

Título: MINERÍA DE DATOS EN EL ANÁLISIS DE LAS FIRMAS DE CULTIVOS AGRÍCOLAS

Nombre: Rodríguez Moreno, Fernando

Universidad: Universidad de Extremadura

Departamento: Expresión gráfica

Fecha de lectura: 22/02/2013

Programa de doctorado: Ciencia y Tecnología Industrial y Gráfica

Dirección:

> **Director:** ÁNGEL MANUEL FELICÍSIMO PÉREZ

Tribunal:

> **presidente:** AURORA CUARTERO SÁEZ

> **secretario:** Vojtěch Lukas

> **vocal:** MARÍA DEL HENAR PRIETO LOSADA

> **vocal:** MARÍA EUGENIA POLO GARCÍA

> **vocal:** CARLOS MARIO CAMPILLO TORRES

Descriptor:

> AGRONOMIA

> RADIOMETRIA

> INTELIGENCIA ARTIFICIAL

El fichero de tesis ya ha sido incorporado al sistema

Localización: BIBLIOTECA CENTRAL DE BADAJOZ

Resumen: Introducción

Los agricultores son cada día más conscientes de la necesidad de aplicar el tratamiento agronómico específico que precisa el cultivo de tal forma que el rendimiento de los insumos sea máximo y el desarrollo sostenible (Barker and Pilbeam 2007; Jones 1998; Marschner 1995; Mengel and Kirkby 2001). De esta forma se obtiene el máximo beneficio hoy y se garantiza la rentabilidad de la actividad agrícola en el futuro.

En la consecución de esta meta la agricultura de precisión juega un papel determinante. La agricultura de precisión pone a disposición del agricultor la última tecnología (Srinivasan 2006; Stafford 1997), gracias a la cual es posible cartografiar de forma precisa, rápida y barata el campo y aplicar un tratamiento diferencial ajustado a las necesidades específicas de cada una de las unidades tierra que lo componen.

La correcta delimitación de las unidades tierra así como el certero diagnóstico agronómico de las mismas se puede conseguir aunando conocimientos agronómicos y técnicas de minería de datos, evidenciarlo es el propósito de esta tesis. Este binomio está fuera del alcance de muchos agricultores por lo que es necesario desarrollar metodologías de muestreo/análisis más sencillas y tratar de desarrollar sistemas de apoyo a la

decisión, de forma que el agricultor pueda encontrar las respuestas que necesita con la ayuda de su consultor agrícola y del laboratorio.

Desarrollo teórico

La tesis comienza evaluando el verdadero potencial de los 21 índices espectrales de vegetación (radiometría) más ampliamente usados para estimar la concentración de un nutriente, el nitrógeno, en planta (Blackburn 2007; Panda et al. 2010; Raymond Hunt et al. 2011; Schellberg et al. 2008). Prueba realizada en unas condiciones que simulan una explotación agrícola real (amplio rango de condiciones de desarrollo). Es preciso conocer si los índices espectrales de vegetación son una solución en la práctica, si empleándolos es posible obtener estimaciones correctas, rápidas y baratas del estado nutricional del cultivo. En caso afirmativo habría que dirigir los esfuerzos a proyectos de demostración con los que animar la transferencia de la tecnología a los agricultores.

En esta tesis se dedica un apartado especial al procedimiento de evaluación de los resultados. Confundir una relación descriptiva con una predictiva conduciría al desarrollo de un modelo cuyas recomendaciones no tendrían ningún valor, lo que haría que los objetivos de la tesis fueran inalcanzables. Para obtener evaluaciones realistas se toman 4 medidas, una ya comentada, experimentar en un amplio rango de condiciones de desarrollo. La segunda medida consiste en forzar al modelo que ha de relacionar el índice espectral de vegetación y la concentración de nutriente a que sea válido durante todo el período durante el cual sería posible actuar para corregir una potencial deficiencia (novedad). La tercera medida es el uso de la validación cruzada en la evaluación de las relaciones y la cuarta medida es el empleo en el ensayo del cultivo que presenta mayores dificultades para el desarrollo de modelos radiométricos, el triticale de doble propósito.

Verato es el cultivar del triticale (X Triticosecale Wittmack) que soporta el pastoreo del ganado durante su desarrollo sin arruinar la cosecha final. Esta peculiaridad dificulta la consecución del objetivo, el desarrollo de un modelo radiométrico eficaz y generalizable, al no poder confiar en el verdor como estimador de la concentración de nitrógeno (después del pastoreo el cultivo amarillea debido a un desajuste entre crecimiento y síntesis de clorofila, no por déficit de nitrógeno). Eso obliga a la búsqueda de rasgos espectrales más directamente relacionados con la concentración de nitrógeno, los cuales respondan correctamente incluso en condiciones complicadas.

