year to another. Three different studies of three different sites linked this inter–annual variation to three different drivers: spring temperature, precipitation and *VPD*. The La Thuile data set provides the required data to re–visit this hypothesis and assign (conditional) probabilities to the potential drivers. The braves among the ecologists who accepted the La Thuile challenge soon had to admit their recklessness: the data set contains at least one dimension too much to allow for the usage of established approaches such multiple–regressions, wavelets or self–organizing maps.

Depending on the selected approach and question at hand, the diurnal variability may be removed from the data set by using variables that are representative at the daily scale. The seasonal variability can be accounted for by de–trending the data which implies that the residuals are then analyzed. Although pre–processing reduces the information content of the data set, it may be acceptable from an ecological point of view. The analysis then proceeds and yields for every site in the analysis an individual time series of regression coefficients, U–matrix or wavelet coefficients for respectively multiple regressions, self–organizing maps and wavelets. Although these outputs look similar, quantitative comparison across sites is hampered by the lack of formal tools to post–process these outputs across sites. This spatial aspect is my so–called too–manieth dimension.

If developed, these tools should as well account for the difference between the calendar year and the timing of the processes under study. For example: the period between the time that *NEE* is dominated by CO_2–uptake and the maximum CO_2 uptake might be of particular interest for an analysis. Although this period is easy to identify, it occurs earlier in the calendar year and last longer in temperate compared to boreal regions. Hence, post–processing techniques should be capable of comparing relationships (described by means of regression coefficients, U–matrix or wavelet coefficients) with a different time and length of occurrence.

Related: Luyssaert et al. (2007); In this study a self–organizing map (SOM) is used to describe the relationships between a set of variables describing the weather condition and the CO_2 exchange between forest and the atmosphere. Although the SOM results in interesting relationships, the analysis comes short in synthesizing these relationships across sites.

3.19 RGB imagery as a useful data source in ecological research

Hella E. Ahrends

Recent studies demonstrate the suitability of standard digital cameras or webcams for phenological observations in forests. This data mining strategy is user–friendly, cost–effective and highly objective. RGB cameras provide archivable documentation, data that can be used for quantitative and qualitative evaluations and the applicability of a technique independent from observer skills. They provide information on e.g. meteorological conditions (such as visibility, fog), changes in canopy structure, visible plant responses to drought stress or snow cover and, thus, are most useful in ecosystem monitoring and ecological research. We suggest that this technique has high potential to bridge the gap between spatially integrated species–averaged information from satellites and point observations at the species level. Some applications, mainly with respect to remote sensing phenology, are shown.

Related: Ahrends et al. (2008)

3.20 Environmental informatics and data mining in analysis of CO_2 data

Mika Sulkava

There are many frameworks for learning from data. Commonly used frameworks include statistical learning, exploratory and confirmatory data-analysis, machine learning, knowledge discovery in databases, data mining, and pattern recognition. Environmental statistics and environmental informatics are fields related to learning from environmental data (Sulkava, 2008). The most significant difference between these fields is the importance of the different frameworks of learning. Data analysis in statistics is often confirmatory, i.e., hypotheses are tested and either confirmed or rejected. In addition, data is usually collected from carefully designed experiments, which involve sampling, randomization, replication, controlling for confounding variables, etc. In many cases environmental data from, e.g., monitoring networks cannot be treated the same way as data from controlled experiments. The characteristics of these frameworks and fields and their suitability to solving scientific problems related to CO_2 fluxes and concentrations are discussed.

Three examples of how methods of environmental informatics—especially data mining and exploratory data-analysis—have been successfully applied in analysis of CO_2 data are presented. First, it is shown how parametric curve fitting, time series segmentations, cross-correlations, sparse regression, and neural networks, such as the Self-Organizing Map (SOM, Kohonen, 2001) have been used in finding drivers of anomalies in CO_2 fluxes of pine forests at different latitudes. These analyses revealed that anomalies in net ecosystem exchange (*NEE*) are dominated by anomalies in gross photosynthesis (*GPP*, Luyssaert et al., 2007). Second, it is presented how time series segmentation and trend detection (e.g., Sulkava et al., 2007) with permutation tests and methods for combining p-values have been used for studying changes in CO_2 concentrations between different parts of Europe. This methodology revealed many statistically significant trends in the differences between a coastal station and inland stations. The time-series were analyzed at different resolutions, which exposed some relatively fast changes in addition to the longer-term trends. Third, the use of clustering of the SOM (Vesanto and Sulkava, 2002) in a sampling problem for studying the representativeness of the European CO_2 flux tower network is demonstrated (Canfora et al., 2009). The analysis showed that the network mostly represents the European domain rather well in respect to the variables considered. However, North-Eastern and South-Eastern climatic and ecophysiological conditions were found to be poorly sampled for several plant functional types. The results of all cases above are encouraging for using data-driven methods of environmental informatics in learning from CO_2 data.

