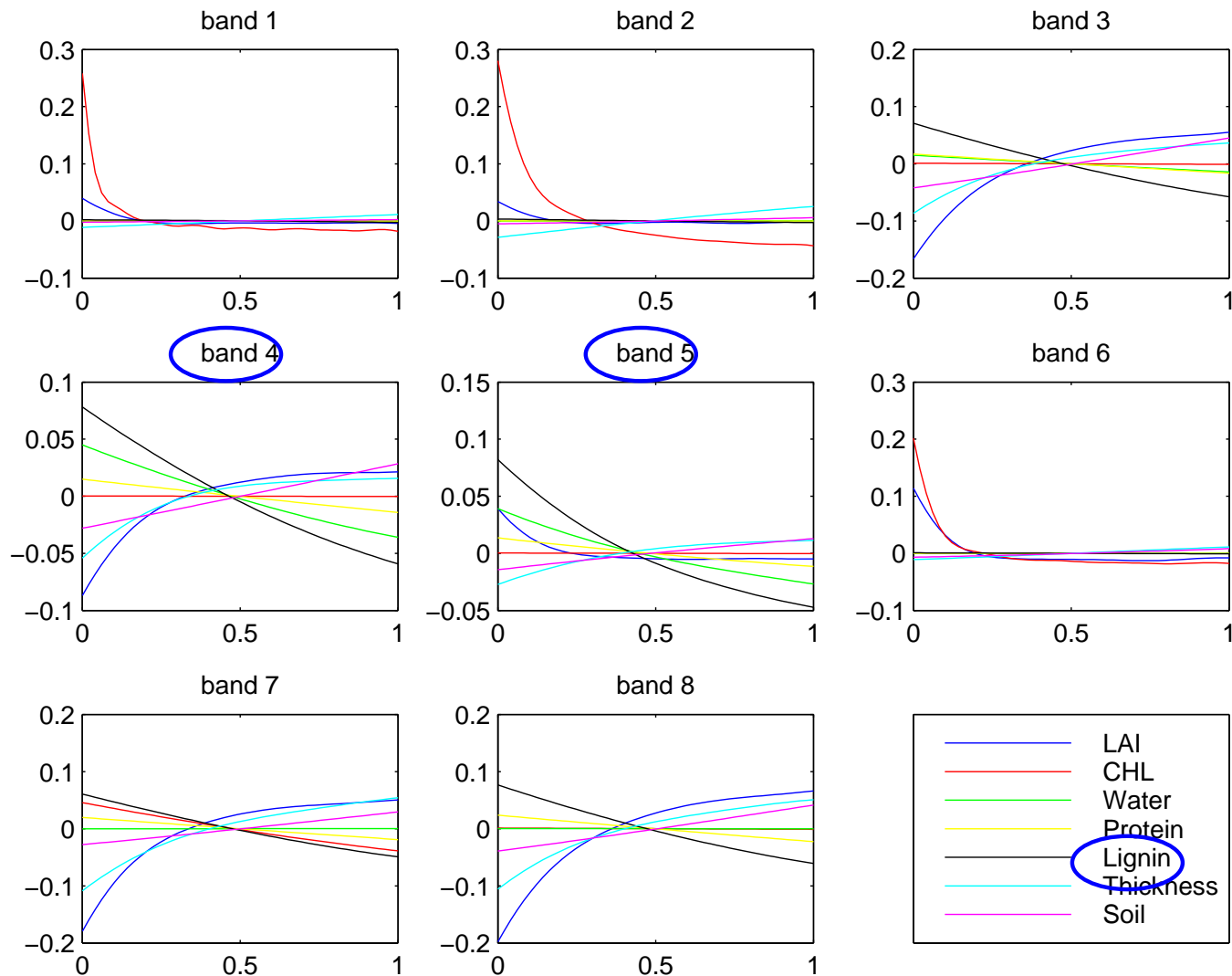


Main Effects for the Leaf-Canopy Model



LAI: affects NIR (bands 7,8); opposite effect in visible (band 5)

Main Effects for the Leaf-Canopy Model



Lignin: affects SWIR (bands 4,5); surprising to domain scientists

Estimating the Uncertainty of the Main Effects

Take $\text{Var}^* \{ \}$ – variances wrt GP predictive distribution as a measure of this uncertainty

$$\text{Var}^* \{ \mathbf{E}(Y \mid u_j) \} = \mathbf{E}^* \{ (\mathbf{E}(Y \mid u_j))^2 \} - (\mathbf{E}^* \{ \mathbf{E}(Y \mid u_j) \})^2.$$