Gracias a estas medidas se pretende determinar el umbral mínimo de eficacia de la metodología, se aumentan las garantías de poder predictivo del modelo y se reducen costes de implementación, facilitando así la transferencia de los resultados a explotaciones agrícolas reales.

En el mismo escenario y con los mismos datos y procedimiento de evaluación se estudia si las dos técnicas de reducción de dimensiones más potentes, Análisis de componentes principales (PCA) (Shlens 2005) y Análisis de componentes independientes (ICA) (Hyvärinen and Oja 2000), son adecuadas para el procesamiento de los datos obtenidos en el muestreo hiperspectral del cultivo (firma espectral completa). Si las técnicas son efectivas entonces concentrarán la información relativa al estado nutricional del cultivo en una decena de nuevas componentes, con las que sería fácil desarrollar modelos de regresión con los que estimar eficazmente la concentración de nitrógeno en planta.

Un resultado positivo en el estudio con las técnicas de reducción de dimensiones difícilmente sería un resultado transferible, esto es así dado el coste del espectroradiómetro necesario para obtener la firma espectral. El

objetivo de este estudio es dar un paso intermedio, comprobar que aún con todas las garantías exigidas es posible desarrollar un modelo efectivo con capacidad de generalización en el espacio y tiempo (dentro de la misma campaña).

La última etapa en esta línea de investigación es la evaluación de la capacidad de los árboles de decisión (Gehrke 2006; Loh 2011; Ruß and Brenning 2010) para estimar la concentración de nitrógeno en planta, empleando para ello la reflectancia de la planta en unas pocas longitudes de onda. Esta investigación se realiza en las mismas condiciones que el estudio con las técnicas de reducción de dimensiones.

Los árboles de decisión evaluados no emplearán más de tres longitudes de onda (reflectancia), de esta forma no será necesaria la participación de un caro espectroradiómetro de campo para realizar la estimación del estado nutricional del cultivo, superando el problema que tiene el trabajo con las técnicas de reducción de dimensiones.

No todo cambio en la concentración de nitrógeno en la planta tiene un efecto sensible en la misma, existiendo por tanto cierta incertidumbre en el cálculo del plan de fertilización. En consecuencia estimar la concentración de nitrógeno en planta sin una precisión de varios decimales, tal y como consigue el laboratorio, no tiene efectos significativos en la gestión agrícola.

Hacer que la salida del árbol de decisión sea un nivel para la concentración de nitrógeno en lugar de un valor concreto puede suponer una mejora en su efectividad. Esto sería así si el algoritmo de clasificación, puntos de ruptura naturales (Jenks 1967), agrupa de tal forma que resulte más fácil encontrar rasgos espectrales distintivos para cada nivel, lo que facilitaría la tarea a los árboles de decisión.

En la otra línea de estudio, el diagnóstico agronómico de los campos conocida la concentración de los nutrientes y otros parámetros (altura, verdor, ζ) de las plantas, se expondrán los trabajos con dos explotaciones agrícolas ubicada en centro Europa (Chequia), sembradas de trigo de invierno.

Lo primero es la identificación de un índice de barata, fácil y rápida determinación que permita una correcta estimación del desarrollo de las plantas. Ese índice puede suministrar valiosa información al servicio de la gestión agrícola y una referencia válida con la que comparar, mediante regresiones no lineales y validaciones cruzadas, las concentraciones de los nutrientes en la búsqueda de una relación con significación estadística fruto de un vínculo causa-efecto que pueda ser empleado en la toma de decisiones.

Dada la dificultad de trabajar con pequeños conjuntos de datos, se evaluará el uso de los factores espaciales (funciones de superficie y variables topográficas) para:

¿Verificar la relación espacial entre las muestras obtenidas en el mismo campo.

¿Realizar una validación espacial de las relaciones encontradas entre el índice de desarrollo del cultivo y los nutrientes.

¿Identificar factores limitantes no-nutrientes (textura, orientación, ζ).

¿Interpolar los datos obtenidos en el muestreo a todo el campo. Esto último es muy importante dado que es posible efectuar el diagnóstico apoyándose en un muestreo de bajo coste del campo y por ello la geoestadística no es una alternativa de garantía por las dificultades para obtener un variograma fiable (Oliver 2010).

Conclusión

En la firma espectral del cultivo existen rasgos relacionados con el nivel de nitrógeno en planta, los cuales permanecen estables durante un amplio intervalo fenológico.

Aún en un amplio rango de condiciones de desarrollo, es posible desarrollar modelos radiométricos eficaces, pero se desconoce si han de ser calibrados cada campaña y si sería posible ajustarlos para operar con medidas de reflectancia del dosel de vegetación.

