

# Régression par analyse des valeurs latentes à noyau

Maxime BERAR<sup>1</sup>, Michel DESVIGNES<sup>2</sup>

<sup>1</sup>Laboratoire Mathématiques Appliquées à Paris 5,  
45 rue des Saints Pères, 75270 Paris Cedex 06, France

<sup>2</sup>Gipsa-lab,  
961 rue de la Houille Blanche, BP 46, 38402 GRENOBLE Cedex, France  
berar@math-info.univ-paris5.fr, michel.desvignes@gipsa-lab.inpg.fr

**Résumé** – Nous présentons dans cet article la régression par analyse des valeurs latentes (LRR) à noyau. En cela similaire à la régression par composantes principales (PCR) ou à la régression aux moindres carrés partiels (PLS), la régression par analyse des valeurs latentes est basée sur la formation de variables latentes, orthogonales entre elles, à partir des prédicteurs. Pour ce faire, la LRR réalise l'Analyse en Composantes Principales de matrices regroupant prédicteurs et variables à estimer. L'astuce noyau nous permet de construire des modèles de régression LRR non linéaires dans des espaces caractéristiques possiblement de grande dimension.

Après avoir décrit l'algorithme de la LRR à noyau, nous le comparons aux algorithmes des régressions PCR et PLS à noyaux ainsi qu'à leur versions linéaires sur des données réelles de reconstruction faciale. La reconstruction faciale consiste à prédire la forme du visage à partir de la forme du crâne. Chaque forme est décrite par un ensemble de points caractéristiques. Sur ces données, la LRR à noyau atteint des résultats de même ordre de grandeur que la PLS à noyau, avec moins de composantes que pour les régressions sur variables latentes linéaires.

**Abstract** – In this paper, a family of regularized least squares regression model is extended by the kernel latent root regression (LRR) model. Similar to principal components regression (PCR) and partial least square (PLS) regression, LRR is a method based on the projection of explanatory variables to the latent variables. It uses Principal Components Analysis on a matrix joining the predictors and the outputs. Kernel methods enable us to build a nonlinear LRR regression models in possibly high-dimensional feature spaces.

We give the theoretical description of the kernel LRR algorithm and we experimentally compare the algorithm with the existing kernel PCR and kernel PLS techniques on a data set used for facial reconstruction. This application consists in estimating the shape of the face know in the shape of the skull. Each shape is described by a set common set of landmarks. On the data set employed, kernel LRR achieves the same results as kernel PLS.

## 1 Introduction

Dans le cas de régressions linéaires sur des données corrélées ou quasi-colinéaires, les régressions sur variables latentes, combinaisons linéaires des prédicteurs, constituent une famille de méthodes souvent utilisées [1]. Parmi ces méthodes, la régression sur composantes principales (PCR) utilise les composantes principales des prédicteurs comme variables latentes. La régression par les moindres carrés partiels (PLS) construit les variables latentes en maximisant un critère basé sur la covariance entre prédicteurs et variables dépendantes. La régression par analyse des valeurs latentes (LRR, *Latent Root Regression* [2]) utilise les composantes principales d'une matrice regroupant les prédicteurs et les variables dépendantes. Les méthodes PCR et PLS ont été étendues au cas où les prédicteurs sont projetés dans un espace caractéristique à l'aide d'un noyau de Mercer : PCR à noyau, PLS à noyau [3]. Dans cet article, nous présenterons la LRR à noyau. Cette nou-

velle méthode sera illustrée et comparée à la PLS, à la PCR et à leurs extensions à noyaux à l'aide de données réelles servant à la reconstruction faciale.

## 2 Régression par analyse des valeurs latentes

### 2.1 Régression sur variables latentes

Soient  $n$  réalisations indépendantes des vecteurs  $\mathbf{x}_i$  dans  $\mathbb{R}^N$ ,  $\mathbf{y}_i$  dans  $\mathbb{R}^L$ , centrées de variance unitaire, et rassemblées dans les matrices des prédicteurs  $\mathbf{X}$  de taille  $n \times N$  et des variables dépendantes  $\mathbf{Y}$  de taille  $n \times L$ . Le modèle de régression linéaire multi-réponse est défini par :

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}, \quad (1)$$

où la matrice  $\mathbf{B}$  de taille  $N \times L$  contient les coefficients de régression, où  $\mathbf{E}$  est une matrice de bruit de taille  $n \times L$ .