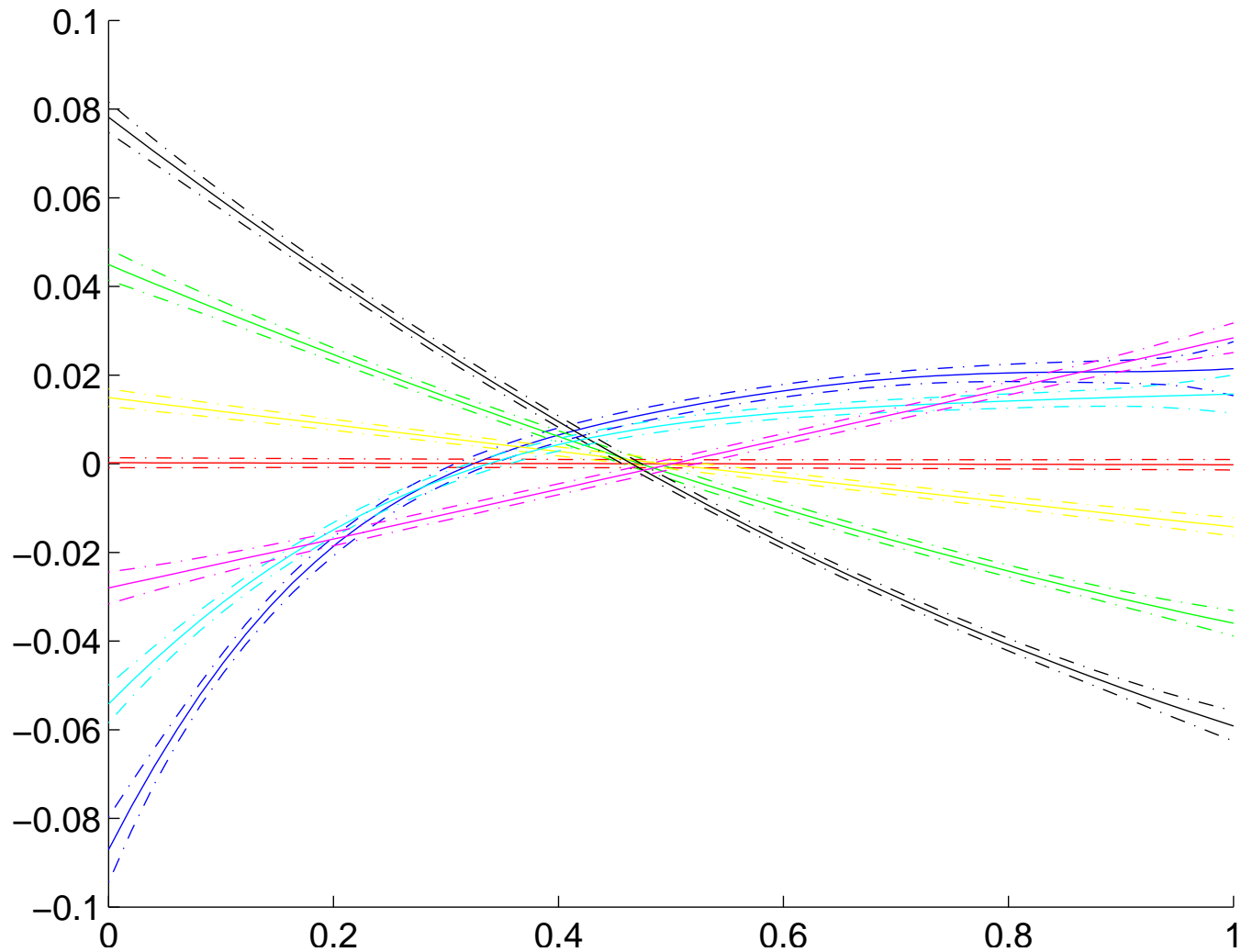
Estimating the Uncertainty of the Main Effects

Take $\text{Var}^* \{ \}$ – variances wrt GP predictive distribution as a measure of this uncertainty

$$\text{Var}^* \{ \mathbf{E}(Y | u_j) \} = \mathbf{E}^* \{ (\mathbf{E}(Y | u_j))^2 \} - (\mathbf{E}^* \{ \mathbf{E}(Y | u_j) \})^2.$$

For our modeling choices (constant mean; exponential or squared exponential correlation function; uniform priors on the inputs) this can be computed analytically.

Main Effects With Uncertainties



The uncertainties due to the GP approximation are small.

Sensitivity Indices Under the GP Approximation

$$S_j = \frac{\text{Var}(\mathbf{E}(Y \mid u_j))}{\text{Var}(Y)}$$

Sensitivity Indices Under the GP Approximation

$$S_j = \frac{\text{Var}(\mathbf{E}(Y \mid u_j))}{\text{Var}(Y)}$$

Cannot compute $\mathbf{E}^* \{S_j\}$ analytically.

Sensitivity Indices Under the GP Approximation

$$S_j = \frac{\text{Var}(\mathbf{E}(Y \mid u_j))}{\text{Var}(Y)}$$

Cannot compute $\mathbf{E}^* \{S_j\}$ analytically.

Approximate by

$$\frac{\mathbf{E}^* \{\text{Var}(\mathbf{E}(Y \mid u_j))\}}{\mathbf{E}^* \{\text{Var}(Y)\}}$$

Sensitivity Indices for the Leaf-Canopy Model

	band; wavelength (<i>nm</i>)							
	1	2	3	4	5	6	7	8
input	469	555	1240	1640	2130	667	748	870
LAI	0.05	0.01	0.43	0.16	0.04	0.28	0.41	0.48
CHL	0.80	0.83	0.00	0.00	0.00	0.56	0.08	0.00
Water	0.00	0.00	0.01	0.12	0.14	0.00	0.00	0.00
Protein	0.00	0.00	0.01	0.02	0.02	0.00	0.02	0.02
Lignin	0.00	0.00	0.19	0.36	0.53	0.00	0.13	0.16
Thick.	0.02	0.05	0.14	0.07	0.05	0.02	0.24	0.18
Soil	0.00	0.00	0.08	0.06	0.03	0.01	0.03	0.06
Total	0.88	0.90	0.86	0.80	0.81	0.87	0.90	0.90

