VALPRED 2021

Assessment of ensemble forecasts (validation des prévisions probabilistes)

CAES du CNRS,24, rue du Coin
73500, Aussois, France
☎ *(+33) 4 79 20 42 05*

Topic

This workshop will deal with the topic of validation of ensemble forecasting. In a nutshell, our goal is to mix around 30 researchers, statisticians, meteorologists, energy providers and other end users who have a common interest in probability forecasting, especially in the assessment of different forecasts. The non-exhaustive list of potential topics will be: scoring rules theory, forecasting methods, extreme value theory, decision theory.

General information and important dates

The school will start on Monday 4th of October at lunch time (12:30) and will end on Thursday 7th of October (after lunch around 13:30). Registration, meals and lodging there will be covered during this period.
Check-in starts on Monday morning (front desk at the Aussois center).

For local information regarding the center, contact Christa Balzer (christa.balzer@caes.cnrs.fr)
Tel : +33 (0) 4 79 20 42 05 Fax : +33 (0) 4 79 20 30 44
or use the website *centre-paul-langevin*
For scientific information regarding the workshop, check the *VALPRED website* , or contact the local organising committee: Clément Dombry (clement.dombry@univ-fcomte.fr), Philippe Naveau (Philippe.Naveau@lsce.ipsl.fr) or Olivier Wintenberger (olivier.wintenberger@upmc.fr).

Transports

The Paul Langevin conference centre is located at an elevation of 1500 m, 7 km from the city of Modane with direct rail links to Paris and Turin. From the train station *Modane*, it is possible to take taxis to go from Modane to Aussois. For example, vans from the *Marius taxi* company can transport many participants coming from the same train (saving money and CO2 but you need to reserve them).

Meals

8:00-9:00 Breakfast
12:30-13:30 Lunch
19:15-20:00 Diner

Internet connections

Wifi code: centrepaullangevin73500

Scientific program

The workshop will take place in the conference room “LA PARRACHÉE".

	MONDAY	TUESDAY	WEDNESDAY	THURSDAY
9 : 00		*Mini-course* **Johanna Ziegel** *Statistical forecast evaluation*	*Mini-course* **Johanna Ziegel** *Statistical forecast evaluation*	*Mini-course* **Johanna Ziegel** *Statistical forecast evaluation*
10 : 00		Coffee break	Coffee break	Coffee break
10 : 30		*Validation* **Jochen Broecker** *Uniform tests for forecast reliability* **Alexander Henzi** *Sequentially valid tests for forecast calibration* **Thibault Modeste** *Validation of dynamic probabilistic forecast*	*Risk* **Amandine Pierrot** *Wind power forecasting: nonlinearity, nonstationarity and varying bounds* **Antoine Heranval** *Application of Generalized Pareto Regression Trees to the cost prediction of floods in France* **Maximilian Aigner** *A competing risks interpretation of Hawkes processes*	*Evaluation* **Juliette Legrand** *Evaluation of binary classifiers for extremes* **Sam Allen** *Evaluating forecast for high impact events using transformed kernel scores*
12 : 30	Lunch	Lunch	Lunch	Lunch
14 : 00	*Machine Learning* **Benedikt Schulz** *Machine learning methods for post-processing ensemble forecasts of wind gusts* **Joseph de Vimarest** *VIKING: Variational Bayesian Variance Tracking* **Olivier Pasche** *EQRNN: Extreme Quantile Regression Neural Networks* **Edossa Merga Terefe** *Extremal Random Forest*	Informal time[1]	Informal time	
16 : 00	Coffee Break	Coffee Break	Coffee Break	
16 : 30	*Spatial extremes* **Philomène Le Gall** *Regional Frequency Analysis applied to climate change detection and attribution* **Pauline Rivoire** *Extreme daily precipitation in Europe estimated using regionalised extreme value distributions* **Carolin Forster** *Non-stationary max-stable models with an application to heavy rainfall data* **Max Thannheimer** *Bayesian Inference for Multivariate Extremal-t Distributions*	*Time series* **Gloria Buriticà** *An analysis of consecutive extremes records for stationary time series* **Paula Gonzalez** *Record analysis of climate models in a non-stationary context* **Nicklas Werge** *AdaVol: An Adaptive Recursive Volatility Prediction Method* **Juraj Bodik** *Detection of causality in time series using extreme values*	*Extremes* **Anne Sabourin** *Tail inverse regression for dimension reduction with extreme response* **Charles Tillier** *Boundary regression models* **Fabien Baeriswyl** *Multivariate regular variation in marked Hawkes processes* **Davit Varron** *A class of empirical measures suited to threshold-based statistics in extreme value theory*	
	18 : 45 Aperitive at the bar	19 : 00 Dinner	19 : 00 Dinner	
	19 : 30 Dinner	20 : 00 *ANR T-REX*[2]	20 : 00 *Fun time*[3]	