Related: Vesanto and Sulkava (2002); Luyssaert et al. (2007); Sulkava et al. (2007); Sulkava (2008); Canfora et al. (2009)

3.21 The evolving patterns of monsoonal precipitation over India

Nishant Malik*, Norbert Marwan & Jürgen Kurths

We present an analysis of a high resolution daily rainfall gridded data set from 1951 to 2007 for India. In the light of changing characteristics of precipitation at the global scale due to atmospheric warming it is of extreme importance to understand how the spatio-temporal behaviour of monsoonal precipitation has evolved over the last few decades. We employ some standard linear trend detection methods and principal component analysis to study the evolution of monsoonal precipitation patterns. We use recurrence plots to calculate spatial complexities and there evolution over the time scales concerned in the data set and correlate them to some new features that has emerged in the rainfall events. Monsoonal precipitation usually occurs in the form of large scale spatial activity and here we show a method to find these scales and divide India into different monsoonal regions based on a measure of event synchronization and then compare it using other measures like cross-correlation and mutual information.

3.22 Information driven ecohydrologic self–organization

Praveen Kumar

Interaction between processes in nature leads to self-organization. For example, the dynamics of vegetation growth is affected by prevailing above and below ground conditions, but the energy, water and carbon fluxes in and out of the vegetation modify this environment itself. This self-organized feedback, facilitated by the variability of the component processes that interact, is continually evolving. Yet, ascertaining the precise role of the variability in this feedback dynamics remains an open question. Towards this goal, using observations of moisture, energy, and carbon fluxes from Fluxnet towers, we measure the predictive information provided by one variable to the

5 Bibliography

Abramowitz, G. 2005. Towards a benchmark for land surface models. *Geophysical Research Letters* **32**, L22702.

Abramowitz, G., R. Leuning, M. Clark, and A. Pitman. 2008. Evaluating the performance of land surface models. *Journal of Climate* **21**, 5468–5481.

Ahrends, H. E., R. Brügger, R. Stöckli, J. Schenk, P. Michna, F. Jeanneret, and W. Wanner. 2008. Quantitative phenological observations of a mixed beech forest in northern Switzerland with digital photography. *Journal of Geophysical Research—Biogeosciences* **113**, G04004.

Arenas, A., A. Diaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou. 2008. Synchronization in complex networks. *Physics Reports* **469**, 93–153.

Atanasova, N., L. Todorovski, D. S., and B. Kompare. 2006. Constructing a library of domain knowledge for automated modelling of aquatic ecosystems. *Ecological Modelling* **194**, 14–36.

Atanasova, N., L. Todorovski, S. Džeroski, and B. Kompare. 2008. Application of automated model discovery from data and expert knowledge to a real–world domain: Lake glumsø. *Ecological Modelling* **212**, 92–98.

Baldocchi, D. 2008. Turner Review No. 15. Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany* **56**, 1–26.

Barbosa, S. M. in press. Changing seasonality in Europe' s air temperature. *The European Physical Journal—Special Topics*.

Barbosa, S. M., M. E. Silva, and M. J. Fernandes. 2009. Multiscale variability patterns in NCEP/NCAR reanalysis sea–level pressure. *Theoreical and Applied Climatology* **96**, 319–326.

Blasius, B., and R. Tönjes. 2005. Quasiregular concentric waves in heterogeneous lattices of coupled oscillators. *Phys. Rev. Lett* **95**, 084101.

Bonan, G. B. 2008. Forest and climate change: forcings, freedbacks, and the climate benefits of forests. *Science* **320**, 1444–1449.

Bub, G., A. Shrier, and L. Glass. 2005. Global organization of dynamics in oscillatory heterogeneous excitable media. *Phys. Rev. Lett.* **94**, 028105.

Bühlmann, P., and T. Hothorn. 2007. Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science* **22**, 477–505.