Sensitivity Indices for the Leaf-Canopy Model

	band; wavelength (<i>nm</i>)							
	1	2	3	4	5	6	7	8
input	469	555	1240	1640	2130	667	748	870
LAI	0.05	0.01	0.43	0.16	0.04	0.28	0.41	0.48
CHL	0.80	0.83	0.00	0.00	0.00	0.56	0.08	0.00
Water	0.00	0.00	0.01	0.12	0.14	0.00	0.00	0.00
Protein	0.00	0.00	0.01	0.02	0.02	0.00	0.02	0.02
Lignin	0.00	0.00	0.19	0.36	0.53	0.00	0.13	0.16
Thick.	0.02	0.05	0.14	0.07	0.05	0.02	0.24	0.18
Soil	0.00	0.00	0.08	0.06	0.03	0.01	0.03	0.06
Total	0.88	0.90	0.86	0.80	0.81	0.87	0.90	0.90

Sensitivity Indices for the Leaf-Canopy Model

	band; wavelength (<i>nm</i>)							
	1	2	3	4	5	6	7	8
input	469	555	1240	1640	2130	667	748	870
LAI	0.05	0.01	0.43	0.16	0.04	0.28	0.41	0.48
CHL	0.80	0.83	0.00	0.00	0.00	0.56	0.08	0.00
Water	0.00	0.00	0.01	0.12	0.14	0.00	0.00	0.00
Protein	0.00	0.00	0.01	0.02	0.02	0.00	0.02	0.02
Lignin	0.00	0.00	0.19	0.36	0.53	0.00	0.13	0.16
Thick.	0.02	0.05	0.14	0.07	0.05	0.02	0.24	0.18
Soil	0.00	0.00	0.08	0.06	0.03	0.01	0.03	0.06
Total	0.88	0.90	0.86	0.80	0.81	0.87	0.90	0.90

Sensitivity Indices for the Leaf-Canopy Model

	band; wavelength (<i>nm</i>)							
	1	2	3	4	5	6	7	8
input	469	555	1240	1640	2130	667	748	870
LAI	0.05	0.01	0.43	0.16	0.04	0.28	0.41	0.48
CHL	0.80	0.83	0.00	0.00	0.00	0.56	0.08	0.00
Water	0.00	0.00	0.01	0.12	0.14	0.00	0.00	0.00
Protein	0.00	0.00	0.01	0.02	0.02	0.00	0.02	0.02
Lignin	0.00	0.00	0.19	0.36	0.53	0.00	0.13	0.16
Thick.	0.02	0.05	0.14	0.07	0.05	0.02	0.24	0.18
Soil	0.00	0.00	0.08	0.06	0.03	0.01	0.03	0.06
Total	0.88	0.90	0.86	0.80	0.81	0.87	0.90	0.90

Wait a Minute. Aren't You Missing Something Here?

What about the uncertainty due to the estimated GP parameters?

Wait a Minute. Aren't You Missing Something Here?

What about the uncertainty due to the estimated GP parameters?

The GP parameters are estimated from a 250 point Latin Hypercube sampling of the input space.

There is significant uncertainty in the estimated GP parameters.

Bayesian Estimation using MCMC to Include GP Parameter Uncertainty

Generate samples of the GP parameters

$$\psi = (\theta, \mu, \sigma^2, \phi)$$

where θ are the predicted outputs at the training points.

Bayesian Estimation using MCMC to Include GP Parameter Uncertainty

Generate samples of the GP parameters

$$\boldsymbol{\psi} = (\boldsymbol{\theta}, \mu, \sigma^2, \phi)$$

where $\boldsymbol{\theta}$ are the predicted outputs at the training points.

Posterior predictive distribution for $\tilde{y} = f(\mathbf{v})$ is

$$p(\tilde{y} | D) = \iint \mathbf{N}(\tilde{y} | \tilde{\boldsymbol{\theta}}, J) \mathbf{N}(\tilde{\boldsymbol{\theta}} | m(\mathbf{v}), S(\mathbf{v})) p(\boldsymbol{\psi} | D) d\boldsymbol{\psi} d\tilde{\boldsymbol{\theta}}$$