1. Casual interactions among all participants are fundamental to built scientific and collaborative networking and to create new ideas. This is a time to chit-chat when hiking in the surroundings of Aussois or anything you like.
2. Informal discussion between the members of the ANR project T-REX .
3. If you have any musical talents, feel free to bring your favorite instrument. Everyone will appreciate a little bit of music or singing.

Detailed program

Tuesday, Wednesday, Thursday 09:00-10:00 – Mini-course

Johanna Ziegel: Statistical forecast evaluation

Predictions for uncertain future events are ubiquitous, and form the basis of many important decisions. Therefore, objective evaluation of predictive performance is important and should not depend on the way in which forecasts are produced. Generally, forecasts are evaluated with respect to absolute criteria (calibration), where assess the statistical compatibility of the forecasts with the observations. Comparison of forecasts should be done with proper scoring rules or consistent scoring functions. I will review notions of forecast calibration and their assessment in prediction problems. The important special case of predictions for binary events will be treated and then real-valued outcomes and multivariate outcomes will be discussed. I will review proper scoring rules, present prominent examples and discuss recent new developments. Furthermore, their limitations concerning tail behavior of forecasts is treated. Finally, I will discuss the important concept of elicitability with regard to the evaluation of point predictions.

Monday 14:00-16:00 – Machine learning (chair O. Wintenberger)

Benedikt Schulz - Machine learning methods for postprocessing ensemble forecasts of wind gusts

Postprocessing ensemble weather predictions to correct systematic errors has become a standard practice in research and operations. However, only few recent studies have focused on ensemble postprocessing of wind gust forecasts, despite its importance for severe weather warnings. Here, we provide a comprehensive review and systematic comparison of eight statistical and machine learning methods for probabilistic wind gust forecasting via ensemble postprocessing, that can be divided in three groups: State of the art postprocessing techniques from statistics (ensemble model output statistics (EMOS), member-by-member postprocessing, isotonic distributional regression), established machine learning methods (gradient-boosting extended EMOS, quantile regression forests) and neural network-based approaches (distributional regression network, Bernstein quantile network, histogram estimation network). The methods are systematically compared using six years of data from a high-resolution, convection-permitting ensemble prediction system that was run operationally at the German weather service, and hourly observations at 175 surface weather stations in Germany. While all postprocessing methods yield calibrated forecasts and are able to correct the systematic errors of the raw ensemble predictions, incorporating information from additional meteorological predictor variables beyond wind gusts leads to significant improvements in forecast skill. In particular, we propose a flexible framework of locally adaptive neural networks with different probabilistic forecast types as output, which not only significantly outperform all benchmark postprocessing methods but also learn physically consistent relations associated with the diurnal cycle, especially the evening transition of the planetary boundary layer.

Joseph de Vilmarest: VIKING: Variational Bayesian Variance Tracking

We consider the problem of time series forecasting and we focus on state-space representations. We introduce a novel approach, named VIKING, to deal with state-space model inference under unknown observation and process noise variances. An augmented dynamical model is considered where the variances are represented as latent variables. The inference relies on the variational bayesian framework. Second-order bounds are used to derive recursive updates.

Olivier Pasche: EQRNN: Extreme Quantile Regression Neural Networks

Predicting conditional quantiles at extreme probability levels is challenging as classical quantile regression methods tend to fail due to the scarcity of training data in the response's tail region. Asymptotic results from extreme value theory, such as peaks over threshold, allow to extrapolate quantile values outside the range of the data. We propose to use neural networks for conditional peaks over threshold parameter estimation, yielding a model capable of both extrapolation, and dependence on a large set of predictors. Extreme Quantile Regression Neural Network models therefore allow modelling extreme conditional quantiles for complex high dimensional mechanisms by taking advantage of the neural network and deep learning methodologies' flexibility and versatility. The proposed methodology is shown to outperform existing methods on simulations.