Les éléments de la matrice  $\mathbf{E}$  sont tels que  $E(\mathbf{E}) = \vec{0}$  et  $var(\mathbf{E}) = \Sigma$ . L'estimation  $\hat{\mathbf{B}}$  des coefficients de régression est au sens des moindres carrés ordinaires :

$$\hat{\mathbf{B}}_{OLS} = \min_B \|\mathbf{Y} - \mathbf{X}\mathbf{B}\|_F^2 = (\mathbf{X}\mathbf{X})^{-1} \mathbf{X}\mathbf{Y}, \quad (2)$$

où  $\|\bullet\|_F^2$  désigne le carré de la norme de Frobenius, ie la somme des carrés des éléments de la matrice, et  $\mathbf{X}$  la transposée de  $\mathbf{X}$ . Etant donné une nouvelle réalisation  $\mathbf{x}_0$ , l'estimation de  $\mathbf{y}_0$  est donnée par

$$\mathbf{y}_0 = \mathbf{x}_0 \hat{\mathbf{B}}. \quad (3)$$

La solution au sens des moindres carrés est

$$\hat{\mathbf{C}}_{OLS} = (\mathbf{T}\mathbf{T})^{-1} \mathbf{T}\mathbf{Y}, \quad (4)$$

où  $\mathbf{T}\mathbf{T}$  est une matrice diagonale. Les méthodes basées sur des variables latentes substituent aux prédicteurs  $\mathbf{X}$  un ensemble de  $r$  vecteurs orthogonaux  $[\mathbf{t}_1 \dots \mathbf{t}_r] = \mathbf{T}$  dans  $\mathbb{R}^n$ , construits comme combinaisons linéaires des prédicteurs ( $\mathbf{T} = \mathbf{X}\mathbf{V}$ ). On cherche la matrice  $\mathbf{C}$  telle que :

$$\hat{\mathbf{C}}_{OLS} = \min_C \|\mathbf{Y} - \mathbf{T}\mathbf{C}\|_F^2 \quad (5)$$

Introduites dans le modèle linéaire

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E} = (\mathbf{X}\mathbf{V})\mathbf{C}_{OLS} + \mathbf{E};$$

on obtient les coefficients de régression suivants :

$$\hat{\mathbf{B}}_{LV} = \mathbf{V}(\mathbf{V}\mathbf{X}\mathbf{X}\mathbf{V})^{-1} \mathbf{V}\mathbf{X}\mathbf{Y}, \quad (6)$$

Les méthodes de régressions linéaires sur variables latentes (PCR, PLS, LRR ...) diffèrent quant à la manière d'obtenir l'ensemble des variables latentes. Le choix du nombre de composantes  $r$  est lui souvent réalisé par validation croisée.

## 2.2 Régression par analyse des valeurs latentes

La régression LRR se base sur les composantes principales de la matrice  $\mathbf{A} = [\mathbf{X}\mathbf{Y}]$  rassemblant les prédicteurs et les variables dépendantes pour construire les variables latentes. L'Analyse en Composantes Principales (ACP) extrait un ensemble de  $N+L$  vecteurs propres  $\mathbf{a}_i \in \mathbb{R}^{N+L}$  associés aux valeurs propres  $\lambda_i$  ( $\lambda_1 \geq \dots \geq \lambda_{N+L} \geq 0$ ), pour lesquels  $\lambda_i \mathbf{a}_i = \mathbf{A}\mathbf{A}\mathbf{a}_i$ . Chaque vecteur propre  $\mathbf{a}_i$  peut être décomposé comme la juxtaposition de deux vecteurs  $\mathbf{t}_i \mathbf{a}_i = [\mathbf{v}_i \mathbf{w}_i]$ ,  $\mathbf{v}_i \in \mathbb{R}^N$  et  $\mathbf{w}_i \in \mathbb{R}^L$ . La LRR utilise les vecteurs  $\mathbf{X}\mathbf{v}_i$  comme variables latentes [2]. Comme ces vecteurs ne sont pas nécessairement orthogonaux, les implémentations récentes de la LRR [4, 5] utilisent une procédure itérative pour construire chaque variable latente, similaire à celle utilisée pour la PLS multi-réponse. Dans ce cadre, la première variable latente est  $\mathbf{t}_1 = \mathbf{X}\mathbf{v}_1$ . Pour obtenir les variables latentes suivantes, on décompose

