ESTIMATING BIOPHYSICAL VARIABLE DEPENDENCES WITH KERNELS

G. Camps-Valls[1], *D. Tuia*[2], *V. Laparra*[1] *and J. Malo*[1]

[1]Image Processing Laboratory (IPL). Universitat de València, Spain.
{gcamps,lapeva,jmalo}@uv.es, http://www.valencia.edu/gcamps
[2]Institute of Geomatics and Analyis of Risk (IGAR), University of Lausanne, Switzerland
devis.tuia@unil.ch, http://devis.tuia.googlepages.com

1. INTRODUCTION

This paper introduces a nonlinear measure of dependence between random variables in the context of remote sensing data analysis. The so-called Hilbert-Schmidt Independence Criterion (HSIC) is a kernel method for evaluating statistical dependence. HSIC is based on computing the Hilbert-Schmidt norm of the cross-covariance operator of mapped samples in the corresponding Hilbert spaces. The HSIC empirical estimator is very easy to compute and has good theoretical and practical properties. We exploit the capabilities of HSIC to explain nonlinear dependences in several remote sensing problems, such as temperature estimation or chlorophyll concentration prediction from spectra. Results show that, when the relationship between random variables is nonlinear or very noisy, the HSIC criterion outperforms other standard methods, such as the linear correlation or multi-information.

2. MEASURING INDEPENDENCE WITH KERNELS

This section presents a criterion for measuring general forms of dependence between random variables. We first start with a simple example of linear dependence to motivate the need for measuring nonlinear dependence. Then the HSIC criterion formulation is presented.

2.1. Linear dependence between random variables

Let us consider two spaces $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ and $\mathcal{Y} \subseteq \mathbb{R}^{d_y}$, on which we jointly sample observation pairs $(\mathbf{x}, \mathbf{y})$ from distribution $\mathbb{P}_{\mathbf{xy}}$. The covariance matrix is defined as $\mathcal{C}_{\mathbf{xy}} = \mathbb{E}_{\mathbf{xy}}(\mathbf{xy}^\top) - \mathbb{E}_{\mathbf{x}}(\mathbf{x})\mathbb{E}_{\mathbf{y}}(\mathbf{y}^\top)$, where $\mathbb{E}_{\mathbf{xy}}$ is the expectation with respect to $\mathbb{P}_{\mathbf{xy}}$, $\mathbb{E}_{\mathbf{x}}$ is the expectation with respect to the marginal distribution $\mathbb{P}_{\mathbf{x}}$, and $\mathbf{y}^\top$ is the transpose of $\mathbf{y}$. The covariance matrix encodes all second order dependences between the random variables. A statistic that summarizes the content of this matrix is its Hilbert-Schmidt norm. The square of this norm is equivalent to the squared sum of its eigenvalues λ_i, $\|\mathcal{C}_{\mathbf{xy}}\|^2_{\text{HS}} = \sum_i \lambda_i^2$. This quantity is zero if and only if there exists no first order dependence between $\mathbf{x}$ and $\mathbf{y}$. Note that the Hilbert-Schmidt norm is limited to the detection of first order relations, and thus more complex (higher-order effects) cannot be captured.

2.2. Measuring dependence with kernels

The nonlinear extension of the notion of covariance was proposed in [1]. Essentially, let us define a (possibly non-linear) mapping $\boldsymbol{\phi} : \mathcal{X} \to \mathcal{F}$ such that the inner product between features is given by a positive definite (p.d.) kernel function $k_x(\mathbf{x}, \mathbf{x}') = \langle \boldsymbol{\phi}(\mathbf{x}), \boldsymbol{\phi}(\mathbf{x}')\rangle$. The feature space $\mathcal{F}$ has the structure of a reproducing kernel Hilbert space (RKHS). Let us now denote another feature map $\boldsymbol{\psi} : \mathcal{Y} \to \mathcal{G}$ with associated p.d. kernel function $k_y(\mathbf{y}, \mathbf{y}') = \langle \boldsymbol{\psi}(\mathbf{x}), \boldsymbol{\psi}(\mathbf{x}')\rangle$. Then, it is possible to define a cross-covariance operator between these feature maps, similarly to the standard (linear) covariance matrix. The cross-covariance operator is a linear operator $\mathcal{C}_{\mathbf{xy}} : \mathcal{G} \to \mathcal{F}$ such that $\mathcal{C}_{\mathbf{xy}} = \mathbb{E}_{\mathbf{xy}}[(\boldsymbol{\phi}(\mathbf{x}) - \mu_x)(\boldsymbol{\psi}(\mathbf{y}) - \mu_y)^\top]$, where $\mu_x = \mathbb{E}_{\mathbf{x}}[\boldsymbol{\phi}(\mathbf{x})]$, $\mu_y = \mathbb{E}_{\mathbf{y}}[\boldsymbol{\psi}(\mathbf{y})]$, and $\mathbf{uv}^\top$ denotes the linear operator $\mathbf{uv}^\top : \mathcal{G} \to \mathcal{F}$, $\mathbf{w} \mapsto \mathbf{u}\langle \mathbf{v}, \mathbf{w}\rangle$. See more details in [2, 3]. The squared norm of the cross-covariance operator, $\|\mathcal{C}_{\mathbf{xy}}\|^2_{\text{HS}}$, is called the Hilbert-Schmidt Independence Criterion