Canfora, E., D. Papale, S. Luyssaert, and M. Sulkava. 2009. Does the European eddy flux tower network represent the climatic and ecophysiological diversity of europe?. In *European Geosciences Union—General Assembly 2009*. A. Richter (ed.). Vol. 11 of *Geophysical Research Abstracts*. European Geosciences Union. Copernicus Publications. Vienna, Austria. EGU2009-9154-3.

Ciais, P., M. Reichstein, N. Viovy, A. Granier, J. Ogée, V. Allard, N. Buchmann, M. Aubinet, C. Bernhofer, A. Carrara, F. Chevallier, N. D. Noblet, A. Friend, P. Friedlingstein, T. Grünwald, B. Heinesch, P. Keronen, A. Knohl, G. Krinner, D. Loustau, G. Manca, G. Matteucci, F. Miglietta, J. M. Ourcival, K. Pilegaard, S. Rambal, G. Seufert, J. F. Soussana, M. J. Sanz, E. D. Schulze, T. Vesala, and R. Valentini. 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003.. *Nature* **437**, 529–533.

Crucifix, M. 2006. Does the Last Glacial Maximum constrain climate sensitivity?. *Geophysical Research Letters* **33**, L18701.

Crucifix, M., and J. C. Rougier. in press. How to cope with climate's complexity? a paleoclimate application with the particle filter. *European Physics Journal–Special Topics*.

Donges, F., Y. Zou, N. Marwan, and J. Kurths. in press. Complex networks in climate dynamics. comparing linear and nonlinear network construction methods. *European Physical Journal–Special Topics*.

Donner, R., and A. Witt. 2006. Characterisation of long–term climate change by dimension estimates of multivariate palaeoclimatic proxy data. *Nonlinear Processes in Geophysics* **13**, 485–497.

Donner, R., D. Barbosa, J. Kurths, and N. Marwan. in press. Understanding the Earth as a complex system—recent advances in data analysis and modelling in Earth sciences. *European Physical Journal—Special Topics*.

Džeroski, S., and L. Todorovski. 2003. Learning population dynamics models from data and domain knowledge. *Ecological Modelling* **170**, 129–140.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, J. G. Lohmann, B. A. Loiselle, G. Manion, G. Moritz, M. Nakamura, Y. Nakazawa, J. M. C. Overton, P. S. J. Townsend Peterson, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151.

Frei, C., R. Schöll, S. Fukutome, J. Schmidli, and P. L. Vidale. 2006. Future change of precipitation extremes in Europe: an intercomparison of scenarios from regional climate models. *Journal of Geophyical Research* **111**, D06105.

Friedlingstein, P., P. Cox, R. Betts, L. Bopp, W. Von Bloh, V. Brovkin, P. Cadule, S. Doney, M. Eby, I. Fung, G. Bala, J. John, C. Jones, F. Joos, T. Kato, M. Kawamiya, W. Knorr, K. Lindsay, H. D. Matthews, T. Raddatz, P. Rayner, C. Reick, E. Roeckner, K. G. Schnitzler, R. Schnur, K. Strassmann, A. J. Weaver, C. Yoshikawa, and N. Zeng. 2006. Climate–carbon cycle feedback analysis: Results from the (CMIP)–M–4 model intercomparison. *Journal of Climate* **19**, 3337–3353.

Hauhs, M., and H. Lange. 2008. Classification of runoff in headwater catchments: A physical problem?. *Geography Compass* **2**, 235–254.

Heimann, M., and M. Reichstein. 2008. Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature* **451**, 289–292.

Hoyer, P. O., D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf. 2009. Nonlinear causal discovery with additive noise models.. In *Advances in Neural Information Processing Systems 21: Proceedings of the 2008 Conference*. D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou. (Eds.). MIT Press. Cambridge, MA. Pp. 689–696.

Intergovernmental Panel on Climate Change 2007. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge Univ. Press, 358 Cambridge, U. K.

Jung, M., M. Reichstein, and A. Bondeau. 2009. Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosciences Discussions* **6**, 5271–5304.

Katul, G., C.-T. Lai, K. Schäfer, B. Vidakovic, A. J., D. Ellsworth, and R. Oren. 2001. Multiscape analysis of vegetation surface fluxes: from seconds to years. *Advances in Water Resources* **24**, 1119–1132.

Kneib, T., T. Hothorn, and G. Tutz. 2009. Variable selection and model choice in geoadditive regression models. *Biometrics*.

Kohonen, T. 2001. *Self-Organizing Maps*. Vol. 30 of *Springer Series in Information Sciences*. third ed. Springer, Berlin, Germany.