Edossa Merga Terefe: Extremal Random Forest

Methods from statistics and machine learning for quantile regression fail in cases where the quantile of interest is so extreme that only few or no training data points exceed it. To overcome this problem, asymptotic results from extreme value theory can be used to extrapolate beyond the range of the data, and several approaches exists that use linear regression, kernel methods or generalized additive models. All of these methods break down if the predictor space has more than a few dimensions or if the regression function of extreme quantiles is complex. We propose a method for extreme quantile regression that combines the flexibility of random forests with the theory of extrapolation. Our extremal random forest estimates the parameters of a generalized Pareto distribution, conditional on the predictor vector, by maximizing a weighted likelihood, where the localizing weights are extracted from a quantile random forest. We further penalize the shape parameter in this likelihood to regularize its variability in the predictor space. A range of simulation setups shows that our ERF outperforms both the classical machine learning methods for quantile regression and the existing regression approaches from extreme value theory.

Monday 16:30-18:30 – Spatial Extremes (chair C. Dombry)

Philomène Le Gall:Regional Frequency Analysis applied to climate change detection and attribution

Rainfall are subject to various spatial features and their intensities can be highly variable, especially at the global scale. A recurrent question from risk analysis and climatologists is to determine how heavy rainfall patterns will differ around 2100. From a statistical point of view, partitioning the globe into homogeneous sub-regions is a delicate task, especially in regards to the modeling of heavy rainfall features in climate change detection and attribution. More precisely, different global climate models have to be combined and two scenarii (with and without anthrogenic forcings) have to be compared. In this work, our main goal is to propose and study a fast and efficient clustering algorithm that partitions the globe into homogeneous sub-regions. The main originality of the method is the coupling between the scale-invariance principle in classical regional frequency analysis with the dependence strength between two locations. Compared to classical regional frequency analysis techniques, a key aspect is that our algorithm does not rely on the *a priori* choice of covariates. While being in compliance with extreme value theory, we do not impose a parametric form on rainfall distributions. We apply this algorithm on multi-model yearly maxima of daily rainfall from 16 Coupled Model Intercomparison Project (CMIP). A method to handle and summarize such ensembles of data is proposed, as well as a comparison of the spatial clustering between two different experiments (with or without anthropogenic forcing). The partitions are compared to those obtained with partitioning method only focusing on margins or dependence. Our clustering algorithm leads to more coherent regions which are climatologically and physically consistent. Finally, we propose a detailed detection and attribution analysis of how the clustering changes between the two scenarii (with and without anthrogenic forcings).

Pauline Rivoire: Extreme daily precipitation in Europe estimated using regionalised extreme value distributions

We provide return level estimates of seasonal daily precipitation for Europe derived from the ERA-5 reanalysis data. This information is needed for a number of hydrological planning purposes (protective infrastructure, dams, retention basins) but is also an area of interest when considering climate change and associated changes in extreme precipitation and associated impacts. Because of the size of the data set (more than 20‘000 land grid points), estimating pointwise return levels is quite onerous and involves the estimation of many parameters. One ambition of statistical climatology is to summarize complex spatio-temporal variations of climatic variables and their extremes using a small number of statistical parameters. Regional Frequency Analysis addresses this challenge by defining homogeneous regions consisting of points with similar distributions of extremes (up to a normalizing factor) and developing so-called sparse regional models, i.e. models that require few parameters. We apply a recently developed method to cluster precipitation data in space and fit a sparse regional model to continent-wide precipitation data with the following objectives: i) come up with a “cooking recipe” of how to apply the method to a continental-scale data set (including discussion of limitations and practical challenges and a git hub repository), ii) illustrate the gain in reducing the number of parameters in the statistical model by the new method, iii) and produce seasonal return-level estimates for precipitation in Europe using both estimates based on the new pooled method for land areas. The homogeneous regions that we identify depend on the local orography. We compare them to homogeneous regions identified in previous works that focused on the national scale. An assessment of local and regional models indicates that a relatively parsimonious model with only a spatially varying scale parameter can compete very well against complex models.

Carolin Forster: Non-stationary max-stable models with an application to heavy rainfall data

In recent years, max-stable processes have become a popular choice for modeling spatial extremes because they arise as the asymptotic limit of rescaled maxima of independent and identically distributed random processes. Apart from few exceptions for the class of extremal-t processes, existing literature mainly focuses on models with stationary dependence structures. In this talk, we propose a novel non-stationary approach that can be used for both Brown-Resnick and extremal-t processes – two of the most popular classes of max-stable processes – by including covariates in the corresponding variogram and correlation functions, respectively. We apply our new approach to extreme precipitation data in Germany and compare the results to existing stationary models in terms of Takeuchi's information criterion (TIC). Our results indicate that, for this case study, non-stationary models are more appropriate than stationary ones.