Bayesian Estimation using MCMC to Include GP Parameter Uncertainty

Generate samples of the GP parameters

$$\psi = (\boldsymbol{\theta}, \mu, \sigma^2, \phi)$$

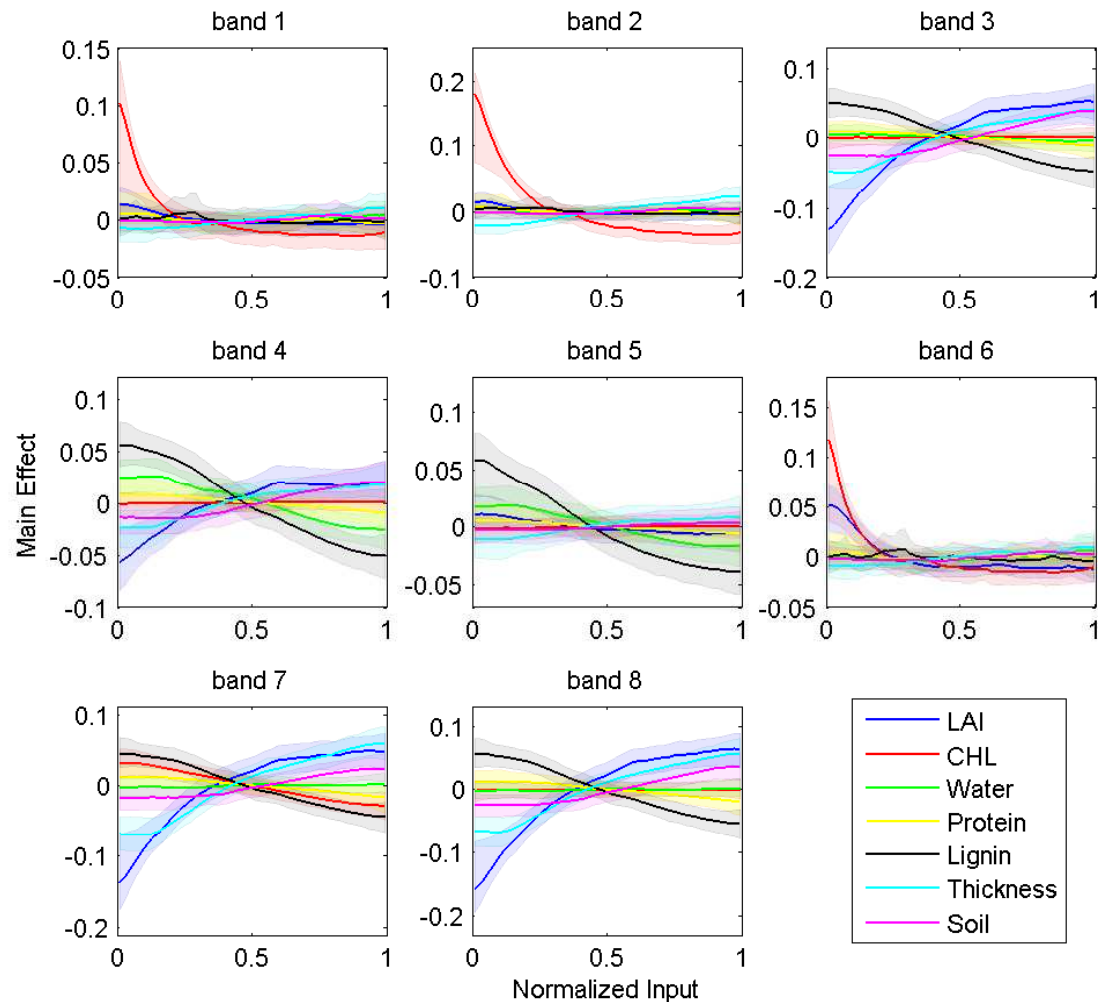
where $\boldsymbol{\theta}$ are the predicted outputs at the training points.

Posterior predictive distribution for $\tilde{y} = f(\mathbf{v})$ is

$$p(\tilde{y} | D) = \iint \mathbf{N}(\tilde{y} | \tilde{\boldsymbol{\theta}}, J) \mathbf{N}(\tilde{\boldsymbol{\theta}} | m(\mathbf{v}), S(\mathbf{v})) p(\boldsymbol{\psi} | D) d\boldsymbol{\psi} d\tilde{\boldsymbol{\theta}}$$

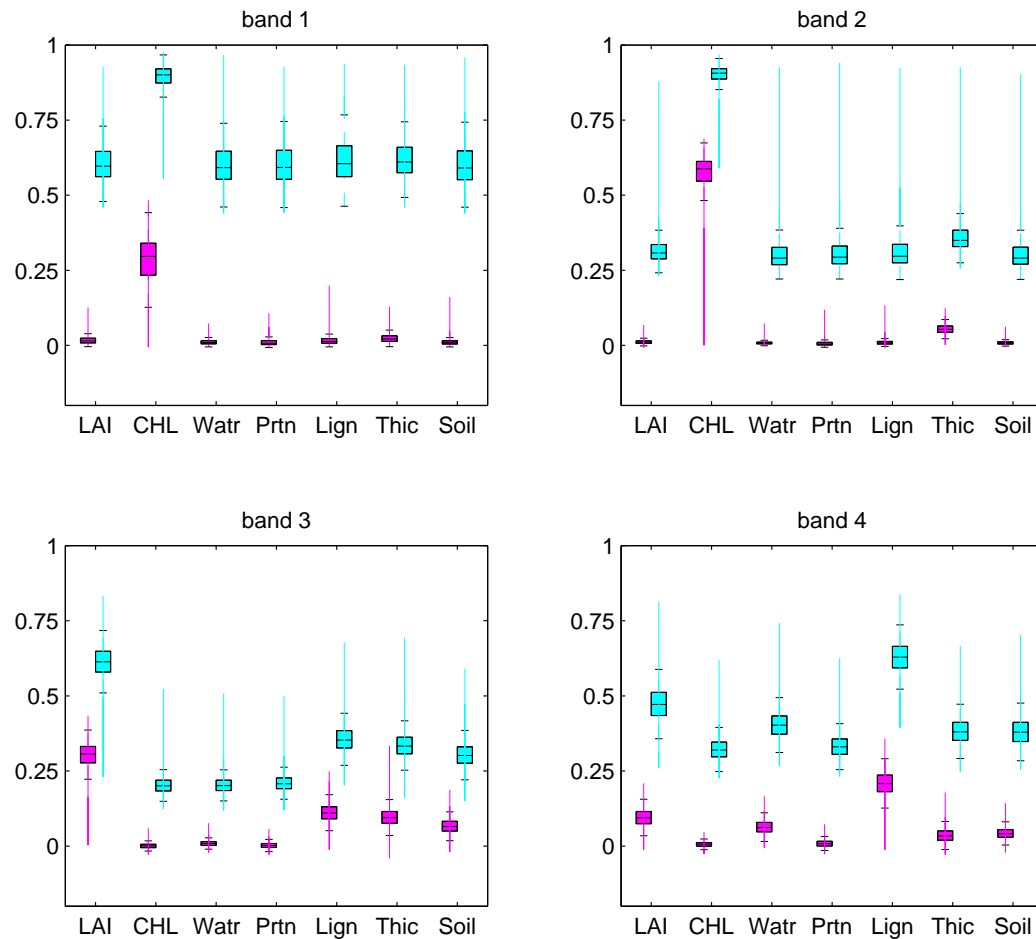
Can use these to estimate the full distribution of the main effects and the sensitivity indices.