Luyssaert, S., I. A. Janssens, M. Sulkava, D. Papale, A. J. Dolman, M. Reichstein, J. Hollmén, J. G. Martin, T. Suni, T. Vesala, D. Lousteau, B. E. Law, and E. J. Moors. 2007. Photosynthesis drives anomalies in net carbon–exchange of pine forests at different latitudes. *Global Change Biology* **13**, 2110–2127.

Mahecha, M. D., and S. Schmidtlein. 2008. Revealing biogeographical patterns by nonlinear ordinations and derived anisotropic spatial filters. *Global Ecology & Biogeography* **17**, 284–296.

Mahecha, M. D., M. Reichstein, H. Lange, N. Carvalhais, C. Bernhofer, T. Grünwald, D. Papale, and G. Seufert. 2007. Characterizing ecosystem–atmosphere interactions from short to interannual time scales. *Biogeosciences* **4**, 743–758.

Marwan, N., M. Romano, M. Thiel, and J. Kurths. 2007. Recurrence plots for the analysis of complex systems. *Physics Reports* **438**, 237–329.

Mooij, J., D. Janzing, P. J., and B. Schölkopf. 2009. Regression by dependence minimization and its application to causal inference in additive noise models. *Proceedings of the 26th International Conference on Machine Learning (ICML 2009)*.

Papale, D., and R. Valentini. 2003. A new assessment of european forests carbon exchanges by eddy fluxes and artificial neural network spatialization. *Global Change Biology* **9**, 525–535.

Piao, S. L., P. Ciais, P. Friedlingstein, P. Peylin, M. Reichstein, S. Luyssaert, H. Margolis, J. Y. Fang, A. Barr, A. P. Chen, A. Grelle, D. Y. Hollinger, T. Laurila, A. Lindroth, and T. Richardson. 2008. Net carbon dioxide losses of northern ecosystems in response to autumn warming. *Nature* **451**, 49–52.

Proulx, R., and L. Parrott. 2008. Measures of structural complexity in digital images for monitoring the ecological signature of an old–growth forest ecosystem. *Ecological Indicators* **8**, 270–284.

Reichstein, M., , D. Papale, R. Valentini, M. Aubinet, C. Bernhofer, A. Knohl, T. Laurila, E. Lindroth, and G. Pilegaard. 2007a. Determinants of terrestrial ecosystem carbon balance inferred from european eddy covariance flux sites. *Geophysical Research Letters* **34**, L01402.

Reichstein, M., and C. Beer. 2008. Soil respiration across scales: The importance of a model-data integration framework for data interpretation. *Journal of Plant Nutrition and Soil Science* **171**, 344–354.

Reichstein, M., E. Falge, D. Baldocchi, D. Papale, R. Valentini, M. Aubinet, P. Berbigier, C. Bernhofer, N. Buchmann, T. Gilmanov, A. Granier, T. Grünwald, K. Havrnkov, D. Janous, A. Knohl, T. Laurila, A. Lohila, D. Loustau, G. Matteucci, T. Meyers, F. Miglietta, J.-M. Ourcival, S. Rambal, E. Rotenberg, M. Sanz, G. Seufert, F. Vaccari, T. Vesala, and D. Yakir. 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology* **11**, 1–16.

Reichstein, M., P. Ciais, D. Papale, R. Valentini, S. Running, N. Viovy, W. Cramer, A. Granier, J. Ogé, V. Allard, M. Aubinet, C. Bernhofer, N. Buchmann, A. Carrara, T. Grünwald, B. Heinesch, P. Keronen, A. Knohl, D. Loustau, G. Manca, G. Matteucci, F. Miglietta, J. M. Ourcival, K. Pilegaard, S. Rambal, S. Schaphoff, G. Seufert, J.-F. Soussana, M.-J. Sanz, E. D. Schulze, T. Vesala, and M. Heimann. 2007b. Reduction of ecosystem productivity and respiration during the european summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis. *Global Change Biology* **13**, 634–651.

Richardson, A., B. Braswell, D. Hollinger, P. Burman, E. Davidson, R. Evans, L. Flanagan, J. W. Munger, K. Savage, S. P. Urbanski, and S. C. Wofsy. 2006. Comparing simple respiration models for eddy flux and dynamic chamber data. *Agricultural and Forest Meteorology* **141**, 219–234.