les espaces engendrés par  $\mathbf{X}$  et  $\mathbf{Y}$  suivant la projection selon  $\mathbf{t}_1$  ( $\Pi_{\mathbf{t}_1}$ ), définissant ainsi les matrices des résidus  $\mathbf{X}_1$ ,  $\mathbf{Y}_1$  :

$$\mathbf{X}_1 = \mathbf{X} - \Pi_{\mathbf{t}_1} \mathbf{X}, \quad (7)$$

$$\mathbf{Y}_1 = \mathbf{Y} - \Pi_{\mathbf{t}_1} \mathbf{Y}, \quad (8)$$

$$\Pi_{\mathbf{t}_1} = \frac{\mathbf{t}_1 \mathbf{t}_1^t}{\mathbf{t}_1^t \mathbf{t}_1} = \frac{\mathbf{X}\mathbf{v}_1 \mathbf{v}_1^t \mathbf{X}}{\mathbf{v}_1^t \mathbf{X}\mathbf{X}\mathbf{v}_1}. \quad (9)$$

On définit récursivement chaque nouvelle variable latente par :

$$\mathbf{t}_{i+1} = \mathbf{X}_i \tilde{\mathbf{v}}_{i+1} = \mathbf{X}\mathbf{v}_{i+1}, \quad (10)$$

où  $\tilde{\mathbf{v}}_{i+1}$  est formé des  $N$  premières coordonnées du premier vecteur propre de  $\mathbf{A}_i \mathbf{A}_i$  :

$$\mathbf{A}_i = [\mathbf{X}_i \mathbf{Y}_i], \quad (11)$$

$$\mathbf{X}_i = \mathbf{X}_{i-1} - \Pi_{\mathbf{t}_i} \mathbf{X}_{i-1} = \mathbf{X} - \sum_{j=1}^{i-1} \Pi_{\mathbf{t}_j} \mathbf{X} \quad (12)$$

$$\mathbf{Y}_i = \mathbf{Y} - \sum_{j=1}^{i-1} \Pi_{\mathbf{t}_j} \mathbf{Y}. \quad (13)$$

Pour un nombre  $r$  de variables latentes, les coefficients de régression sont alors :

$$\hat{\mathbf{B}}_{LRR} = \sum_{i=1}^r \frac{\mathbf{v}_i \mathbf{v}_i^t}{\mathbf{v}_i^t \mathbf{X}\mathbf{X}\mathbf{v}_i} \mathbf{X}\mathbf{Y}. \quad (14)$$

## 3 Régression par analyse des valeurs latentes à noyau

Soit la fonction caractéristique  $\Phi : x \in \mathbb{R}^N \rightarrow \Phi(x) \in \mathcal{F}$ . Notre but est de construire un modèle de régression linéaire LRR dans  $\mathcal{F}$ . A ce modèle correspond une régression non-linéaire dans  $\mathbb{R}^N$ . Cette transformation  $\Phi$  n'est cependant connue que par le noyau de Mercer  $k$  correspondant au produit scalaire dans  $\mathcal{F}$  :

$$k(x_i, x_j) = \Phi(x_i) \bullet \Phi(x_j),$$

Parmi ces noyaux, on trouve le noyau gaussien

$$k(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$$

de paramètre  $\sigma$  ou encore les noyaux polynomiaux de degré  $d$

$$k(x_i, x_j) = (1 + x_i \bullet x_j)^d.$$

L'espace caractéristique peut être de très grande dimension voir infini dans le cas de noyaux gaussiens et  $\Phi$  peut être implicite (noyau gaussien) ou explicite (noyau polynomiaux). Comme pour l'ACP à noyau [6], la matrice noyau  $\mathbf{K}$  est définie par  $K_{ij} = k(x_i, x_j)$ . La variable latente  $\mathbf{t}_1$  est obtenue à partir du premier vecteur propre  $\mathbf{u}_1$  de  $\mathbf{A}\mathbf{A}$  et de la matrice noyau  $\mathbf{K}$  :