Median and 95% Probability Bands of the Posterior Distributions of the Main Effects



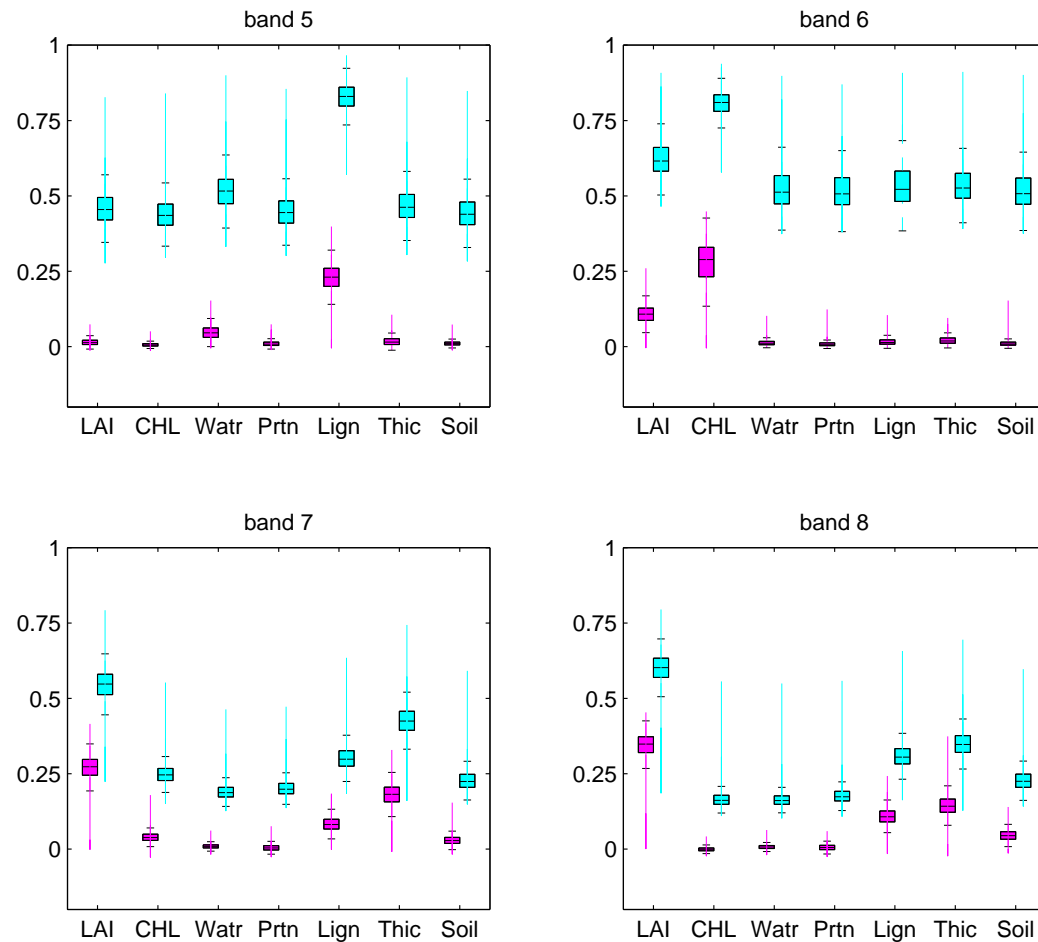
Uncertainties are larger, especially at the extreme values of the inputs.
Basic behaviour is the same.

Distributions of the Sensitivity Indices



Large values of Total Sensitivity Indices when first order SI is close to zero indicate important interaction effects.