This paper has been partially supported by the Spanish Ministry for Education and Science under projects TEC2006-13845, CSD2007-00018 and AYA2008-05965-C04-03, and by the Swiss National Foundation under grants no. PBLAP2-127713 and 200021-126505.

(HSIC) and can be expressed in terms of kernels [1]. Given a sample dataset $Z = \{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_m, \mathbf{y}_m)\}$ of size m drawn from $\mathbb{P}_{\mathbf{xy}}$, an empirical estimator of HSIC is [1]:

$$\text{HSIC}(\mathcal{F}, \mathcal{G}, \mathbb{P}_{\mathbf{xy}}) = \frac{1}{m^2}\text{Tr}(\mathbf{K}_x\mathbf{H}\mathbf{K}_y\mathbf{H}), \tag{1}$$

where Tr is the trace (the sum of the diagonal entries), $\mathbf{K}_x$, $\mathbf{K}_y$ are the kernel matrices for the (possibly multidimensional) random variables $\mathbf{x}$ and $\mathbf{y}$, respectively, and $H_{ij} = \delta_{ij} - \frac{1}{m}$ centres the variables in $\mathcal{F}$ and $\mathcal{G}$. Here δ represents the Kronecker symbol, where $\delta_{i,j} = 1$ if $i = j$, and zero otherwise.

Note that the actual HSIC is the Hilbert-Schmidt norm of an operator mapping between potentially infinite dimensional spaces, and thus would give rise to an infinitely large matrix. However, due to the *'kernelization'*, the empirical HSIC only depends on computable matrices of size $m \times m$.

In [1], several statistical tests of independence based on the empirical HSIC estimator (1) were proposed. The test should discern between the null hypothesis $\text{H}_0 : \mathbb{P}_{\mathbf{xy}} = \mathbb{P}_{\mathbf{x}}\mathbb{P}_{\mathbf{y}}$ (factorization means independence), and the alternative hypothesis $\text{H}_1 : \mathbb{P}_{\mathbf{xy}} \neq \mathbb{P}_{\mathbf{x}}\mathbb{P}_{\mathbf{y}}$. This is done by comparing the test statistic HSIC with a given threshold. Among the possibilities to define such a threshold over the HSIC estimate, a reasonable one is to approximate the null distribution as a two-parameter gamma distribution, $\widehat{\text{HSIC}} \sim \frac{x^{a-1}e^{-x/b}}{mb^a\Gamma(a)}$, where $a = \mathbb{E}[\widehat{\text{HSIC}}]^2/\mathbb{V}[\widehat{\text{HSIC}}]$ and $b = \mathbb{V}[\widehat{\text{HSIC}}]/\mathbb{E}[\widehat{\text{HSIC}}]$. Note that to compute the test quantile, we need empirical estimates of both $\mathbb{E}[\widehat{\text{HSIC}}]$ and $\mathbb{V}[\widehat{\text{HSIC}}]$, whose detailed expressions can be found in [1] (Theorems 3 and 4). Then, the threshold θ is computed through the inverse cumulative density function (cdf) of the $1 - \alpha$ value, where α is the adopted significance level (typically, $\alpha = 0.05$ or $\alpha = 0.01$). Two random variables are then considered dependent if $\widehat{\text{HSIC}} \geq \theta$, and independent otherwise.

3. EXPERIMENTAL RESULTS

This section is devoted to the design and application of different statistical dependence approaches based on HSIC to establish the dependence between biophysical variables. The population HSIC is zero at independence, so the sample is unlikely to be independent when the empirical HSIC is large. The significance of the result is tested through the HSIC threshold, which is a user-specified quantile of the empirical HSIC distribution at independence. When he empirical HSIC exceeds this threshold, $\widehat{\text{HSIC}} \geq \theta$, we reject the independence hypothesis.

3.1. Temperature estimation

We consider the case of temperature estimation from thermal infrared (TIR) remotely sensed data. In this scenario, land/sea surface temperature (T_s) and emissivity (ε) are the two main geo-biophysical variables to be retrieved from TIR data, since most of the energy detected by the sensor in this spectral region is directly emitted by the land surface.