Max Thannheimer: Bayesian Inference for Multivariate Extremal-t Distributions

Multivariate extreme value distributions are popular models for spatial extreme events. Due to the complicated structure of the full likelihood in higher dimensions, likelihood based inference is challenging. One common approach is the usage of pairwise likelihood estimators which are easier to handle but reduce the efficiency. In this paper, we follow the approach by Dombry et al. (2017) and Thibaud et al. (2016) and introduce a Bayesian framework with partitions as latent variables which allows for the use of full likelihood information. We focus on the popular extremal-t model and discuss a computationally efficient implementation of MCMC algorithms for sampling from the posterior distribution. The efficiency of the resulting estimators is compared to the pairwise likelihood estimator in a simulation study.

References:

- Clement Dombry, Sebastian Engelke, and Marco Oesting. Bayesian inference for multivariate extreme value distributions. Electronic Journal of Statistics, 11(2):4813-4844, 2017.
- Emeric Thibaud, Juha Aalto, Daniel S. Cooley, Anthony C. Davison, and Juha Heikkinen. Bayesian inference for the BrownResnick process, with an application to extreme low temperatures. The Annals of Applied Statistics, 10(4):2303-2324, 2016.

Tuesday 10:30-12:00 – Validation (chair P. Naveau)

Jochen Broecker: Uniform tests for forecast reliability

Calibration (or reliability) of forecasting systems is a hypothesis about an entire functional relationship. A calibrated probability forecast for binary events for instance should equal the conditional probability of the event given the forecast, whatever the value of the forecast. A new class of tests is discussed which use the supremum of the cumulative deviations from this relationship as test statistic. The new tests are compared with established regression based tests (analytically and numerically), and applications to operational weather forecasting systems are presented.

Alexander Henzi: Sequentially valid tests for forecast calibration

Forecasting and forecast evaluation are inherently sequential tasks, and forecast quality is often monitored continuously. However, the classical statistical methods for forecast evaluation are static, in the sense that tests for forecast calibration are only valid if the evaluation period is fixed in advance. Recently, e-values have been introduced as a new tool for statistical testing. E-values are valid in sequential settings and they can easily be transformed into p-values. We propose e-values for testing probabilistic calibration of forecasts. The e-values are competitive in power when compared to extant methods, and they provide useful additional insights in the evaluation of probabilistic weather forecasts.

Thibault Modeste: Validation of Dynamic Probabilistic Forecast

Forecast and its evaluation are major task in statistic. In real applications, forecast often take the form of a dynamic process evolving over time and this sequential point of view must be taken into account. We propose and discuss a minimal framework for dynamic probability forecast and its evaluation. Proper scoring rules are a crucial concept for probability forecast evaluation and we show, under minimal assumptions, that they can still be used in the dynamic framework because they are minimized, in the sense of the long term average score, by the ideal forecast. Another strategy for forecast evaluation is calibration theory based on the probability integral transform. Here ideal forecast is characterized by conditional calibration and we present some new tests based on regression trees and give some asymptotic behaviour of these tests. Then we compare to the ones proposed by Straehl and Ziegel (EJS 2017) in the framework of cross-calibration.

Tuesday 16:30-18:30 – Time series (chair V. Chavez)

Gloria Buritica: An analysis of consecutive extremes records for stationary time series

We study short periods with consecutive extreme records known as extreme blocks and focus on functionals acting on these blocks. For example, the extremal index: a classical measure of clustering of extremes, can be recovered this way. We derive large deviation principles of extreme blocks and reconsider cluster inference using blocks estimators based on exceedances of the p-norm of blocks.

Paula Gonzalez: Record analysis of climate models in a non-stationary context

Extreme climate events, such as heatwaves and heavy precipitations, have high societal and environmental impact. In the context of climate change, estimating how the likelihood of extreme climate records will be evolving in the future is important for mitigation and adaptation policies. For a given climate variable, the probability of breaking a record has been modeled using Extreme Event Attribution methodologies. Yet these approaches do not take into consideration temporal non-stationarity. I will present how Record Event Attribution can be combined with non-stationary inference techniques in order to project future record probabilities taking into account non-linear trends.