La metodología SIOPP ha probado ser eficaz realizando el diagnóstico agronómico de dos explotaciones agrícolas reales en las que los problemas eran complejos. Con las evidencias disponibles y dado el bajo coste añadido que implica su implementación, ya se considera una alternativa al uso de rangos de suficiencia o relaciones entre nutrientes para la interpretación de los análisis foliares.

Cuando no existan zonas en el campo con un buen desarrollo del cultivo, zonas de referencia, SIOPP no podrá realizar el diagnóstico. Las condiciones de desarrollo del cultivo han de estar alejadas del óptimo, esa primera corrección, obvia revisando un tratado de agronomía, es un requisito de la metodología en casos extremos. En caso de existir severa heterogeneidad (márgenes abruptos) o si hay que diagnosticar plantaciones con centenares de hectáreas de extensión, SIOPP puede presentar problemas ante la debilidad de modelar barreras y/o infinidad de gradientes. En ese caso sería necesario realizar un estudio previo para la delimitación de las zonas de gestión, esta es una de las líneas de trabajo futuras.

Bibliografía

- Abdelhamid, M., Horiuchi, T. & Oba, S. (2003). Evaluation of the SPAD value in faba bean (*Vicia faba* L.) leaves in relation to different fertilizer applications. *Plant Production Science*, 6(3), 185-189, doi:10.1626/pps.6.185.
- Agbangba, E. C., Olodo, G. P., Dagbenonbakin, G. D., Kindomihou, V., Akpo, L. E. & Sokpon, N. (2011). Preliminary DRIS model parameterization to access pineapple variety 'Perola' nutrient status in Benin (West Africa). *African Journal of Agricultural Research*, 6(27), 5841-5847, doi:10.5897/ajar11.889.
- Alves Mourão Filho, F. D. A. (2005). DRIS and sufficient range approaches in nutritional diagnosis of "Valencia" sweet orange on three rootstocks. *Journal of Plant Nutrition*, 28(4), 691-705, doi:10.1081/pln-200052645.
- Amundson, R. & Koehler, F. (1987). Utilization of DRIS for diagnosis of nutrient deficiencies in winter wheat. *Agronomy Journal*, 79(3), 472-476.
- Arregui, L. M., Lasa, B., Lafarga, A., Irañeta, I., Baroja, E. & Quemada, M. (2006). Evaluation of chlorophyll meters as tools for N fertilization in winter wheat under humid Mediterranean conditions. *European Journal of Agronomy*, 24(2), 140-148, doi:10.1016/j.eja.2005.05.005.
- Bannari, A., Khurshid, K. S., Staenz, K. & Schwarz, J. W. (2007). A comparison of hyperspectral chlorophyll indices for wheat crop chlorophyll content estimation using laboratory reflectance measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10), 3063-3074, doi:10.1109/tgrs.2007.897429.
- Barker, A. V. & Pilbeam, D. J. (2007). *Handbook of plant nutrition*: CRC press.
- Blackburn, G. A. (2007). Hyperspectral remote sensing of plant pigments. *Journal of Experimental Botany*, 58(4), 855-867, doi:10.1093/jxb/erl123.
- Camacho, M. A., da Silveira, M. V., Camargo, R. A. & Natale, W. (2012). Normal nutrient ranges by the CHM, DRIS and CND methods and critical level by method of the reduced normal distribution for orange-pera. *Revista Brasileira de Ciencia do Solo*, 36(1), 193-200, doi:10.1590/s0100-06832012000100020.
- Cammarano, D., Fitzgerald, G., Basso, B., Chen, D., Grace, P. & O'Leary, G. (2011). Remote estimation of chlorophyll on two wheat cultivars in two rainfed environments. *Crop and Pasture Science*, 62(4), 269-275, doi:10.1071/cp10100.
- Castelli, F. & Contillo, R. (2009). Using a chlorophyll meter to evaluate the nitrogen leaf content in flue-cured tobacco (*Nicotiana tabacum* L.). *Italian Journal of Agronomy*, 4(2), 3-11.