Table 1. Dependence relations tested with the estimated HSIC and its threshold θ, mutual information, I, and the Pearson's correlation coefficient $\boldsymbol{\rho}$ (and the corresponding p-value) for a significance level of $\alpha = 0.05$.

Relation	HSIC	θ	I	ρ	p-value
$T - T_s$	82.34	0.81	0.4826	0.994	0
$T - W$	56.67	0.66	0.1494	0.748	0
$T - \varepsilon$	0.60	0.55	0.0013	0.056	0.23
$T_s - \varepsilon$	0.34	0.55	0.0009	0.009	0.46
$W - \varepsilon$	0.29	0.62	0.0004	0.016	0.36

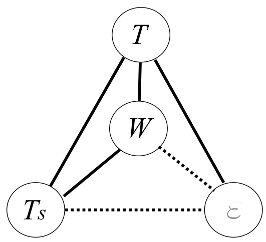
Both quantities, T_s and ε, are coupled and constitute a typical problem in remote sensing referred as to the "temperature and emissivity separation problem". On the one hand, models for estimating land temperature, T_s, typically involve simple parameterizations of at-sensor brightness temperatures, T, the mean and/or differential emissivities ($\bar{\varepsilon}$ and $\Delta\varepsilon$), and the total atmospheric water vapor content, W. Knowing the surface emissivities and the water vapor is difficult as it involves complex radiative inversion models [4]. On the other hand, models for estimating the ε have been devised, either by retrieving ε only from thermal infrared (TIR) data, T, or from visible near-infrared (VNIR) data [5].

We evaluate the capabilities of the HSIC criterion to assess the dependence and interaction between the involved variables in the process, i.e. W, T, ε and T_s. From physical knowledge, it is assumed that T_s and ε *depend* of W and T, and that improved models for temperature estimation involve estimating the surface emissivities as well. In addition, it is also known that spectral emissivity is nearly *independent* of T for most common materials and temperature of the terrestrial environment[1]. To assess such relations, we consider synthetic data simulating ASTER sensor

[1]Only a small marginal dependence is typically observed for high temperature values.

conditions [4]. We included in the data a total amount of 101 ε spectra for natural surfaces extracted from the ASTER spectral library and 61 atmospheric profiles extracted from the Thermodynamic Initial Guess Retrieval database. Simulations were performed at 0^o so no angular effects due to the wide swath angle of low-resolution sensors are considered. By taking into account the five different T acquired values, the 61 atmospheric profiles, and the 108 emissivity spectra, a total of 6588 data points is available.

Fig. 1. Identified dependence graph between variables by HSIC. Thick (dashed) lines represent strong (weak) variable dependence.



The dependence graph between the considered variables (W, T, T_s, ε) was assessed estimating both the HSIC and the Pearson's correlation coefficient $\boldsymbol{\rho}$ between all pairwise variable combinations, and the multi-information, I estimated with the method in [6], which does not explicitly estimate the densities. Note, however, that even though the mutual information is an excellent dependence measure, its finite sample empirical estimate is not easy, especially for continuous variables as in this case.

We randomly selected 500 samples from the dataset, and show the averaged results for 10 realizations in Table 1. In all cases, we fixed the significance value at $\alpha = 0.05$. It can be noticed that all methods detect stronger dependence for T-T_s and T-W, as expected. More interesting is the case of the relation between the at-sensor brightness temperatures, T and the emissivity ε: HSIC identifies a dependence, yet marginally, while the other methods yield a low multi-information value, as well as very low (but not statistically significant) linear correlation. This result justifies the need of identifying and exploiting nonlinear relations in emissivity estimation models, as done elsewhere [4]. Finally, HSIC succeeds in identifying independence between emissivity and T_s or W. In these particular cases, low correlations are obtained but these are not statistically significant, $p > \alpha$. The identified dependence graph by the proposed HSIC method is shown in Fig. 1. Note that the expected dependences have been correctly identified from data.

3.2. Oceanic Chlorophyll-a concentration estimation

The estimation of oceanic chlorophyll concentration from radiance remote sensing data allows us to assess the global ocean phytoplankton mass. Accurate prediction methods lead to understanding and modeling ocean color and dynamics, and to better describe the relationship between the remote sensing reflectance and the backscattering-to-absorption ratio. In this context, we aim at analyzing the nonlinear statistical dependency between the reflectance channels and the measured Chl-a concentration.

For this purpose, we used the SeaBAM dataset [7], which gathers 919 *in-situ* measurements of chlorophyll concentration around the United States and Europe related to five different multispectral remote sensing reflectance bands that correspond to some of the SeaWiFS wavelengths (412, 443, 490, 510 and 555 nm). The chlorophyll concentration values span an interval between 0.019 and 32.787 mg/m^3.