$$\lambda_1 \mathbf{u}_1 = \mathbf{A}\mathbf{A}\mathbf{u}_1 = (\mathbf{K} + \mathbf{Y}\mathbf{Y})\mathbf{u}_1 \quad (15)$$

$$\mathbf{t}_1 = \Phi(\mathbf{X})\mathbf{v}_1 = \frac{1}{\sqrt{\lambda_1}} \Phi(\mathbf{X})\mathbf{t}_1 \Phi(\mathbf{X})\mathbf{u}_1 = \mathbf{K}\mathbf{u}_1. \quad (16)$$

où  $\Phi(\mathbf{X})$  de dimension  $n \times \mathcal{F}$  contient les  $n$  éléments  $\Phi(x_i)$ . En appliquant un schéma itératif similaire à celui utilisé précédemment, on obtient une suite de variables latentes

$$\mathbf{t}_i = \mathbf{K}\mathbf{u}_i = \mathbf{K}_i \tilde{\mathbf{u}}_i$$

avec  $\tilde{\mathbf{u}}_i$  premier vecteur propre de  $\mathbf{A}_i \mathbf{t}_i \mathbf{A}_i = \mathbf{K}_i + \mathbf{Y}_i \mathbf{t}_i \mathbf{Y}_i$  et  $\mathbf{K}_i = (\mathbf{Id} - \Pi_{\mathbf{t}_i}) \mathbf{K}_{i-1} (\mathbf{Id} - \Pi_{\mathbf{t}_i})$ .

Les coefficients de régressions  $\mathbf{B}_\Phi$  ne sont pas toujours explicites. Par contre, pour un nombre  $r$  de composantes, l'estimation de  $\mathbf{y}_0$  est :

$$\mathbf{t}_i \hat{\mathbf{y}}_0 = \Phi(\mathbf{x}_0) \mathbf{B}_\Phi \quad (17)$$

$$= \Phi(\mathbf{x}_0) \mathbf{t}_i \Phi(\mathbf{X}) \sum_{i=1}^r \frac{\mathbf{u}_i \mathbf{t}_i \mathbf{u}_i}{\mathbf{t}_i \mathbf{u}_i \mathbf{t}_i \mathbf{K} \mathbf{K} \mathbf{u}_i} \mathbf{t}_i \mathbf{K} \mathbf{Y} \quad (18)$$

$$= \mathbf{t}_i \mathbf{k}_0 \sum_{i=1}^r \frac{\mathbf{u}_i \mathbf{t}_i \mathbf{u}_i}{\mathbf{t}_i \mathbf{u}_i \mathbf{t}_i \mathbf{K} \mathbf{K} \mathbf{u}_i} \mathbf{t}_i \mathbf{K} \mathbf{Y}. \quad (19)$$

avec  $\mathbf{k}_0$ , vecteur contenant les quantités  $\{\mathbf{k}(\mathbf{x}_0, \mathbf{x}_j)\}_{j=1}^n$ .

## 4 Application à la reconstruction faciale

La reconstruction faciale cherche à prédire le visage correspondant à un crâne retrouvé sans tissu. Pour cette étude, nous disposons de 45 maillages du crâne et du visage, extraits de scanners de femmes entre 25 et 40 ans [7]. Sur chaque maillage, on mesure la position de couples (crâne, visage) de points anatomiques ou géométriques, situés sur des géodésiques ([8] (figure 1)). Les prédicteurs sont les 135 points du crâne, les variables dépendantes ceux du visage. La reconstruction faciale (dans un premier temps) consiste à appliquer les modèles de régression estimés aux points extraits sur un crâne retrouvé sans tissu ( $\mathbf{x}_0$ ) [9]. Une partie des méthodes actuelles sont basées sur l'utilisation d'un modèle linéaire, construit par ACP, de la variabilité de la position de points caractéristiques du crâne et du visage [9, 10], influencées par les travaux sur les modèles statistiques de formes [11].