Nicklas Werge: AdaVol: An Adaptive Recursive Volatility Prediction Method

Quasi-Maximum Likelihood (QML) procedures are theoretically appealing and widely used for statistical inference. While there are extensive references on QML estimation in batch settings, it has attracted little attention in streaming settings until recently. An investigation of the convergence properties of the QML procedure in a general conditionally heteroscedastic time series model is conducted, and the classical batch optimization routines extended to the framework of streaming and large-scale problems. An adaptive recursive estimation routine for GARCH models named AdaVol is presented. The AdaVol procedure relies on stochastic approximations combined with the technique of Variance Targeting Estimation (VTE). This recursive method has computationally efficient properties, while VTE alleviates some convergence difficulties encountered by the usual QML estimation due to a lack of convexity. Empirical results demonstrate a favorable trade-off between AdaVol's stability and the ability to adapt to time-varying estimates for real-life data.

Juraj Bodik: Detection of causality in time series using extreme values

We deal with the following problem: Let us have two stationary (possibly nonlinear) time series with heavy-tailed marginal distributions. We want to detect whether there is some Granger causality present. Even more, we want to determine the minimal lag, i.e. the time how much it takes for information to travel from one time series to another. We will examine the asymmetry in extremes between the cause and effect, and present a statistic that can estimate such asymmetries. The basis of the idea stands by the so-called causal tail coefficient for time series, which in some way represents the behaviour in extremes of one series conditioned on the presence of an extreme in the other.

Wednesday 10:30-12:00 – Risk (chair Y. Goude)

Amandine Pierrot: Wind power forecasting: nonlinearity, nonstationarity and varying bounds

There is increasing interest in very short-term and higher-resolution wind power forecasting (from mins to hours ahead), especially offshore. Statistical methods are of utmost relevance, since weather forecasts cannot be informative for those lead times.
Wind power generation is a stochastic process that is not only non-stationary and non-linear but also theoretically double-bounded by zero and the nominal power of the turbine. Accommodating those aspects may lead to improving both point and probabilistic forecasts. For example generalized logit-normal distributions are naturally suitable and flexible for double-bounded and non-linear processes and have proved to be most appropriate especially when coming to probabilistic wind power forecasting. However in practice the upper bound which censors wind power generation may also change in time, while being unknown, for example in case of curtailment actions.
First we introduce the full maximum likelihood framework in using generalized logit-normal distributions for wind power short-term forecasting. Then we perform a simulation study to show the impact of a wrong fixed bound assumption on synthetic time series scaled by a bound which varies over time. We focus on both point and probabilistic forecasting and evaluate the accuracy of the forecasts in the worst and best cases. In the former it is assumed that the upper bound is fixed while in the latter it is assumed that one has perfect knowledge of the actual time-varying bound. Then we define a new modelling framework in which the bound can be adaptively estimated from the data. In particular we consider Monte Carlo Expectation-Maximization algorithms.

Antoine Heranval: Application of Generalized Pareto Regression Trees to the cost prediction of floods in France

In this work we use Generalized Pareto Regression Trees the cost prediction of floods in France on a real dataset. The aim of this study is to improve the cost prediction of an event of floods, shortly after its occurrence, for the entire French market. Indeed, following a natural catastrophe, it can be difficult to evaluate the scale and the cost of an event. In order to do that we use GPD Regression Trees to have a special focus on the extreme events and to gain further insight on the heterogeneity of the severity of these events. Thanks to a partnership with the French Federation of Insurance (FFA), essentially with one of his dedicated technical body, the association of French insurance undertaking for natural risk knowledge and reduction (Mission Risques Naturels, MRN), we have access to a large volume of events. These events represent all the events of floods that have been acknowledged as in state of natural catastrophe in France for the past 20 years. The cost of these events comes from the claims reported by the insurance company.

Maximilian Aigner: A competing risks interpretation of Hawkes processes

We give a construction of the Hawkes process as a piecewise competing risks model. We argue that the most natural interpretation of the self-excitation kernel is the hazard function of a defective random variable. This establishes a link between desired qualitative features of the process and a parametric form for the kernel, which we illustrate using examples from the literature. Two families of cure rate models taken from the survival analysis literature are proposed as new models for the self-excitation kernel. Finally, we show that the competing risks viewpoint leads to a general simulation algorithm which avoids inverting the compensator of the point process or performing an accept-reject step, and is therefore fast and quite general.