We analyzed the pairwise dependencies among all variables (spectral radiances and the *in situ* Chlorophyll-a concentration measurement). We used all available data and tested dependencies with HSIC (and the corresponding estimated threshold value), mutual information, I, and the Pearson's correlation coefficient $\boldsymbol{\rho}$ (and the corresponding p-value) for a significance level of $\alpha = 0.05$. Results are summarized in Fig. 2. Looking at the correlation coefficient dependency matrix, a strong (linear) correlation is observed between spectral reflectance at 555 nm and chlorophyll concentration, but also among reflectances at 412, 443, and 490 nm. These relations are not used in the literature for developing bio-optical models of chlorophyll concentration estimation, which typically focus on ratios between reflectances at 490 nm or 510 nm versus the 555 nm band.

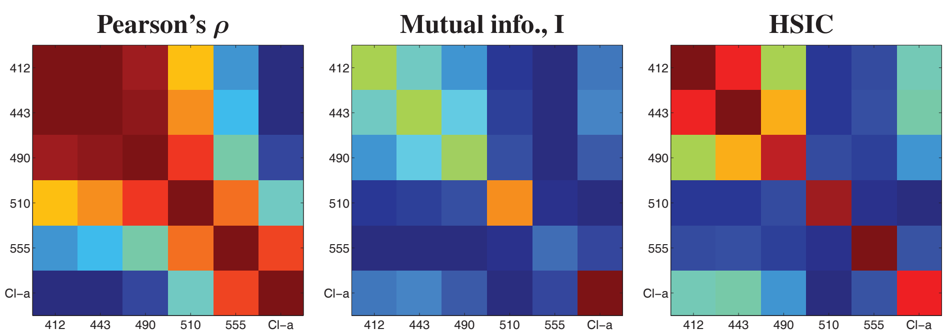


Fig. 2. Pairwise dependency relations between spectral channels and Chlorophyll-a concentration tested with different methods.

The correlation coefficient successfully identifies these relations, and goes in line with the most widely used OC4 model [7]. The mutual information confirms some of these relations, particularly between pairs 412-443 nm and 490-443 nm. Concerning the HSIC dependency estimations, one can observe that most of the channels (except 510 nm) have a high dependency on Chl-a concentration. However, some dependencies are more evident: the two bands depending most on Chl-a are at 443 nm and 412 nm, which matches the parameterizations considered in Morel-1 and Morel-3 bio-optical models, and the more advanced OC4 model. This may be due to the good signal-to-noise ratios of these bands. It is also shown that channels at 443 and 412 nm are tightly dependent, which could constitute a masking effect to be tested, probably due to the high noise level in those specific bands.

4. CONCLUSIONS

This paper presented a kernel method for remote sensing variable nonlinear dependence estimation. The method has very good theoretical and practical properties and can discover nonlinear dependencies between random variables when the space is not sufficiently sampled or in noisy situations. More examples and theoretical analysis will be given at the time of the conference.

5. ACKNOWLEDGMENTS

The authors wish to thank J. C. Jiménez-Muñoz from the Image Processing Laboratory (IPL) at the Universitat de València (Spain) for providing the ASTER simulated data.

6. REFERENCES

[1] A. Gretton, O. Bousquet, A. J. Smola, and B. Schölkopf, "Measuring statistical dependence with Hilbert-Schmidt norms," in *Proceedings Algorithmic Learning Theory*, S. Jain and W.-S. Lee, Eds., 2005.

[2] C. Baker, "Joint measures and cross-covariance operators," *Trans. Am. Math. Soc.*, vol. 186, pp. 273–289, 1973.

[3] Kenji Fukumizu, Francis R. Bach, and Michael I. Jordan, "Dimensionality reduction for supervised learning with reproducing kernel hilbert spaces," *J. Mach. Learn. Res.*, vol. 5, pp. 73–99, 2004.

[4] J. A. Sobrino, J. C. Jimenez-Muñoz, G. Sòria, M. Romaguera, L. Guanter, J. Moreno, A. Plaza, and P. Martínez, "Land surface emissivity retrieval from different VNIR and TIR sensors," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 2, pp. 316–327, Feb 2008.

[5] J. C. Jimenez-Muñoz and J. A. Sobrino, "Split-window coefficients for land surface temperature retrieval from low-resolution thermal infrared sensors," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 806–809, Oct 2008.

[6] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E*, vol. 69, no. 6, pp. 066138, 2004.

[7] J. E. O'Reilly, S. Maritorena, B. G. Mitchell, D. A. Siegel, K. Carder, S. A. Garver, M. Kahru, and C. McClain, "Ocean color chlorophyll algorithms for SeaWiFS," *Journal of Geophysical Research*, vol. 103, no. C11, pp. 24937–24953, Oct 1998.