Une procédure de type "leave-one-out" est utilisée. L'erreur calculée est la moyenne de l'erreur sur la position de chaque point. Nous comparons les résultats de la LRR à noyau à ceux des régressions sur variables latentes (PCR, PLS, LRR [4]) et ceux de leurs extensions aux méthodes à noyau (PCR à noyau, PLS à noyau [3]). Pour les méthodes à noyau, un noyau gaussien avec  $\sigma$  choisi comme la somme des écart-type des données est utilisé.

Pour les méthodes linéaires, les résultats présentés figure 2 corroborent la similarité des résultats entre LRR et PLS observée de la littérature [4] : des erreurs moyennes minimale de 3.9 mm pour 9 composantes pour la LRR ou la PLS. La PCR nécessite un nombre supérieur (13) de composantes pour une erreur de même ordre. Toutes ces méthodes sont instables pour un nombre de composantes

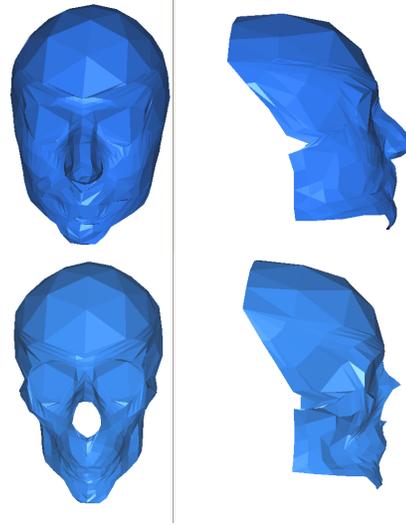


FIG. 1 – Maillages du crâne et de la face, formé par les points extraits

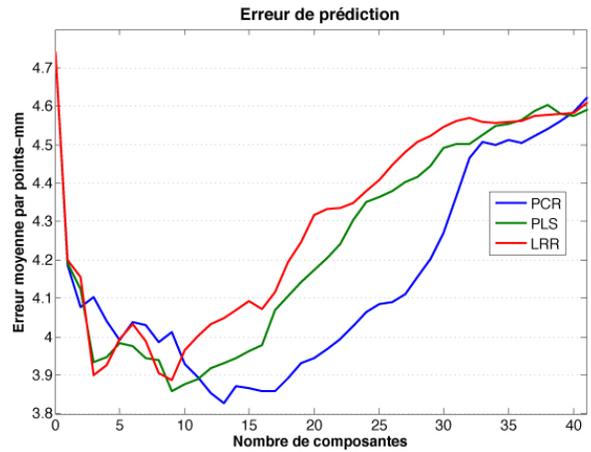


FIG. 2 – Erreur de prédiction pour les méthodes linéaires

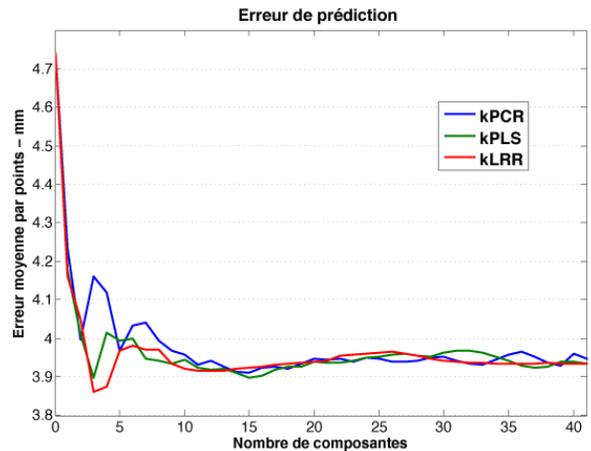


FIG. 3 – Erreur de prédiction pour les méthodes à noyau

élevé. La validation croisée est donc une étape importante dans la mise en place du modèle.

Les méthodes à noyau présentent des résultats plus stables, exposés figure 3. Les erreurs minimales sont du même ordre de grandeur que pour les méthodes linéaires, mais la sélection du bon nombre de composantes aura moins d'influence sur les résultats. Pour peu de composantes, les erreurs de prédiction de la LRR à noyau et de la PLS à noyau sont moins élevées que pour la PCR à noyau. La LRR à noyau présente donc des résultats proches de la PLS à noyau comme dans le cas linéaire.