Wednesday 16:30-18:30 – Extremes (chair A.-L. Fougères)

Anne Sabourin: Tail inverse regression for dimension reduction with extreme response

We consider the problem of dimensionality reduction for prediction of a target $Y \in R$ to be explained by a covariate vector $X \in R^p$ with a particular focus on extreme values of Y which are of particular concern for risk management. The general purpose is to reduce the dimensionality of the statistical problem through an orthogonal projection on a lower dimensional subspace of the covariate space. Inspired by the sliced inverse regression (SIR) methods, we develop a novel framework (TIREX, Tail Inverse Regression for EXtreme response) relying on an appropriate notion of tail conditional independence in order to estimate an extreme sufficient dimension reduction (SDR) space of potentially smaller dimension than that of a classical SDR space. We prove the weak convergence of tail empirical processes involved in the estimation procedure and we illustrate the relevance of the proposed approach on simulated and real world data.

Charles Tillier: Boundary regression models

In the context of nonparametric regression models with one-sided errors, we consider parametric transformations of the response variable in order to obtain independence between the errors and the covariates. In view of estimating the tranformation parameter, we use a minimum distance approach and show the uniform consistency of the estimator under mild conditions. The performance of the estimator is demonstrated on a small sample in a simulation study.

Fabien Baeriswyl: Multivariate regular variation in marked Hawkes processes

In this talk, we discuss the tail asymptotics of the clusters of marked Hawkes processes, that are self-exciting point processes describing situations in which points tend to cluster together. These models are notably used in insurance (e.g. to model claim arrivals in view of computing ruin probabilities), in epidemiology to model the spread of a disease or, more recently, in modelling tweets popularity (by considering the cascade of retweets) on Twitter. In particular, we emphasise how (multivariate) regular variation transfers from the immigrant and offspring processes as well as from the mark distribution to the distribution of the cluster itself, generalising a result due to Basrak, Wintenberger and Žugec (2019). We illustrate how the individual components of the process (i.e. the variables governing both the average number of children of an event and its mark) impact the model.

Davit Varron: A class of empirical measures suited to threshold-based statistics in extreme value theory

We propose a class of random measure that behave like the empirical measure when conditioned to another random phenomenon. We prove consistency (Glivenko-Cantelli) and asymptotic Gaussianity (Donsker) results under standard assumptions related to Vapnik-Chervonenkis theory. We then illustrate its utility by application to Peaks-Over-Threshold and Multivariate Regular Variation models.

Thursday 10:30-11:30 – Evaluation (chair M. Zamo)

Juliette Legrand: Evaluation of binary classifiers for extremes

In this work, we address the question of how to assess and compare classifiers with respect to their capacity to capture extreme occurrences. This is also related to the topic of scoring rules used in forecasting literature. In this context, we propose and study a risk function adapted to extremal classifiers. The inferential properties of our empirical risk estimator are derived under the framework of multivariate regular variation. We apply our framework to the analysis of extreme river discharges in the Danube river basin. (joint work with Philippe Naveau and Marco Oesting)

Sam Allen: Evaluating forecasts for high-impact events using transformed kernel scores

It is often of interest to evaluate a forecaster's ability to predict outcomes that have a large impact on the forecast user. Although weighted scoring rules have become a well-established tool to achieve this, such scores have been studied almost exclusively in the univariate case, with interest typically placed on extreme events. However, events considered to be extreme from a statistical perspective often do not generate a large impact; instead, an impact might result from the interaction of several moderate events, none of which are necessarily extreme. Compound weather events provide a good example of this. To assess forecasts made for high-impact events such as these, this work extends existing results on weighted scoring rules by introducing weighted multivariate scores. To do so, we utilise kernel scores. Kernel scores form a very general class of scoring rules based on conditionally negative definite kernels, and we leverage well-known properties of kernels in order to introduce a new approach to weight such scores. Furthermore, we demonstrate that the threshold-weighted continuous ranked probability score (twCRPS), arguably the most well-known weighted scoring rule, is a kernel score. This result leads to a convenient representation of the twCRPS when the forecast is an ensemble, and also permits a generalisation to other kernel scores, allowing us to introduce, for example, a threshold-weighted energy score and threshold-weighted variogram score. To illustrate the additional information that these weighted multivariate scoring rules provide, results are presented for a case study in which the weighted scores are used to evaluate MeteoSwiss daily precipitation accumulation forecasts, with particular interest on events that could lead to flooding.