Richardson, A. D., D. Y. Hollinger, J. D. Aber, S. V. Ollinger, and B. H. Braswell. 2007. Environmental variation is directly responsible for short– but not long–term variation in forest–atmosphere carbon exchange. *Global Change Biology* **13**, 788–803.

Roweis, S., and L. Saul. 2000. Nonlinear dimensionality reduction by locally linear embedding.. *Science* **290**, 2323–2326.

Ruddell, B. L., and P. Kumar. 2009a. Ecohydrologic process networks: 1. identification. *Water Resources Research* **45**, W03419.

Ruddell, B. L., and P. Kumar. 2009b. Ecohydrologic process networks: 2. analysis and characterization. *Water Resources Research* **45**, W03420.

Schmid, H., R. Luder, B. Naef-Daenzer, R. Graf, and N. Zbinden. 1998. *Schweizer Brutvogelatlas. Verbreitung der Brutvgel in der Schweiz und im Frstentum Liechtenstein 1993–1996*. Sempach, Schweizerische Vogelwarte.

Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär. 2006a. Land–atmosphere coupling and climate change in Europe. *Nature* **443**, 205–209.

Seneviratne, S. I., R. D. Koster, Z. Guo, P. A. Dirmeyer, E. Kowalczyk, D. Lawrence, P. Liu, C.-H. Lu, D. Mocko, K. W. Oleson, and D. Verseghy. 2006b. Soil moisture memory in AGCM simulations: Analysis of Global Land-Atmosphere Coupling Experiment (GLACE) data. *Journal of Hydrometeorology* **7**, 1090–1112.

Stoy, P. C., A. D. Richardson, D. D. Baldocchi, G. G. Katul, J. Stanovick, M. D. Mahecha, M. Reichstein, M. Detto, B. E. Law, G. Wohlfahrt, N. Arriga, J. Campos, J. H. McCaughey, L. Montagnani, K. T. Paw U, S. Sevanto, and M. Williams. 2009. Biosphere-atmosphere exchange of co_2 in relation to climate: a cross-biome analysis across multiple time scales. *Biogeosciences Discussions* **6**, 4095–4141.

Strobl, C., A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* **9**, 307.

Strobl, C., A.-L. Boulesteix, Z. A., and T. Hothorn. 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* **8**, 25.

Sulkava, M. 2008. *Learning from environmental data: methods for analysis of forest nutrition time series*. D.Sc. thesis. Helsinki University of Technology, Dissertations in Computer and Information Science, Report D24. Espoo, Finland.

Sulkava, M., S. Luyssaert, P. Rautio, I. A. Janssens, and J. Hollmén. 2007. Modeling the effects of varying data quality on trend detection in environmental monitoring. *Ecological Informatics* **2**, 167–176.

Tenenbaum, J. B., V. de Silva, and J. C. Langford. 2000. A global geometric framework for nonlinear dimensionality reduction. *Science* **290**, 2319–2323.

Todorovski, L., and S. Džeroski. 2006. Integrated knowledge–driven and data-driven approaches to modeling. *Ecolological Modeling* **194**, 3–13.

van der Maaten, L., and G. Hinton. 2008. Visualizing data using t–SNE. *Journal of Machine Learning Research* **9**, 2579–2605.

Vesanto, J., and M. Sulkava. 2002. Distance matrix based clustering of the self-organizing map. In *Artificial Neural Networks – ICANN 2002: International Conference, Proceedings*. J. R. Dorronsoro (ed.). Vol. 2415 of *Lecture Notes in Computer Science*. Springer-Verlag. Madrid, Spain. Pp. 951–956.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta. 2003. Towards reduced uncertainty in conceptual rainfall–runoff modeling: Dynamic identifiability analysis. *Hydrol. Processes* **17**, 455–476.

Williams, M., A. D. Richardson, M. Reichstein, P. C. Stoy, P. Peylin, H. Verbeeck, N. Carvalhais, M. Jung, D. Y. Hollinger, J. Kattge, R. Leuning, Y. Luo, E. Tomelleri, C. Trudinger, and Y.-P. Wang. 2009. Improving land surface models with fluxnet data. *Biogeosciences Discussions* **6**, 2785–2835.

Yang, F., K. Ichii, M. A. White, H. Hirofumi, A. Michaelis, P. Votava, A.-X. Zhu, A. Huete, S. W. Running, and R. R. Nemani. 2007. Developing a continental scale measure of gross primary production by combining MODIS and AmeriFlux data through support vector machine approach. *Remote Sensing of Environment* **110**, 109–122.