La qualité de ces résultats pour une reconstruction faciale est insuffisante. La description fournie par les couples de points (crâne, visage) n'est pas assez précise pour permettre une identification des personnes et la mesure de l'épaisseur de la peau est très dépendante de la qualité de la reconstruction de la surface osseuse [7]. Des points caractéristiques supplémentaires sont nécessaires pour obtenir une représentation plus fine du visage. Des méthodes de mise en correspondance ou des modèles de distribution de points sont alors nécessaires [9]. Cependant, les apports de la régression par analyse des valeurs latentes à noyau observés sur ce premier jeu de données devraient permettre d'améliorer les résultats observés sur des jeux de données plus complets et obtenus par la régression par analyse des valeurs latentes (figure 4).

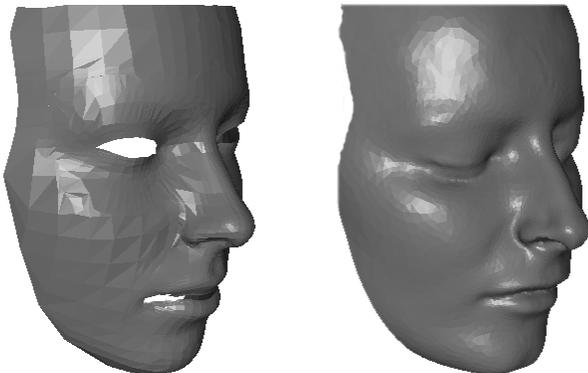


FIG. 4 – Exemple de reconstruction faciale.

## 5 Conclusions et perspectives

Nous avons présenté dans cet article l'extension à noyau de la régression par analyse des valeurs latentes. Celle-ci a été comparée sur des données réelles aux autres régressions sur variables latentes à noyaux (PLS à noyaux, PCR à noyaux) ainsi qu'à leurs versions linéaires. Les résultats sont conformes à ceux attendus au vu de la littérature : sur ces données, la LRR à noyaux est équivalente à la PLS à noyaux et l'utilisation d'un noyau gaussien offre une

certaine stabilité, notamment quant au choix du nombre de composantes .

Une des limites du formalisme noyau présenté ici est de ne pas projeter conjointement les prédicteurs et les variables dépendantes. Une perspective plus générale de ces travaux serait de réaliser cette projection conjointe pour toutes ces méthodes.

## Références

- [1] A. Burnham, R. Viveros et J. MacGregor. *Frameworks for latent variable multivariate regression*, Journal of Chemometrics, Vol. 10, 31-45,1996.
- [2] R. Gunst, J. Webster et R. Mason. *A comparison of least squares and latent root regression estimator*. Technometrics, Vol. 18, 75-83, 1976.
- [3] R. Rosipal et L. Trejo. *Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Space* Journal of Machine Learning Research, Vol. 2, 97-123, 2001.
- [4] E. Vigneau et M. Qannari. *A new algorithm for latent root regression analysis*, Computational Statistics & Data Analysis, Vol. 41, 231-242, 2002.
- [5] S. Bougeard , M. Hanafi et E. Qanari *Multiblock latent root regression. Application to epidemiological data*, Computational Statistics, Vol. 22, 209-222(14), 2007.
- [6] B. Schölkopf, A. Smola et K. Müller. *Nonlinear Component Analysis as a Kernel Eigenvalue Problem*, Neural Computation, Vol. 10, 1299-1319,1998.
- [7] F. Tilotta et al. *Construction and analysis of a head CT-scan database for craniofacial reconstruction* Forensic Science, 2008.
- [8] Y. Wang, B. Peterson et L. Staib. *Shape-Based 3D Surface Correspondence Using Geodesics and Local Geometry* IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 644-651, 2000.
- [9] M. Berar, M. Desvignes, G. Bailly et Y. Payan. *3D Semi Landmarks based statistical face reconstruction* Journal of Computing and Information Technology, Vol. 14 (2), pp 31-43, 2006.
- [10] P. Claes, D. Vandermeulen, S. De Greef, et al. *Craniofacial reconstruction using a combined statistical model of face shape and soft tissue depths : methodology and validation*. Forensic Science International, 2006.
- [11] T. Cootes, C. Beeston, G. Edwards et C. Taylor, *A Unified Framework for Atlas Matching Using Active Appearance Models*. Proc. Information Processing in Medical Imaging, pp. 323-333,1999.