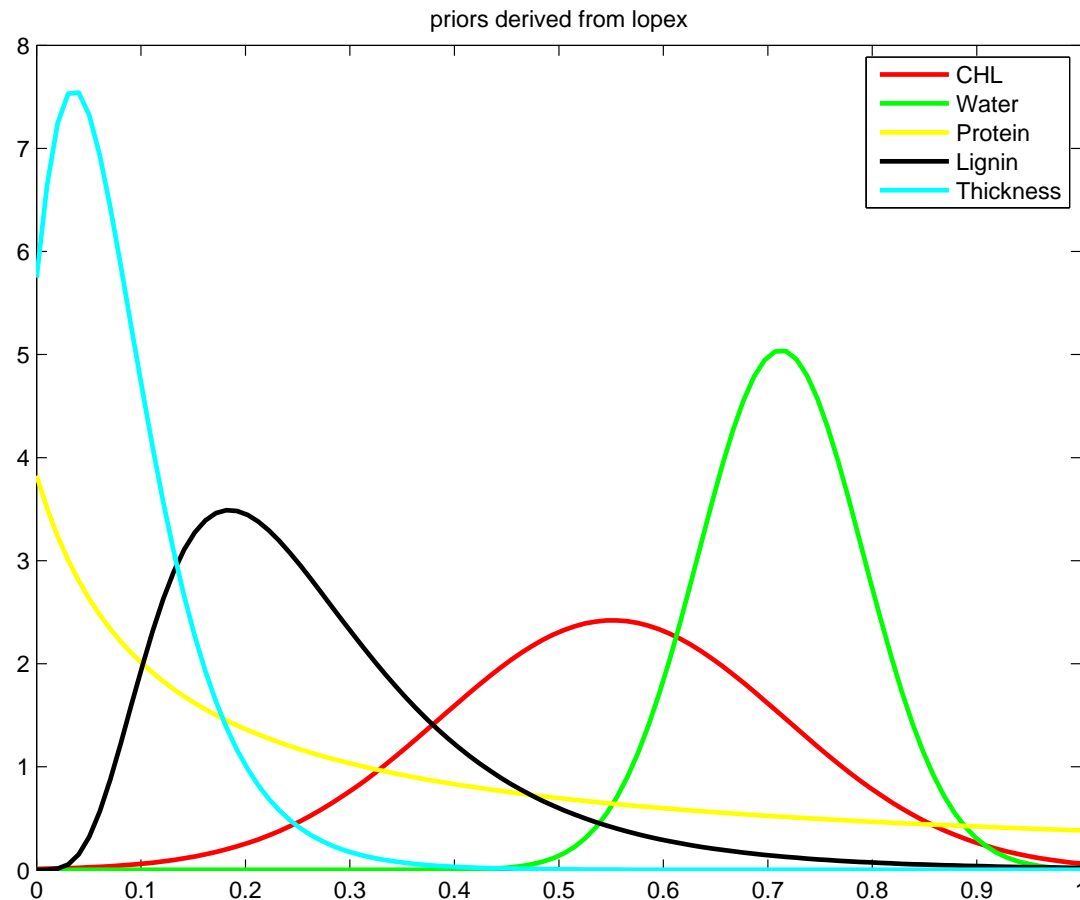
Distributions of the Sensitivity Indices



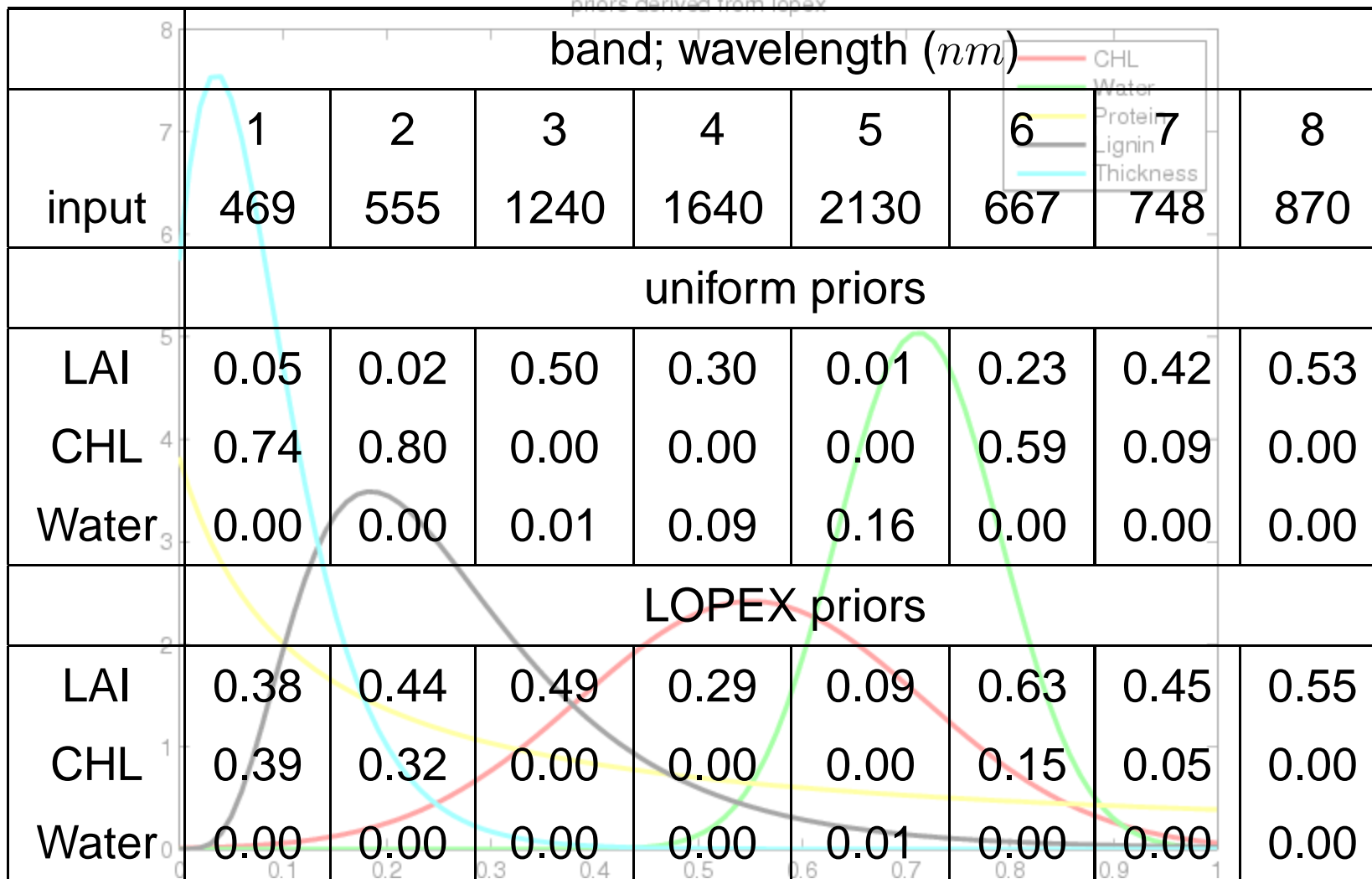
Large values of Total Sensitivity Indices when first order SI is close to zero indicate important interaction effects.