- Dagbénobakin, G., Agbangba, C., Glèlè Kakai, R. & Goldbach, H. (2010). Preliminary diagnosis of the nutrient status of cotton (*Gossypium hirsutum* L) in Benin (West Africa). *Bulletin de la Recherche Agricole du Bénin*, 67, 32-44.
- De Assis Carvalho Pinto, F., Da Silva Junior, M. C., De Queiroz, D. M. & Vieira, L. B. (2007). Pasture nitrogen status identification using a balloon remote sensing platform. Minneapolis, MN: American Society of Agricultural and Biological Engineers.
- Gehrke, J. (2006). Classification and regression trees. In J. Wong (Ed.), *Encyclopedia of data warehousing and mining* (Vol. 1, pp. 141): Information Science Publishing.
- Gnyp, M. L., Fei, L., Miao, Y., Koppe, W., Jia, L., Chen, X. et al. (2009). Hyperspectral data analysis of nitrogen fertilization effects on winter wheat using spectrometer in North China Plain. *CORD Conference Proceedings*, 1-4, doi:10.1109/WHISPERS.2009.5289007.
- Goel, P. K., Prasher, S. O., Landry, J. A., Patel, R. M., Viau, A. A. & Miller, J. R. (2003). Estimation of crop biophysical parameters through airborne and field hyperspectral remote sensing. *Transactions of the American Society of Agricultural Engineers*, 46(4), 1235-1246.
- Gómez-Casero, M. T., López-Granados, F., Peña-Barragán, J. M., Jurado-Expósito, M., García-Torres, L. & Fernández-Escobar, R. (2007). Assessing nitrogen and potassium deficiencies in olive orchards through discriminant analysis of hyperspectral data. *Journal of the American Society for Horticultural Science*, 132(5), 611-618.
- Graeff, S., Claupein, W., Pfenning, J. & Liebig, H. P. (2009). Evaluation of different active and passive sensor systems to adapt N fertilizer applications in broccoli (*Brassica oleracea* convar. botrytis var. italica). *Acta Horticulturae*, 824, 133-140.
- Heege, H. J., Reusch, S. & Thiessen, E. (2008). Prospects and results for optical systems for site-specific on-the-go control of nitrogen-top-dressing in Germany. *Precision Agriculture*, 9(3), 115-131, doi:10.1007/s11119-008-9055-3.
- Hyvärinen, A. & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural networks*, 13(4-5), 411-430.
- Jenks, G. F. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography*, 7, 186-190.
- Jeong, K. Y., Whipker, B., McCall, I. & Frantz, J. (2009). Gerbera leaf tissue nutrient sufficiency ranges by chronological age. *Acta Horticulturae*, 843, 183-190.
- Jones, J. B. (1998). *Plant nutrition manual*: CRC.
- Li, F., Miao, Y., Hennig, S. D., Gnyp, M. L., Chen, X., Jia, L. et al. (2010). Evaluating hyperspectral vegetation indices for estimating nitrogen concentration of winter wheat at different growth stages. *Precision Agriculture*, 11(4), 335-357, doi:10.1007/s11119-010-9165-6.
- Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 14-23.
- Marschner, H. (1995). *Mineral nutrition of higher plants*: Academic Press.
- Mengel, K. & Kirkby, E. A. (2001). *Principles of Plant Nutrition*: Kluwer Academic Publishers.
- Oliver, A. (2010). *Geostatistical applications for precision agriculture*: Springer.
- Panda, S. S., Hoogenboom, G. & Paz, J. O. (2010). Remote sensing and geospatial technological applications for site-specific management of fruit and nut crops: A review. *Remote Sensing*, 2(8), 1973-1997, doi:10.3390/rs2081973.
- Raymond Hunt, E., Daughtry, C. S. T., Eitel, J. U. H. & Long, D. S. (2011). Remote sensing leaf chlorophyll

- content using a visible band index. *Agronomy Journal*, 103(4), 1090-1099, doi:10.2134/agronj2010.0395.
- Rengel, Z. (1998). *Nutrient use in crop production*: CRC.
- Reuter, D., Robinson, B., Mader, P. & Tlustos, P. (1998). *Plant Analysis: An Interpretation Manual*. *Biologia Plantarum*, 41(2), 317-317.
- Ruß, G. & Brenning, A. (2010). Data mining in precision agriculture: Management of spatial information. In E. Hüllermeier, R. Kruse & F. Hoffmann (Eds.), *IPMU'10 Proceedings of the Computational intelligence for knowledge-based systems design and 13th international conference on Information processing and management of uncertainty* (Vol. 6178, pp. 350-359, *Lecture Notes in Computer Science*). Dortmund: Springer Berlin / Heidelberg.
- Schellberg, J., Hill, M. J., Gerhards, R., Rothmund, M. & Braun, M. (2008). Precision agriculture on grassland: Applications, perspectives and constraints. *European Journal of Agronomy*, 29(2-3), 59-71, doi:10.1016/j.eja.2008.05.005.
- Shlens, J. (2005). *A tutorial on principal component analysis*: Systems Neurobiology Laboratory, University of California at San Diego.
- Srinivasan, A. (2006). *Handbook of precision agriculture: Principles and applications*: Taylor & Francis.
- Stafford, J. V. (1997). *Precision Agriculture '97: Spatial variability in soil and crop*: BIOS Scientific Pub.
- Stalenga, J. (2007). Applicability of different indices to evaluate nutrient status of winter wheat in the organic system. *Journal of Plant Nutrition*, 30(3), 351-365, doi:10.1080/01904160601171207.