Reducing the Uncertainty on the Inputs

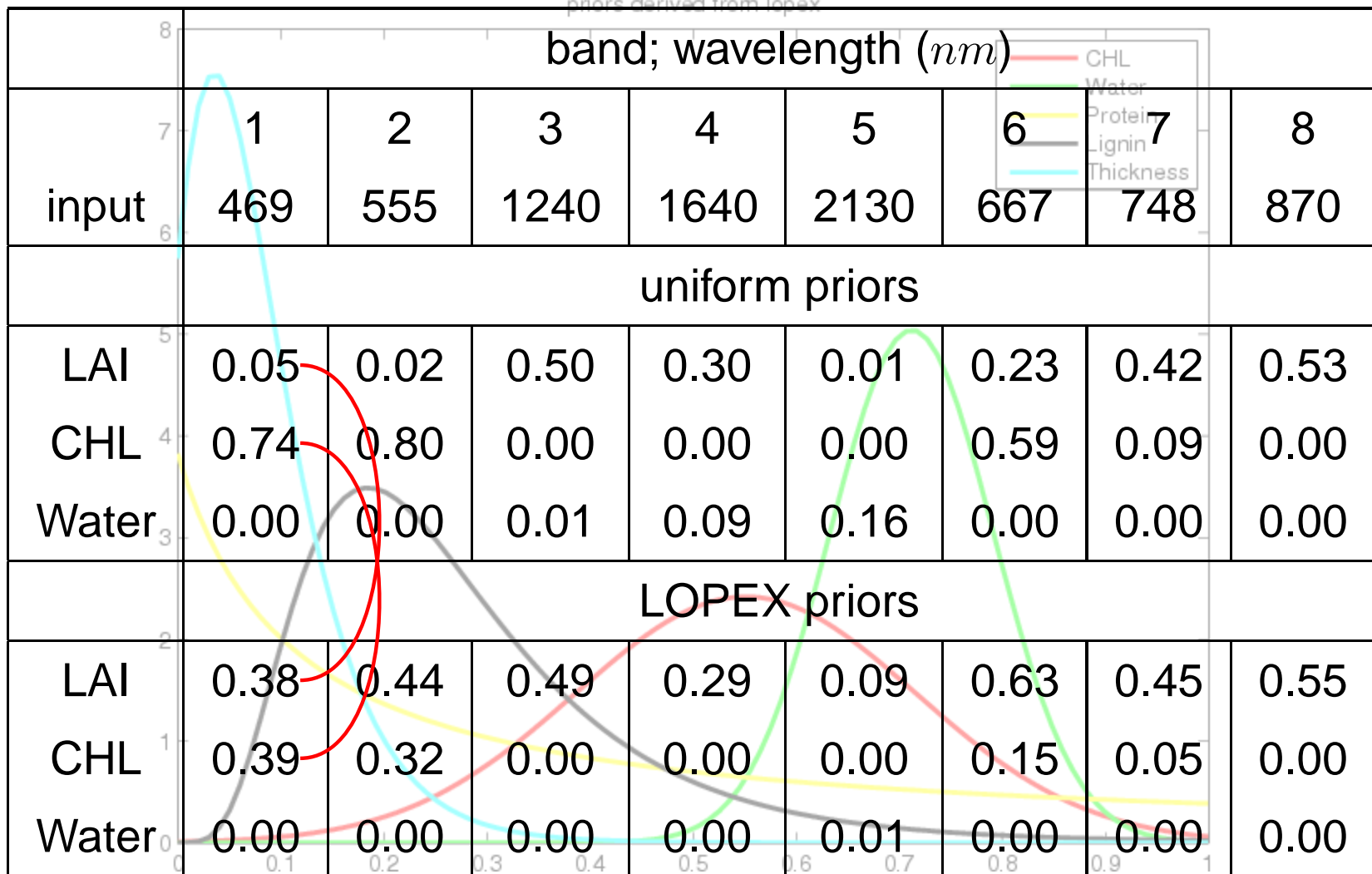
Better priors on the inputs from analysis of the LOPEX (Leaf Optical Properties EXperiment) database.



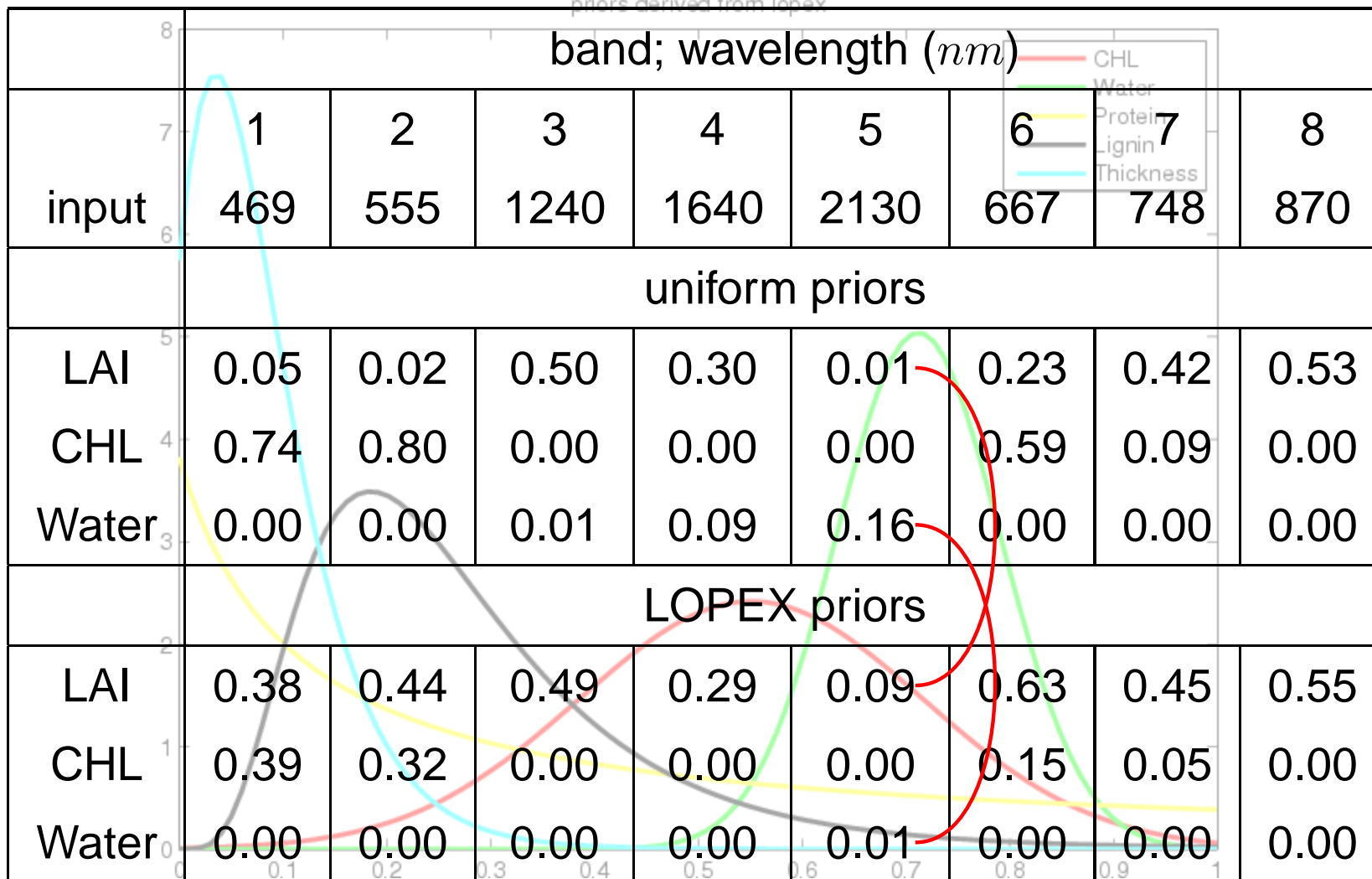
Reducing the Uncertainty on the Inputs



Reducing the Uncertainty on the Inputs



Reducing the Uncertainty on the Inputs



Current and Future Work

Inverting the LCM to produce LAI estimates from MODIS data, using the GP as a fast approximation to the LCM.

Current and Future Work

Inverting the LCM to produce LAI estimates from MODIS data, using the GP as a fast approximation to the LCM.

Studying the effect of including field data as priors on the model inputs when performing inversion.

Current and Future Work

Inverting the LCM to produce LAI estimates from MODIS data, using the GP as a fast approximation to the LCM.

Studying the effect of including field data as priors on the model inputs when performing inversion.

Calibrating the LCM by estimating a bias function from areas where there are both field data and remote sensed data.

Conclusions

Developed main effects and sensitivity indices for the LCM
RTM

Extended the framework to account for uncertainty in the
estimated GP emulator.

Provides insight for model improvement.

Results provided new information to the domain scientists.

Extending this work to validation and inversion.

Conclusions

Developed main effects and sensitivity indices for the LCM
RTM

Extended the framework to account for uncertainty in the
estimated GP emulator.

Provides insight for model improvement.

Results provided new information to the domain scientists.

Extending this work to validation and inversion.

Questions?